



LIBRARY
Michigan State
University

This is to certify that the
dissertation entitled

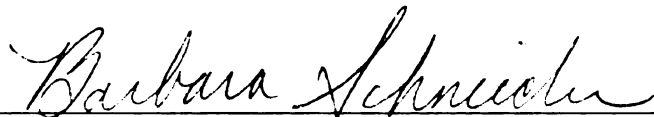
TEACHER TURNOVER AND UNDERSUPPLY: AN
ORGANIZATIONAL ANALYSIS

presented by

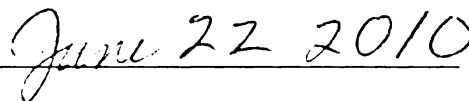
Venessa Ann Keesler

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Measurement and Quantitative
Methods



Major Professor's Signature



Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**TEACHER TURNOVER AND UNDERSUPPLY: A SCHOOL ORGANIZATIONAL
ANALYSIS**

By

Venessa Ann Keesler

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Measurement and Quantitative Methods

2010

ABSTRACT

TEACHER TURNOVER AND UNDERSUPPLY: A SCHOOL ORGANIZATIONAL ANALYSIS

By

Venessa Ann Keesler

This dissertation investigates the three components of teacher labor supply in the state of Michigan: teacher demand, teacher supply and retention, and teacher undersupply. These factors of teacher labor supply are considered as organizational characteristics of schools. In the first study, school-and subject-specific teacher demand to meet a specified set of graduation requirements was estimated using a formula I developed. This formula is tested under varying assumptions, and a revised formula is produced. The key finding from this study is the importance of using a projected enrollment estimate when calculating demand.

The second and most critical study focuses on school-level teacher retention. Using multilevel modeling and propensity score matching, the relationship between school-level teacher retention and student achievement and mobility is studied. This analysis finds that although high school-level teacher retention has a positive relationship with student achievement, this can be explained when including other covariates. However, high school-level teacher retention impacts student mobility, where students in schools with high school-level teacher retention are less likely to change schools. As student mobility is an important predictor for student achievement, school-level teacher retention impacts student achievement indirectly through the mechanism of student mobility. This finding was supported by the results of the propensity score match, which

demonstrates that, when comparing schools with a similar propensity to have high teacher retention, those schools that actually experienced high teacher retention had a decrease in student mobility. In other words, the “treatment” of high teacher retention is effective for decreasing a school’s rate of student mobility. These findings were both tested using sensitivity analyses and found to be reasonably robust to the potential impact of unmeasured confounds.

The final analysis investigates the predictors of teacher undersupply and the impacts of teacher undersupply. This study finds that higher levels of teacher undersupply are related to lower student achievement in mathematics, English language arts, and science. Although the state of Michigan does not have a net undersupply of teachers, this analysis provides evidence that the localized cases of undersupplied schools may struggle to produce student achievement gains due to their undersupplied status.

*Dedicated to my beautiful children, Rose and Jeffrey Jr., and to my husband, Jeffrey Sr.
With all my love*

ACKNOWLEDGEMENTS

This dissertation is the product of multiple years of experience with and exposure to these issues as part of the extensive supports provided to me by my advisor, Dr. Barbara Schneider. Barbara has been my dissertation advisor, mentor, and supervising professor for my entire graduate school experience, and has given me numerous opportunities to grow and learn, has challenged me to develop as a scholar and as a leader, and has provided the necessary encouragement and reinforcements I needed to complete this course of study and this dissertation. She brought me into this work on teacher labor supply through her REL-Midwest grant, and has given me abundant opportunities to engage with these data, with the state of Michigan, and with other researchers around these important topics. I am indebted to her for her unflagging support through this process.

I am also grateful for the ongoing support and encouragement of my academic advisor, Dr. Kenneth Frank. Ken's consistent methodological expertise, coupled with his ability to balance a substantive focus with rigorous methods, has been an invaluable resource through this dissertation process, and through my graduate career. I am grateful for the support of my committee, Dr. Richard Prawat and Dr. Cassandra Guarino, and appreciate their time and contribution to this work. Dr. Kimberly Maier, although not officially named as my advisor or committee member, has provided support and encouragement at key moments, and I appreciate her selflessness with her time and her concern.

No one writes a dissertation or completes graduate school without the devoted and long-suffering support of family and friends, and I am no exception. My husband, Jeff,

has been an unwavering support to me throughout the last five years and has been a true partner in this endeavor. My parents, Steve and Peggy, and my in-laws, Tom and Carolyn, have provided endless support, including loving and fabulous child care to my children. I want to acknowledge as well the support of my entire family: Stephen Rose, Najwa El-Nachef and Aaron Keesler, Brian and Abigail Keesler, Jeffrey Gulick, Louis and Janet Gulick, Jean Rose and all other members of the Rose, Gulick, Keesler and Craven families. Many friends have also stood by me and provided emotional and moral support, and I want to specifically thank Sara Gulick, Marisa Feyen, Jessica Reid, Heidi Ball, and all of the other friends who have listened to my dissertation stories. Leigh Ann Halas deserves special thanks, as always, for not only listening and supporting me, but also for helping me with the intricacies of formatting and references, and for being a lifelong and wonderful friend.

Finally, I want to thank my colleagues. I have received invaluable support from fellow (and former) graduate students, including Nathan Jones, Erin Grogan and Adam Maier. Christina Mazuca has been a listening ear through this process. I also want to thank and acknowledge Freda Lynn for her early work on the REL-Midwest project, and her excellent Stata knowledge that has been a huge asset for anyone working with these state data. Thanks go as well to Thomas Howell of CEPI and Joseph Martineau and Steven Viger of OEAA for their advice and guidance, and to Patricia King of OEAA for her listening ear.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	xi
CHAPTER 1: TEACHER SUPPLY, DEMAND, AND UNDERSUPPLY: BACKGROUND TO THE PROBLEM	1
Background to the Problem	1
Addressing teacher supply and demand using state administrative data	5
Foundations for the dissertation: is there an adequate supply of teachers to meet the Michigan Merit Curriculum?	7
Calculating supply	9
Calculating undersupply	9
Investigating Teacher Demand, Supply and Turnover, and Undersupply Using State Administrative Data: Dissertation Outline	11
CHAPTER 2: DATA, METHODS, AND MEASURES	13
Data	13
The Registry of Educational Personnel	14
The Single Record Student Database	15
Student assessment data	16
School-level data	17
Sample	18
Methods: Multilevel Modeling and Propensity Score Matching	19
The multilevel models to be estimated	20
Analytic strategy for multilevel models	24
Sensitivity analyses	25
Propensity scores: brief discussion of propensity scores as an analytic technique	26
Propensity scores: weighting method	29
Measures	32
Key predictors of interest: school-level turnover and undersupply	32
School-level covariates	33
Student-level covariates	34
CHAPTER 3: DESCRIPTIVE ANALYSES OF MICHIGAN’S TEACHER SUPPLY, TURNOVER, AND UNDERSUPPLY	36
Investigating Michigan’s Teacher Labor Supply: Previous Analyses	37
What are the demographic characteristics of Michigan’s teaching force?	37
License and endorsement	40
Teacher license and endorsement: are teachers teaching out of field? Is there a reserve supply of teachers to meet the demands	

of the MMC?	42
Endorsements and subject assignments—potential reserve	43
Michigan’s Overall Teacher Labor Supply	43
School-Level Teacher Retention by School Characteristics	44
School-level Undersupply by School Characteristics	46
Mathematics undersupply	47
English language arts and science undersupply	49
CHAPTER 4: ESTIMATING TEACHER DEMAND USING STATE ADMINISTRATIVE	
DATA: CHALLENGES AND RESOLUTIONS	68
Background to the Problem: Estimating Teacher Demand	68
Estimating demand to respond to changes in curricular requirements	70
Demand Formula, Assumptions, and Alterations	72
Assumptions	72
Testing each assumption	74
Testing and Validating the Formula	79
Analytic Strategy	80
Results	81
Part I: Choosing an enrollment estimate	81
Part II: Distributional assumptions for class size and courses per FTE: Addressing the issues with ratios	88
Do different demand formula assumptions produce estimates that vary by school characteristics?	97
How does the categorization of undersupply differ in each of the four situations?	98
Conclusions	102
CHAPTER 5: TEACHER CHURN RATE: THE RELATIONSHIP BETWEEN TEACHER	
RETENTION AND STUDENT ACHIEVEMENT	115
Introduction	115
Background to the Problem	116
Teacher turnover as an organizational characteristic of schools	117
Student retention: the other level of retention	120
Teacher turnover in high schools and its relationship to student achievement	121
The relationship between school characteristics, teacher turnover, and student outcomes: hypotheses and research questions	124
Analytic Methods	126
Data and Measures	127
Data source	127
Measures	128
Analytic Approach	131
The multilevel models to be estimated	132
Sensitivity analyses	136
Results	136

Descriptive statistics of key school variables by teacher	
retention rate.....	136
Mathematics achievement outcomes.....	137
Student mobility as an outcome.....	140
Sensitivity analyses.....	142
Limitations.....	144
Conclusions.....	145
CHAPTER 6: PROPENSITY ANALYSES OF THE EFFECTS OF TEACHER TURNOVER.....	154
Estimating the Effect of High School-Level Teacher Retention.....	154
Propensity scores: weighting method.....	155
Checking the weights/trimming the weights.....	156
Checking for balance.....	157
Propensity scores: stratification method.....	158
Propensity scores: nearest neighbor, kernel matching, and radius matching.....	158
Results.....	159
Sensitivity Analyses.....	161
Conclusions from propensity score sensitivity analyses.....	167
Limitations of this Analysis.....	168
Case Study Schools: An Exploratory Analysis.....	170
CHAPTER 7: THE CORRELATES AND IMPACTS OF SCHOOL-LEVEL TEACHER UNDERSUPPLY.....	182
Background to the Problem: Undersupply as an Organizational Characteristic of Schools.....	182
Analytic Methods.....	185
Part I: Model predicting undersupply.....	185
Findings: predictors of subject specific undersupply.....	186
The relationship between teacher undersupply and student achievement.....	189
Data source.....	190
Analytic Approach.....	192
The multilevel models to be estimated.....	192
Sensitivity analyses.....	194
Results.....	194
Mathematics achievement outcomes.....	194
English language arts achievement outcomes.....	197
Science achievement outcomes.....	197
Comparison of main effects under different demand estimations.....	199
Sensitivity Analyses.....	200
Conclusions and Limitations.....	202
CHAPTER 8: MOVING BEYOND TEACHER DEMAND, SUPPLY, AND UNDERSUPPLY: CONCLUSIONS AND DISCUSSIONS.....	217
Estimating Teacher Demand.....	217

Teacher Retention	218
Teacher Undersupply	219
Policy Implications	221
Looking Ahead	222
APPENDIX A: COMPLETE LIST OF VARIABLES AND SUMMARY STATISTICS	224
APPENDIX B: ADDITIONAL HISTOGRAMS FROM CHAPTER 4: ESTIMATING TEACHER DEMAND USING STATE ADMINISTRATIVE DATA: CHALLENGES AND RESOLUTIONS	233
APPENDIX C: ADDITIONAL INFORMATION ON PROPENSITY SCORE METHODS DISCUSSION OF VARIOUS PROPENSITY SCORE METHODS UTILIZED	244
APPENDIX D: SENSITIVITY ANALYSES	266
BIBLIOGRAPHY	269

LIST OF TABLES

Table 3.1: Characteristics of Michigan Public School Teachers by Full-Time Equivalency Status.....	51
Table 3.2: Characteristics of Michigan Schools by Teacher Full-Time Equivalency Status.....	53
Table 3.3: Characteristics of Teachers by Minority and Non-Minority Status.....	54
Table 3.4: Characteristics of Teachers by Education Degree Attainment	55
Table 3.5: Characteristics of Teachers by Locale	57
Table 3.6: License Demographics for Michigan's Instructional Workforce.....	59
Table 3.7: Out of Field in Core Subjects	61
Table 3.8: Teaching Assignment by Endorsement Pattern (Potential "Reserve" of Teachers).....	62
Table 3.9: Teacher Labor Supply in Michigan, 2006-2008.....	63
Table 3.10: Mean Teacher Retention Rate by High School Characteristics.....	64
Table 3.11: Mean Mathematics Teacher Undersupply by High School Characteristics	65
Table 3.12: Mean English language arts Teacher Undersupply by High School Characteristics.....	66
Table 3.13: Mean Science Teacher Undersupply by High School Characteristics	67
Table 4.1: Mean Mathematics Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions.....	105
Table 4.2: Mean English Language Arts Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions.....	107
Table 4.3: Mean Science Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions.....	109
Table 4.4: Categorization of Undersupply Under Four Sets of Assumptions	111
Table 4.5: Comparison of Cross-Categorizations of Undersupply Under Four Sets of Assumptions.....	112

Table 4.6: "Misclassification" Between Situation 1 and Situation 2 ("A" from Table 4.4)	113
Table 4.7: "Misclassification" Between Situation 2 and Situation 4 ("C" From Table 4.4)	114
Table 5.1: Variable Description and Summary Statistics	147
Table 5.2: Hierarchical Linear Model Predicting Student Mathematics Achievement as a Function of School-Level Teacher Retention	149
Table 5.3: Hierarchical Generalized Linear Model Predicting Student Mobility (1=different school from freshman-junior year)	151
Table 6.1: Testing for Balance in Propensity Score Estimation	174
Table 6.2: Logistic Regression for Estimating the Treatment Propensity Score (Treatment = School-Level Teacher Retention > 85%)	176
Table 6.3: Estimated Effect of High School-Level Teacher Retention on Average School-Level Student Mobility	177
Table 6.4: Sensitivity Analyses using Mantel-Haenszel Bounds for Effect on Student Mobility: Nearest Neighbor Matching	178
Table 6.5: Sensitivity Analyses using Mantel-Haenszel Bounds for Effect on Student Mobility: Stratification Method	179
Table 6.6: Comparison of School and Student Characteristics Between High- and Low-Retention Case Study Schools, and Between Case Study and Non-Case Study Schools	180
Table 7.1: OLS Regression Predicting Mathematics Teacher Undersupply	204
Table 7.2: Logistic Regression Predicting Significant Mathematics Teacher Undersupply	205
Table 7.3: OLS Regression Predicting ELA Teacher Undersupply	206
Table 7.4: Logistic Regression Predicting Significant ELA Teacher Undersupply	207
Table 7.5: OLS Regression Predicting Science Teacher Undersupply	208
Table 7.6: Logistic Regression Predicting Significant	209
Table 7.7: Hierarchical Linear Model Predicting Student Mathematics Achievement as a Function of Mathematics Teacher Undersupply	210

Table 7.8: Hierarchical Linear Model Predicting Student English Language Arts Achievement as a Function of English Language Arts Teacher Undersupply	211
Table 7.9: Hierarchical Linear Model Predicting Student Science Achievement as a Function of Science Teacher Undersupply	212
Table 7.10: Comparison of Main Effects for Math Undersupply on Student Mathematics Achievement Using Different Demand Formula Assumptions	216
Table 7.11: Comparison of Main Effects for English Language Arts Undersupply on Student ELA Achievement Using Different Demand Formula Assumptions	216
Table 7.12: Comparison of Main Effects for Science Undersupply on Student Science Achievement Using Different Demand Formula Assumptions	216
Appendix A: Complete List of Variable Descriptions and Summary Statistics	225

LIST OF FIGURES

Figure 3.1: Comparison of Racial Composition of Michigan’s Teaching Force to Michigan and the USA.....	38
Figure 4.1: Comparison of Mathematics Demand with Varying Enrollment.....	82
Figure 4.2: Comparison of Mathematics Undersupply with Different Enrollment.....	85
Figure 4.3: Comparison of Demand Estimates Under Four Estimation Scenarios.....	91
Figure 4.4: Comparison of Undersupply Under Four Estimation Scenarios.....	95
Figure 5.1: Conceptual Model for the Relationship Between Teacher Retention, Student Mobility, and Student Achievement Outcomes.....	124
Figure 5.2: Distribution of School-Level Teacher Retention Rates	129
Figure 5.3: Timing of the Longitudinal Data for Establishing a Causal Path.....	153
Figure 6.1: Box and Whisker Plot to Demonstrate Distribution of Weights	157

CHAPTER 1: TEACHER SUPPLY, DEMAND AND UNDERSUPPLY:

BACKGROUND TO THE PROBLEM

Ensuring an adequate supply of qualified teachers to meet instructional demands is an area of critical importance in education policy research. Teacher supply became a matter of research focus in the early 1980s, with two key demographic trends—increasing student enrollments and increasing teacher retirements—as the cause of a predicted teacher shortage (Darling-Hammond, 1984; Ingersoll, 2001). Some of the early research on the topic reflected a growing concern regarding increasing enrollments and simultaneous teacher retirements, and sought to establish methods and procedures for estimating the supply and demand accurately (Boe & Gilford, 1992; Arnold, Choy, & Bobbitt, 1993). The decades since have been devoted to attempting to quantify the size of this teacher supply (Ingersoll & Perda, 2009), to produce more rigorous methods for estimating supply (Santiago, 2002), and dealing with issues inherent in supply, such as teacher attrition and retention (Borman & Dowling, 2008; Dolton & van der Klaauw, 1999; Harris & Adams, 2007; Imazeki, 2005; Shen, 1997; Stinebrickner, 1998; 2002) and teacher labor market matching (Boyd, Lankford, Loeb, & Wyckoff, 2006; Stinebrickner, 2001).

Teacher demand has been influenced in part by the advent of merit curriculums. In order to meet the increased rigor necessitated by the No Child Left Behind legislation (2001), states have focused resources on ensuring that teachers are teaching in-field, and have also raised standards, both in terms of teacher quality and curricular requirements. An example of such curricular changes is expanded graduation requirements, such as increasing the number of math courses required to graduate, which helps equip students

for the transition to postsecondary education and the labor market. This focus is ongoing in other nationwide efforts, such as the Common Core Standards and the America Diploma Project, both of which seek to increase standards and expectations so that students finish high school with rigorous preparation necessary for success in postsecondary education and in the workforce. All of these changes in curricular requirements have an impact on the demand for teachers, both in terms of sheer quantity of teachers as well as teachers trained with specific skill sets.

Finally, the current economic climate and the dramatic “shocks” to the system make predicting teacher supply and demand even more problematic. The teaching force in America is not necessarily structured to be quickly responsive to sudden systemic shocks. Given recent economic developments, there are significant changes related to student population and enrollment, with some states losing students at a rapid rate due to out-migration. There are also new issues on the demand side, as schools may need more teachers but may be severely restricted by funding challenges.

In understanding teacher labor supply, there are three major components: demand, supply and retention, and the impact of undersupply. This dissertation addresses the overall question of teacher labor supply and its impact on student achievement and mobility by focusing on these three areas. To estimate teacher demand, a previously developed and utilized demand is tested and refined, and then used in undersupply analyses. While teacher supply in the original formula accounted for the number of FTEs of each subject area in each school, it did not take into account issues of teacher “churn rates.” For a school with undersupply, or even with an adequate supply, what are the teacher churn rates, and is there a revolving door of teacher entry and departure that

affects student achievement and mobility directly, and that interfaces with teacher undersupply? Finally, using the revised demand formula and considering teacher retention as a component of teacher supply, what is the impact of school-level teacher undersupply on student achievement outcomes?

This study conceptualizes teacher demand, supply, and undersupply from an organizational perspective, considering teacher labor supply at the school level as an organizational characteristic of each high school, and investigates the impact of this organizational characteristic on the effectiveness of the organization itself, as measured by student achievement. Much of the work on teacher demand, supply, and undersupply has been conducted at the state or national level, using nationally representative data sets. In fact, much of the seminal work on issues related to teacher supply utilized surveys such as Schools and Staffing, and suggested an upcoming national teacher shortage (Boe & Gilford, 1992; Darling-Hammond, 2000; Ingersoll, 2001, 2003). While it is important to understand overall teacher labor supply and demand at the national and state level, this approach can mask important differences that occur on a school-by-school basis. By contrast, the studies presented here all begin by conceiving of teacher supply, demand, and undersupply as school organizational characteristics, and utilize state administrative data from Michigan to take into account distinct organizational factors related to each factor of teacher labor supply (demand, supply, and undersupply) and student achievement.

As an organizational characteristic of the school, teacher supply and demand can contribute to the ability of a school to build trust among members, and to develop a sense of community. This draws on the sociology of education, which has shown that the

presence of a sense of community and cohesion among families, teachers, and students is important for the success of schools (e.g., Bryk & Schneider, 2002; Durkheim, 1961; Grant, 1988; Parsons, 1959; Rosenholtz, 1989, Waller, 1932). A body of evidence suggests that the community of the school has important implications for school performance and effectiveness (Bryk, Lee, & Smith, 1990; Coleman & Hoffer, 1987; Rosenholtz, 1989). Specifically, the communal nature of certain types of schools, such as private schools, can create an environment that reinforces shared values and that leads to higher levels of social capital (Coleman & Hoffer, 1987). This sense of community is created by shared values, trust, and reciprocity between individuals in the school (Bryk & Schneider, 2002). Schools in which there is a stronger sense of community and cooperation between students and teachers have been shown to have higher achievement levels, as well as a more equitable distribution of achievement (Lee, Bryk, & Smith, 1993; Lee & Smith, 1997).

How does teacher labor supply relate to the organizational culture of the school? More specifically, how can teacher labor supply be expected to impact student achievement as an organizational characteristic of a school? Prior research has shown that much of the variation in student achievement is due to within-school, rather than between-school, factors. Therefore, in order to estimate supply and demand and the impact those factors have on student achievement, the most appropriate level of analysis is at the school level. Teacher demand, particularly teacher demand as understood in the context of statewide graduation requirements, will vary by each school and subject, as will teacher supply. An important corollary to teacher supply is teacher turnover, as the relationship between student achievement and teacher supply is likely affected by teacher

turnover. Schools with high rates of turnover may be more disorganized as a culture, which in turn may lead to both decreased student achievement, as well as increased student mobility, as students respond to the disorganized nature of the school and leave in greater numbers. There may also be an indirect effect of teacher turnover on student achievement via student mobility, as teacher turnover may lead to increased student turnover, which is known to be linked to decrease achievement outcomes. The third component of teacher labor supply is undersupply. School-level undersupply in each subject is hypothesized to have a negative impact on student achievement outcomes.

Addressing teacher supply and demand using state administrative data

Since 2005, the Institute for Education Sciences (IES) has been awarding statewide longitudinal data systems (SLDS) grants to states under the Educational Technical Assistance Act of 2002. The purpose of these grants is to help states to manage, analyze and utilize education data. Fourteen states were awarded grants in 2005; 12 additional states and the District of Columbia received grants in 2007; and under the American Reinvestment and Recovery Act of 2009, 27 states were awarded an additional \$265 million in funding to further the work of the longitudinal data systems in the states. The result of this substantial effort on behalf of the federal government is that many states have rigorous, extensive, longitudinal data on students and teachers (Institute for Education Sciences, n.d.). Moreover, the SLDS grants, NCLB, and the new requirements under the American Reinvestment and Recovery Act and the Race to the Top competition all require that states, districts, and schools make use of the data they are collecting in order to influence policy and to support decision making.

One potentially fruitful use of this state administrative data is the study of teacher labor supply. These teacher data are comprehensive; it is possible to know who applies, who is granted a license, and what those licenses are in, along with certain demographic information on those individuals. This helps to address a repeated concern in teacher supply research, which is that the use of data from SASS or other nationally representative data sets do not allow for the direct estimation of the absolute size of a teacher shortage (Murphy, DeArmound, & Guin, 2003). Ingersoll and Perda (2009) point out that one of the problems with estimating supply for teachers is a lack of comprehensive data on “entry, licensing, and preemployment preparation”(p. 3).

The state of Michigan, via the Center for Educational Performance and Instruction (CEPI), collects and reports information related to the entire teaching workforce, and collects this information in the Registry of Educational Personnel. This database includes employment records for any individual hired by any educational entity in the state of Michigan. The data are longitudinal; they have been collected in the REP since 2002, but have become more rigorous and reliable within the last four years. This data source is discussed in greater detail in Chapter 2: Data and Methods.

Michigan also has a wealth of student data available to the research community. Student demographic data and achievement data are available from the 2003-2004 school year to the present, linked by a unique identifier code.¹ The achievement data include student test scores and performance levels of the state-administered MEAP (for grades 3-8) and MME (for high school). These data are discussed in greater detail in the data and methods chapter.

¹ Student demographic and achievement data are available prior to 2003-2004, but they are not linked by a unique identifier code, as this was not in place prior to 2003-2004.

These data are well-suited for use in analyzing this question for several reasons. First, they are universe data, and include *all* students, teachers, and schools in the state of Michigan. This removes a great deal of uncertainty regarding the generalizability of results from a sample to a larger population.² Secondly, as they include all teachers and any observed attrition indicates a teacher leaving the Michigan teaching workforce, rather than leaving the study sample, it is possible to obtain more accurate information about the movement of teachers between schools and districts within the state, although information on teachers moving out of state is still be difficult to track. Finally, Michigan, like many states, has invested significant resources in the construction, maintenance, and improvement of its longitudinal data systems, and is vitally interested in using these data more frequently to assess questions related to student achievement and statewide policy. Michigan's recent Race to the Top Round 2 application focused heavily not only on making data available for use by researchers, but on making data available for use to truly drive instruction, target professional development, and inform policy (Michigan's Race to the Top application, 2010). For this reason, these data represent a timely and critical resource to be utilized for academic research, in order to inform the state while also supporting their efforts in this area. Given the national interest in using state administrative data to address questions in educational policy, this study is also an important contributor to the national discussion of these issues.

Foundations for the Dissertation: Is there an adequate supply of teachers to meet the Michigan Merit Curriculum?

² Although the data are universe data in one sense, they still represent a sample, albeit a very well-defined one. These data can be considered one of the *possible* populations of teachers, students, and schools that could represent Michigan. One would expect the sampling error to be very small to negligible, but from a theoretical point of view, it is generally understood that one never has a *true* population, even when in possession of very complete data.

In 2006, Michigan instituted the Michigan Merit Curriculum, which required four years of math and English, three years of science and social studies, and two years of world language for high school graduation (Michigan Department of Education, 2006). The state needed a feasible, easily utilized formula that enabled them to estimate where the demand might be located, in order to inform both policy and to target resources. This method, designed to help the state make an administrative decision regarding whether or not it has enough teachers to accommodate the increase in graduation requirements, needed to meet three requirements; it needed to be 1) easily manipulated, 2) able to be completed in a short period of time, and 3) able to utilize existing data. The following method fits these three characteristics.

To estimate the number of teachers needed to fill the FTEs prescribed by the demands of the Michigan Merit Curriculum, I developed the following formula³:

$$D_i = \frac{a(x_i)/y}{z} \quad (1)$$

where

- D_i = number of teachers needed to meet graduation requirements in a subject area
- a = proportion of the student body that needs to be enrolled in each subject each year in order to meet graduation requirements
- x_i = total student enrollment in each school (including special education students)
- y = class size (assumed to be 25)
- z = number of periods taught per FTE per day (assumed to be 5)⁴

³ See Keesler, Wyse & Jones (2008b) on the IES website (http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/techbrief/tr_00508.pdf). This formula has been vetted by IES and is considered as a promising tool for use by other states. While the report was a collaborative effort, the development of the formula was an individual contribution on my part. This was also part of a technical report under the REL-Midwest Collaboration (Keesler, Wyse, & Jones, 2008a).

⁴ The formula and related assumptions will be discussed in greater detail in Chapter 4, Estimating Teacher Demand.

A key benefit of this demand formula is that it can be adjusted by the practitioner to meet specific conditions of a state, district or school. Class size and number of FTEs per teacher can be increased or decreased, enrollment can be adjusted, and specific curricular requirements of the school (outside of what is required by the state) can be factored in. One of the main objectives of the dissertation is to produce a formula that is rigorous but that is easily used by practitioners, allowing them to be more strategic and cost-efficient with resources, and to use available data in a real-time, meaningful manner.

Calculating Supply

To calculate supply, it is necessary to have access to data regarding teacher and building assignment; these are compliance data, i.e., information that each state is required to collect in order to receive federal funding. Supply is calculated by summing FTEs in each subject assignment within each school. Subject assignments are recorded by FTEs, or the proportion of each full-time employee (FTE) assigned to a particular task. For example, a teacher may teach math for half of the day and physics for the other half. In that case, her FTE would be reported as 0.5 math, 0.5 science. Similarly, a teacher may have both a teaching and administrative assignment. Therefore, it is important to calculate supply not by the number of “teachers” but by the total FTE within each subject area within each school.

Calculating Undersupply

Once demand is estimated for each school and supply is calculated, then one can subtract supply from demand. Positive numbers indicate an undersupply of FTE in a given subject area. The researcher should then designate a cut point at which undersupply is determined to be significant—where this “lack” is likely affecting practice, and not

able to be easily covered by adjusting teaching assignments or reconfiguring classes or other administrative tactics. Here, a cut point of greater than one FTE of undersupply is designated as significant undersupply. In the Michigan study, undersupply was calculated on a subject-specific basis, in order to take into account variation on availability of qualified teachers in different subjects. This method of estimating supply and demand is known as a behavioral model which can be used to answers questions related to the effects of policy changes by linking demand and supply estimates to relevant conditions and policies (Boe & Gilford, 1992).

In Michigan in 2007, there were 223 high schools undersupplied in math, 64 in English Language Arts, 41 in science, 39 in social studies, 52 in both mathematics and English Language Arts, and 9 undersupplied in all four core areas (Keesler, Wyse, & Jones, 2008a, 2008b). Undersupply seems to be more prevalent in mathematics and ELA. This undersupply is most likely a consequence of MMC's new requirements that students take four years of mathematics and ELA.⁵ There appears to be an association between schools that are undersupplied and failing AYP. While many schools that are undersupplied are able to meet their AYP requirements, there appears to be a relationship between school undersupply and failing AYP. Failing to meet AYP targets could be more a function of the demographic profile of the schools than whether a school is undersupplied or not. We cannot determine from the analyses presented here whether meeting AYP is a function of the demographics of the school, issues of undersupply, or both.

Finally, although the number of schools that are undersupplied is relatively small,

⁵ One of the reasons we suspect that many of the schools have an undersupply in mathematics and ELA is that, according to multiple school websites, graduation requirements prior to the MMC often required students to take 2-3 years of mathematics and 2-3 years in ELA.

the number of students affected by undersupply is not insignificant. For example, 72,798 students attend the 61 schools that were undersupplied in mathematics and ELA in 2007.

Investigating Teacher Demand, Supply and Turnover, and Undersupply Using State Administrative Data: Dissertation Outline

The method outlined above for calculating teacher demand, supply, and undersupply forms the foundation for this dissertation. This method is unique in several ways. First, it utilizes state administrative data which, as described above, offers unique advantages over nationally representative data sets, including the full specification of an entire teaching population. Secondly, supply and demand calculations are performed on a school-by-school basis. Given the local nature of both demand and teacher supply, calculating these factors at the school level gives increased precision and allows for a fine-grained analysis of the issue. Again, much of the work on teacher supply and demand produces estimates at the state or national level, which is far too aggregated to show the intricacies of supply and demand on the local level (for example, the most recent NCES publication using the SASS Teacher Follow Up Survey, Marvel, Lyter, Peltola, Strizek, Morton, & Rowland, 2007). Following Ingersoll, I view issues of supply, demand, undersupply and turnover as organizational characteristics of schools, which suggests that there are school-level differences and interactions between the organization itself and these factors which are critically important to understanding their import. Finally, this method is administratively useful as well as policy-relevant in that the assumptions can be changed by a state or district to more accurately reflect local condition, and this analysis can be run by non-technical staff who are interested in generating supply and demand calculations for their district or school.

This dissertation generally follows a three-essay model, although focusing more on developing a coherent overall empirical narrative as opposed to three standalone essays. As each analysis is necessary for consequent analyses, they are three critical components of the overall story of school-level teacher labor supply and its relationship to student achievement in the state of Michigan.

The next chapter presents, in greater detail, the data and analytic methods utilized in the dissertation. Chapter 3 provides descriptive analyses of the teaching workforce in Michigan at the individual level; the relationship between school-level teacher retention rates and school characteristics; and finally, the relationship between school-level undersupply and school characteristics. Chapter 4 introduces the demand formula and presents a series of modifications, and then tests the original formula against the modified formula. Chapter 5 undertakes a multilevel analysis of the impact of school-level teacher retention on student achievement and student mobility, and Chapter 6 extends this analysis into a quasi-experimental framework to estimate the effect of school-level teacher retention as a treatment. Utilizing the findings from the demand formula re-estimation (Chapter 4) and integrating the school-level teacher retention rate (Chapter 5), Chapter 7 estimates the impact of school-level teacher undersupply on student achievement and student mobility. The dissertation concludes with discussion and conclusions, as well as suggested policy impacts and outlined next steps (Chapter 8).

CHAPTER 2: DATA, METHODS, AND MEASURES

This chapter presents an overview of the data and sample, methods, and variables used throughout this dissertation. The majority of the detail regarding these elements will be presented here and referenced in their appropriate chapters.

Data

This dissertation utilizes data from several Michigan administrative data sources:

(1) The Michigan Registry of Educational Personnel (REP), (2) the Michigan Single Record Student Database (SRSD), (3) Michigan assessment data, and (4) Michigan teacher data. Each of these sources is discussed below.¹

These data represent an important, relatively new source of data for education research. Federal and state governments have spent vast sums of money developing and implementing these systems, and ensuring that they contain quality data. It is incumbent upon the research community begin to make use of these data, both in partnership with the states themselves, as well as in individual research work. These data are markedly different from other large-scale data sets. As they contain information on every student, teacher and school in the state, they are *universe* data.² While they contain an immense number of observations, they can be sparse in terms of number and types variables. A researcher can be creative, as was done here, and create variables as well as link with other data sets, but unlike studies such as those produced by NCES, these data do not contain multiple measures of things like social issues, expectations and plans, or family

¹ All data were obtained from Michigan's Center for Educational Performance and Information, and from the Office of Educational Assessment and Accountability in the Michigan Department of Education. These data are available to researchers who submit appropriate applications, research plans, and undergo human subjects review by a state review board.

² Although the data are universe data in one sense, they still represent a sample, albeit a very well-defined one. See footnote #2 in Chapter 1 for a discussion of this issue.

background measures. Missing data is a different issue as well. In a nationally representative survey, missing data can be very critical in terms of who is missing and to what extent that biases the findings. In these data sets, there is very little to no missing data, as many of these fields are required collections and missing data is not allowed.

When data is identified as “missing,” it is necessary to revisit that variable and understand whether a missing value truly means “no response” or whether it is a variable that would not be expected to have a value for that given student, teacher or school.

Finally, it is important to note that, although this work specifically uses Michigan administrative data, both the findings and the lessons learned from working with these types of state data are expected to be applicable in other state contexts.

The Registry of Educational Personnel (REP)

The Registry of Educational Personnel is Michigan’s longitudinal administrative database that contains information on every educational personnel member in the state of Michigan. This dataset includes demographic, teaching assignment, and licensure and endorsement information for all teachers. For this study, teacher data from 2005-2008 is utilized. It is possible to use these data to define the population of teachers in a given year, and understand their demographic and school characteristics; study the relationship between licenses, endorsements, and subjects taught, in order to test for out-of-field teaching; and to look at the distribution of types of licenses across types of teachers and also types of schools.³

The demographic information on teachers includes the following variables: age, gender, race, highest education degree, major and minor, higher education institution

³ The REP is stored in a relational file database, and thus creating research files for analysis requires “unstacking” the files to the appropriate level. The technical reports produced by the REL-Midwest provide further detail on the mechanics of this process.

attended and whether or not a teacher is “highly qualified.”^{4,5} Teaching assignment data provides information on each assignment held by a given teacher, including the subject, number of FTEs taught for that assignment, and number of classes taught by each teacher. Finally, the license and endorsement portion of the database provides information on the type of license a teacher holds (provisional, professional, vocational, administrative, etc.), the date that license was issued, and information on which specific endorsements the teacher holds.⁶ There are over 2,000 types of endorsements, covering every subject area, and teachers can hold up to seven endorsements, although most hold between one and three (Lynn & Keesler, 2008). This database includes a field for teacher preparation institution, but it is missing a relatively large amount of data (approximately 40%) because older teachers did not submit the information when they entered the system and the field only reports the most recent degree. There is also a highly qualified field, although 98% of teachers are considered and reported as highly qualified, so this is not necessarily a highly informative field. There is extensive information related to assignments—subject taught, number of FTEs in various subject.

The Single Record Student Database (SRSD)⁷

The SRSD is Michigan’s student database, which contains information on student demographics, such as gender, date of birth, racial/ethnic identification, and personally

⁴ Not all of the fields are “compliance fields,” which means that districts are not required to submit information for them. Age, race, gender, and highest education are all compliance fields, and are missing less than 2% of the information, but noncompliance codes like major/minor have approximately 40% missing. These variables should be used with caution, understanding their limitations.

⁵ “Highly qualified indicates that a teacher has at least a bachelor’s degree and has passed the Michigan teacher certification exams in both basic skills and in their subject area taught, or has demonstrated subject-area specific knowledge and training, such as graduate coursework in that subject or National Board Certification (Watkins, 2003).

⁶ This is known as the License 2000, or L2K, database, which is now fully contained in the REP.

⁷ Beginning in 2010, the SRSD will be replaced with the Michigan Student Data System (MSDS), with appropriate crosswalks between the two systems.

identifying information. Michigan is able to link all student demographic and assessment data using a unique identifier code (UIC), which enables longitudinal analyses of the data. The SRSD is collected annually by all intermediate school districts and submitted to the state of Michigan in order to comply with federal reporting requirements.

Student Assessment Data

Michigan, like most states, has a statewide assessment system, known as the MEAS (Michigan Educational Assessment System). This includes yearly testing in reading and mathematics in grades 3-8, and one grade in high school, grade 11, as well as alternative assessments for students with moderate to severe cognitive impairments, second-language learners, and different assessments for elementary and middle schools, and high schools.⁸ The high school test is the Michigan Merit Examination, a three-day test administered to all juniors that includes mathematics, English language arts, science, social studies and reading, and also includes the intact ACT and WorkKeys tests.

When measuring the relationship between the main predictors of interest in this study (school-level teacher retention and undersupply), and student achievement and mobility outcomes, it is desirable to have a pretest measure of student achievement, as this is an important covariate in regression analyses. In fact, Cook, Shadish and Wong (2008) find that “we can trust estimates from observational studies that match intact treatment and comparison groups on at least pretest measures of outcome.” For an observational study such as this one, including a pretest measure of student ability is critical to minimize the chance that the inferences are due to bias.

⁸ In compliance with NCLB, Michigan assesses mathematics and reading yearly grades 3-8 and grade 11. Other subjects must be assessed once per level (elementary, middle, and high school). Writing is assessed in grades 4 and 7. Science is assessed in 5th, 8th, and 11th grade. Social studies is assessed in grades 5, 6, and 9.

However, in Michigan, the state's assessments are not vertically scaled. As a consequence, the score that a student receives on their 8th grade achievement test is not on the same scale as the 11th grade achievement test, and therefore, methods such as gain scores or regression-based growth models are not well-supported by the data. When using a pre-test score as a predictor, however, it is possible to use the assessment data in this format, because although the scores are not linked to the same scale, they can serve as a reasonable prior measure of ability as an independent variable in regression-based analytic techniques. Therefore, for this analysis, student test scores on one administration of the Michigan Merit Examination are used—the Spring 2009 administration, and a pre-test score from their last assessment data—8th grade MEAP—is also used.⁹

School Level Data

While the teacher data file does not include school information, it is possible to link rich school-level data in using the Michigan school code. This school code has a corresponding NCES identification number, which allows for the linking of data from the Common Core of Data and other NCES sources. For this study, the school data file is constructed using the state of Michigan's Educational Entity Master, the overall master file that defines every educational entity in the state of Michigan and its current status (i.e. open, closed). State-released files on free and reduced lunch, pupil headcount, AYP

⁹ A better outcome measure for tracking student achievement growth is a specially developed measure by the state of Michigan which assigns each student one of five performance level change indicators with respect to their achievement scores from year to year—maintaining, increasing, significantly increasing, decreasing, or significantly decreasing. However, these data are not available for high school students, because of concerns with constructing this measure using the high school test (the MME) and the last test prior to high school (8th grade MEAP). These tests are administered three years apart from each other, and the scales and constructs are extremely different. Middle and elementary school content is determined by Grade Level Content Expectations (GLCEs), while high school content is determined via the Michigan Merit Curriculum. Therefore, calculating the performance level change categories seems inadvisable.

and school accreditation grade are linked in, as is information from the Common Core of Data. Only high schools are included in the school file.

Sample

Although data are available on all teachers, students, and schools, the sample for this dissertation is restricted to high schools only. The Michigan Merit Curriculum is a set of *high school* graduation requirements, and therefore, analyzing teacher supply, demand and undersupply in relation to these requirements requires analyzing high schools only. Only high schools that are local education agency (LEA) schools or public school academy (PSA) schools are analyzed—on other words, standard public high schools and charter high schools. Of those schools, I only analyze those defined as “regular” schools, as opposed to alternative or other types of schools, like special education centers. This is due to concerns about generalizability, in that alternative schools consist of unique configurations of students and teachers, and are fundamentally different than “regular” schools.¹⁰ If schools have a retention rate of zero, they are retained in the analysis only if they are not closed schools, with the exception of a handful of schools that were truly closed for restructuring one year and reopened the following year. Schools with less than three teachers are excluded from the analysis as well, as calculating a retention rate with less than three teachers leads to skewed results. This yields a sample of 580 high schools used in these analyses.¹¹

¹⁰ In a separate analysis of teacher retention in alternative high schools only, I find that higher levels of teacher retention rates are actually associated with lower student achievement, although this effect is not statistically significant than zero.

¹¹ The sample for both the school-level teacher retention and the school-level undersupply analyses is defined by the decisions made related to retention rates described above. It was necessary to analyze the same population of schools for both the retention and undersupply analyses, as the relationship between retention and undersupply was investigated. Additionally, the restrictions placed on the sample from the retention analyses served primarily to restrict the high schools to traditional high schools for whom this

Teachers were included in the population of instructional staff for a given year based on whether or not they had an “instructional” subject assignment, as defined by the state of Michigan. All instructional assignment records were tagged, and these were then flattened into a person-level file for each year. The population of teachers was 111,974 in 2005; 111,055 in 2006; 109,915 in 2007, and 107,793 in 2008.

For the student population, only students with both a valid, non-zero MME and MEAP assessment score were retained, and those who either did not have a matching pre-test score, had a score of zero, or took MI-Access (Michigan’s alternative assessment for students with moderate to severe cognitive difficulties), were eliminated from the analysis. This yields a sample of 96,556 students.

Methods: Multilevel Modeling and Propensity Score Matching

This dissertation focuses on two important relationships: (1) estimating the relationship between school-level teacher retention and student achievement and mobility outcomes, and (2) estimating the relationship between school-level teacher undersupply to meet the Michigan Merit Curriculum and student achievement, taking into account the school-level “churn” rate. In order to investigate these relationships, two types of methodological strategies are utilized: multilevel models and propensity score matching. Multilevel modeling, particularly with longitudinal data, allows for the estimation of the impact of school- and student-level characteristics on given outcomes. Multilevel models are used in tandem with propensity score matching techniques in order to provide additional evidence regarding the relationship observed, and to test for a relationship in a quasi-experimental framework.

demand formula is most neatly applied. However, it is possible to re-define the sample and conduct undersupply analyses on a different sub-population of high schools.

Below, I present an overview of each strategy, including a brief discussion of why that strategy was selected. The methods are then referenced in their appropriate chapters. The analysis of teacher retention uses multilevel modeling with sensitivity analyses, and propensity score matching with sensitivity analyses, to examine the relationship between school-level teacher retention, student mobility and student achievement, while the analysis of school-level undersupply uses multilevel modeling with sensitivity analyses to investigate the relationship between school-level subject-specific undersupply and student achievement.

The Multilevel Models to Be Estimated

To test the effects of the two main predictors of interest, school-level teacher retention and undersupply, on key outcomes, multilevel models (i.e., models with random effects) with high school students nested within schools are estimated. Unlike fixed-effects models, multilevel models do not simply control for effects of organizational contexts (e.g., schools), because they include the capacity to simultaneously test effects at multiple levels (e.g., the effects of student prior ability at the individual level and of teacher retention rate at the school level), as well as cross-level interactions to find, for example, an interaction between a student's race or gender and the teacher retention rate of the school (see Raudenbush & Bryk 2002).

Two types of multilevel models are estimated. The first are hierarchical linear models (HLMs), which test the relationship between the key predictors (teacher retention and teacher undersupply) and the continuous student achievement outcomes, and the second are hierarchical generalized linear models (HGLMs), used only in the teacher retention analysis in order to predict student mobility, which is a dichotomous outcome,

as a function of teacher retention. The general structure of both types of models (specified below) is a means-as-outcomes model, where the grand mean for each school is predicted by both student- and school-level characteristics. The theoretical framework of this study suggests that although the main predictors of interest, teacher retention and teacher undersupply, are school-level variables, these predictors do not act separately from student-level characteristics and that student performance is a function of both student- and school-level characteristics. It is well known that much of the impact on student achievement is related to individual student background characteristics, but a growing body of literature suggests that school characteristics have an impact on student performance as well (Rumberger & Thomas, 2000).

In the teacher retention multilevel models, teacher retention predicts student achievement both directly and as a crosslevel interaction with student mobility. The purpose of these models is to test whether or not teacher retention has a direct effect on student achievement, as well as to test whether student mobility interacts with teacher retention in this relationship.

The general model specification for the HLM models is as follows:

Level 1 model: (1)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_j Q' + u_{0j}$$

where

Y_{ij} = outcome (mathematics scale score) for each student i in school j

β_{0j} = each school mean, represented as a function of the grand mean, student mobility, the matrix of student-level predictors, the school teacher retention rate, and the matrix of school level predictors

β_{1j} = coefficient for student mobility

β_j = vector of coeffs for school j
 Z' = vector of student covariates for school j
 γ_{00} = grand mean (intercept)
 γ_{01} = effect of teacher retention on B_{0j} (each school mean)
 γ_j = vector of school-level predictors)
 Q' = vector of school covariates
 u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$
 r_{ij} = level 1 variance (student error term), r_{ij} distributed iid $N(0, \sigma^2)$

In the final teacher retention model, a crosslevel interaction term is added. The specification for this model is the same as above, with the student mobility slope predicted by teacher retention rate. The slope is not allowed to vary (i.e. is fixed).

Level 1 model: (2)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_j Q' + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{teacher retention})$$

This model is modified slightly in the teacher undersupply analyses, as outlined below:

Level 1 model: (3)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{undersupply}) + \gamma_j Q' + u_{0j}$$

where

Y_{ij} = outcome (mathematics, ELA or science scale score) for each student i in school j

β_{0j} = each school mean, represented as a function of the grand mean, student mobility, the matrix of student-level predictors, the school teacher undersupply rate, and the matrix of school level predictors

β_{1j} = coefficient for student mobility

β_j = vector of coeffs for school j

Z' = vector of student covariates for school j

γ_{00} = grand mean (intercept)

γ_{01} = effect of subject-specific undersupply on B_{0j} (each school mean)

γ_j = vector of school-level predictors

Q' = vector of school covariates

u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$

r_{ij} = level 1 variance (student error term), r_{ij} distributed iid $N(0, \sigma^2)$

For the HGLMs in the teacher retention analysis, student mobility is a binary outcome variable; therefore the use of a standard level 1 multilevel model is inappropriate (Raudenbush & Bryk, 2002). Student mobility is indicated by whether or not a student remained in the same school from the fall of 2006, which was their freshmen year, until the spring of 2009, when they took the high school achievement test (1=changed schools).

The general structure of the HGLM models follows below.

Level 1 structural model:¹²

(4)

$$\eta_{ij} = \beta_{0j} + \beta_j Z'$$

Level 2 model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_{02}(\text{student mobility 2007 cohort}) + \gamma_j Q' + u_{0j}$$

$$\beta_{pj} = \gamma_{po} \text{ for } p > 0$$

η_{ij} = the log odds of remaining in the same school for each student i in school j

β_{0j} = each school mean, represented as a function of the grand mean, the matrix of student-level predictors, the school teacher retention rate, and the matrix of school level predictors

β_j = vector of coeffs for school j

¹² The level 1 model in HGLM consists of three parts: a sampling model, a link function, and a structural model. The sampling model assumes that Y_{ij} , given the predicted value μ_{ij} , is distributed NID (μ_{ij}, σ^2).

The level-1 predicted value, μ_{ij} , can be transformed so that the predictions remain within the given

interval, which produces the transformed predicted value η_{ij} . This transformed predicted value is now related to the predictors of the model through the linear structure model. Combining the sampling model, link function, and level 1 structural model reproduces the familiar level-1 HLM model (Raudenbush & Bryk, 2002). The level 1 variance is now heteroskedastic

\mathbf{Z}' = vector of student covariates for school j
 γ_{00} = grand mean (intercept)
 γ_{01} = effect of teacher retention on B_{0j} (each school mean)
 γ_{02} = effect of 2007 cohort student mobility on B_{0j} (each school mean)
 γ_j = vector of school-level predictors)
 \mathbf{Q}' = vector of school covariates
 u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$

Analytic Strategy for Multilevel Models

A baseline for all multilevel models is established by estimating an unconditional random effects ANOVA. This allows for the calculation of the intraclass correlation, or the proportion of variance that is between schools. Following the estimation of this baseline model, a series of multilevel models are estimated. For the teacher retention analysis, four different multilevel models are estimated (see Table 5.2). The first, a bivariate model, estimates the bivariate relationship between school-level teacher retention and student achievement in mathematics. The second (Model 2) includes a second school-level predictor shown in the preliminary analyses to be highly correlated with the student achievement, percent free lunch, in order to control for other key school-level factors. This model also introduces a pretest measure of mathematics achievement at the student level, in order to account for student prior ability. The final two models are fully specified models with all predictors at level 1 and level 2, with the final model containing a crosslevel interaction between teacher retention and students who remain in the same school, to test the hypothesis that there is a multiplicative effect of teacher retention and student retention on student achievement.¹³

¹³ All Level 1 predictors, with the exception of pretest score and same school, are grand mean centered, in order to control for their effect, rather than partial out the impact attributable to student and school. A model with all Level 1 predictors group mean centered was run (output not reported); the effects for gender

A similar modeling scheme is utilized in the models predicting student mobility, beginning with a bivariate model with teacher retention predicting student mobility (see Table 5.3). The second model includes only the school-level measure of prior student mobility for the 2007 cohort. The final model is a full model, including all level 1 predictors (gender, race and program eligibility) as well as level-2 predictors.

In the teacher undersupply analysis, a very similar set of models were estimated as described above. Following the estimation of a baseline model, a series of multilevel models were estimated, for mathematics, English language arts, and science. Four different multilevel models are presented in Table 7.7 (mathematics), Table 7.8 (English language arts), and Table 7.9 (science). The first, a bivariate model, estimates the bivariate relationship between subject-specific school-level teacher undersupply and student achievement in that subject. The second (Model 2) includes a second predictor shown in the preliminary analyses to be highly correlated with the student achievement, free lunch eligibility, as well as a pretest measure of mathematics achievement and the student mobility indicator at the student level, in order to account for student prior ability. The third model adds in the school-level teacher retention rate. The final model includes all predictors at level 1 and level 2.

Sensitivity Analyses

When approaching applied analytic work using observational data, there is a concern regarding the impact of an unobservable characteristic on the outcome, one that might invalidate the inferences drawn from the study. When using state administrative data, this is a concern as well, as state data is rich in observations but does not include a

and race are largely at the student level, not at the school-mean level. Therefore, the decision was made to grand mean center race and gender in the reported models.

large number of variables. Sensitivity analyses are conducted to test the robustness of the inferences to the influence of other unobserved characteristics. Observational studies all vary in the degree to which they are sensitive to hidden bias, where some are very sensitive and others are relatively insensitive, even to substantial bias (Rosenbaum, 2005).

The presence of unobserved confounding variables cannot be “solved” but the sensitivity of the findings to this potential bias can be quantified, as a way to understand and contextualize the rigor of the results. Therefore I will characterize the robustness of these inferences to the potential impact of confounding variables (Frank, 2000). Following Frank (2000) and Frank & Sykes, et al. (2008), the sensitivity analyses techniques used in this paper consider how unknown quantities could affect estimates. However, as opposed to reporting how violations of assumptions produce a range of estimates, the focus remains on the extent to which an assumption must be violated to invalidate an inference. As a result, the indices reported here seek to quantify the robustness of the original inference (Frank & Sykes et al, 2008; Frank & Min, 2007; Frank, 2000).¹⁴

Propensity Scores: Brief discussion of propensity scores as an analytic technique

Propensity scores matching is one of a set of relatively new analytic techniques that have arisen out of the need to estimate treatment effects from observational data. Built upon the work of statisticians, such as Rosenbaum and Rubin (1983), and econometricians, such as Heckman, propensity score analysis allows for the estimation of

¹⁴ For more information regarding the development of these indices and their application to this study, see the Technical Appendix D. For more information on the these indices and their development and application across a variety of contexts, see Frank (2000), Frank & Min (2007), and Frank & Sykes, et al (2008).

causal effects from non-experimental data. Because of their utility in isolating effects when randomized studies are not possible (as is often the case), these methods are becoming part of the analytic methods of a number of disciplines, including education (Guo & Fraser, 2010).

To provide a brief overview of propensity score matching as a method, the goal is to approximate the counterfactual: how might the outcome have varied if a school had or had not received the treatment? Since the counterfactual can never be observed in one individual unit, propensity score matching finds similar units with different treatment assignments conditions, and compares their outcomes. One way to do this is to match the observations on a set of variables. However, this quickly becomes difficult due to the high dimensionality (Guo & Fraser, 2010; Rosenbaum, 2002; Rosenbaum & Rubin, 1983). Instead, a propensity score can be used as a balancing measure that summarizes the information in the covariates, and represents the estimated probability of receiving the treatment as a function of variables that predict treatment assignment (Morgan & Harding, 2006; Rosenbaum, 2002; Rosenbaum & Rubin, 1983). In this case, the treatment is considered high school-level teacher retention. Propensity scores are one of the main methods in matching strategies for estimating the counterfactual. The inherent difficulty with the counterfactual is that it can never be observed on one person or unit—a person/unit cannot receive both the treatment and the control under the same conditions. Therefore, the counterfactual is usually estimated in the aggregate.

Propensity scores are often presented as an improvement over ordinary multiple regression, as a technique that is quasi-experimental in nature and that can account for possible selection effects. Where OLS regression makes no claims to approximating an

experiment, propensity scores do, and are often described as a way to do an experiment when true random assignment is impossible, as is so often the case in social science and education research (i.e. when it is unethical, impractical, and/or impossible to randomize people into treatment and control conditions).

While propensity score matching is presented in an entirely different framework and so conceptually is more aligned with the idea of experimental design, in reality, it is not different in a statistical sense than regular OLS with predictors. One reason for this is that the same data is being used to estimate both types of analyses. The prediction equation in propensity score matching is using the same information from covariates that could be included in a regular multiple regression framework. That information is encoded in the propensity score, but still yields the same kind of estimates for individuals in the data set. In other words, Individual A's values on key covariates like SES, race, gender and family composition are the same in the prediction equation for the probability to receive treatment as they are in a multiple regression framework predicting the outcome.

What propensity score matching adds in applied contexts is an intuitive framework for isolating potential effects, as well as an efficient method for comparing treatment effects on groups that are matched on all available and observed covariates. While including all predictors in a multiple regression framework can yield similar estimates (for example, Frank, 2008), it does not allow the researcher to explain impacts in a clear treatment/control context. Presentation of findings in treatment/control contexts can be more intuitive to audiences, particularly when those audiences are policy makers or practitioners. A regression equation with a large number covariates can be confusing and

overwhelming; a comparison of means between treatment and control (where assignment is predicted by those covariates) can be more digestible. This class of methods focus on controlling for selection bias by using the observed covariates in ways to minimize this bias, and thus they offer advantages over OLS regression (Guo & Fraser, 2010). More importantly, when targeting analyses toward a policymaking audience, it can help isolate the potential effect of implementing a new policy—as is the goal here.

However, it is important not to overstate the “power” of propensity scores. They do not remove the issue of unobserved confounding variables (which is a key criticism of regular regression) because those unobserved confounding variables are still unobserved, and assignment into treatment and control conditions based on covariates is only as good as the covariates. While they can control for overt selection bias, they are not able to control for hidden selection bias (Guo & Fraser, 2010). For this reason, propensity score matching is used in this analysis in conjunction with sensitivity analyses that help quantify how robust the treatment effect is to hidden bias, as suggested by Rosenbaum (2002, 2005) and Rosenbaum and Rubin (1983).

Propensity Scores: Weighting Method

Propensity scores allow for the comparison of those in a “treatment” group (schools with high teacher retention) and those in a control group (schools with low teacher retention) who have similar propensities to receive the treatment. The treatment, high teacher retention, is defined as a school with over 85% of the teachers retained from year to year.¹⁵ The propensity score is the probability of a school receiving the treatment

¹⁵ Tests were done to estimate the best threshold for the defining the high teacher retention treatment level. Several thresholds were included in a series of logistic regressions to see which were significant. Additionally, propensity scores were estimated for several thresholds and the distribution of treatment and

(high teacher retention), given a set of covariates. For this analysis, a weighting approach to propensity scores is used, in order to allow all information to be retained in the analysis and not lost as can be the case when doing a case by case matching approach to propensity scores (Hirano & Imbens, 2001; Morgan & Harding, 2006; Robins, Hernan, & Brumback, 2000; Robins & Rotnitzky, 1995; Robins, Rotnitzky & Scharfstein, 2000).

The propensity score is calculated by estimating a logistic regression predicting treatment (i.e. high teacher retention).¹⁶ This yields a propensity score, which is the propensity of receiving treatment given the set of covariates, and is defined as follow (see Rosenbaum & Rubin, 1983):

$$e(x) \equiv \Pr\{t = 1 \mid x\} \quad (5)$$

where t = whether or not the school received the treatment of high teacher retention, and $\Pr\{t=1|x\}$ is the probability of receiving the treatment given that set of covariates.

After estimating the propensity score, several weights are calculated. The first is the weight for the treatment effect for those schools on the margin of indifference, or the effect of treatment at the margin of indifference (EOTM) (Heckman, 2005). In other words, this weight focuses on those schools that might be considered most likely to respond to a change in policy on teacher retention rates. Schools that had low teacher retention but had a high propensity for having high teacher retention might be responsive to shifts in statewide policy regarding professional development or incentives for teachers to remain in schools. This weight is calculated as follows:

control cases in each of the blocks was examined. An 85% threshold yielded the most reasonable results, and therefore is used here as the treatment.

¹⁶ Propensity scores are estimated using the `pscore` program in the Stata statistical analysis software, created by Sascha Becker and Andrea Ichino (2002).

$$\omega(t, x) = \frac{t}{e(x)} + \frac{1-t}{1-e(x)} \quad (6)$$

This weight is then included in the multilevel models specified previously, weighting the level-2 observations by their propensity to receive the treatment.¹⁷ This means that schools that have high retention are weighted by $\frac{1}{e(x)}$ and schools that have low retention are weighted by $\frac{1}{1-e(x)}$. Therefore, schools with high retention are weighted more when they have a lower propensity of having high retention, and low retention schools are weighted more when they have a high propensity of having high retention. This focuses attention on those schools that either received the treatment and had low propensity to do so; or those who did not receive the treatment despite having a high propensity to do so.

Weights were also constructed that allow attention to be focused on the treatment effect for the treated and the treatment effect for the control. Looking at the treatment effect for the treated asks, “How well did the treatment work for those who received it?” while looking at the treatment effect for the control asks, “How well might this treatment have worked for those in the control group?” These weights are constructed as follows:

Treatment effect for the treated:

$$\omega(t, x) = t + \frac{1-t}{1-e(x)} \quad (7)$$

Treatment effect for the control:

¹⁷ In order to preclude extreme observations from exerting undue influence on the models, weights were trimmed at 18.

$$\omega(t, x) = \frac{t}{e(x)} + 1 - t \quad (8)$$

Propensity Scores: Stratification Method

I also investigate this relationship by looking at the average treatment effect within four strata.¹⁸ This estimates the average treatment effect on the treated (ATT) using stratification matching. The strata were defined during the estimation of the propensity score. By construction, the covariates in each block are balanced and the assignment to treatment can be considered random. The ATT is computed only on the region of common support, and is computed using a weighted (by the number of treated) average of the block-specific treatment effects. In turn these are computed as the difference in average outcomes of treated and controls within the same block for which the all control variables are balanced (Becker & Ichino, 2002).¹⁹

Measures

Key predictors of interest: School-level teacher turnover and undersupply²⁰

These analyses use two important predictors constructed from the data: school-level teacher “churn rates,” and school-level teacher undersupply. As this is an organizational analysis of teacher supply and demand, characteristics that may be studied at the individual level (i.e. teacher retention) are considered as an organizational characteristic of the schools.

School-level teacher churn rates are calculated by identifying the number of teachers who remain in the workforce and in the same school from one year to the next.

¹⁸ This is computed using Stata’s `atts` command within the `pscore` program. See Becker & Ichino (2002).

¹⁹ For comparison, I estimate these effects with nearest neighbor, kernel, and radius matching as well. This is described further in Chapter 6 and in the Appendix C.

²⁰ A complete list of all variables used throughout the dissertation is included in Appendix A.

This is estimated over four years, which is three retention time points. This is then aggregated to a churn rate; the proportion of the teaching force in a given school that remains the same from one year to the next. This rate is averaged over the three time points, and a “difference” indicator is included in analyses to control for the direction of change in average retention rates.

School-level teacher undersupply is calculated using the method outlined above, with modifications. Demand is estimated using a modified demand formula (Chapter 4), then supply is calculated by summing the FTEs in each subject for each school, and undersupply is calculated by subtracting supply from demand. This variable is then used as a predictor in multilevel analyses predicting student achievement. It is also used in conjunction with the school-level teacher retention rates calculated above, in order to take into account the impact of a “revolving door” of teachers on the relationship between undersupply and outcomes.

School-level Covariates

The school-level predictors include those variables shown by research to be related to both teacher retention and student outcomes: the percent minority students in the school, percent free/reduced lunch students, locale (a dummy variable comparing city, town and rural schools to suburban schools), school size (a dummy variable comparing small schools of under 300 students and large schools of over 1000 students to medium schools), indicator variables for charter and magnet status.

Additionally, I use a set of workforce composition variables, including the proportion of teachers with professional licenses (by FTEs), proportion minority teachers (by FTEs), and proportion of highly qualified teachers (by FTEs).²¹

For the student mobility analysis, a school-level covariate for the 2007 cohort mobility is used. This variable indicates the proportion of students in each school who took the MME in the spring of 2007 and who were in the same school in the fall of 2005.²² It is used as a prior measure of school-level student mobility, in order to help establish a stronger causal ordering. This measure is prior to the final year of the teacher retention measure (2008), as well as prior to the student achievement outcome measures from 2009.

Student-Level Covariates

The following student predictors are utilized: mathematics or English language arts pretest score from the 2005-2006 8th grade achievement test, the MEAP test,²³ gender, race, free and reduced lunch eligibility, special program eligibility and an indicator variable for whether or not students have been in the same high school since 9th grade. These student level predictors are all well-known to have important relationships with student achievement outcomes, and are a standard set of student background covariates to include in such an analysis. Gender is a dichotomous variable (1=female). Race is categorized into American Indian/Pacific Islander, Asian, Hispanic, Black and

²¹ School means of all student predictors are also included on the intercept when those student level predictors are included in the model and are group-mean centered.

²² This variable indicates the proportion of 2007 juniors who were in the same school in the fall of 2005, i.e. as sophomores. Data was not available for the fall of 2004, when those students were freshmen.

²³ By legislative mandate, Michigan only administers a high school achievement test at one time point in high school—in 11th grade. Therefore, the only available pre-test measure of prior achievement is the 8th grade MEAP score. Only students with valid MEAP and MME scores of a non-zero value were retained in the analysis (n=96,556).

multi-ethnic, with white as the reference category. Free and reduced lunch eligibility is a dichotomous variable (1=free/reduced lunch eligible). The same school variable used as an outcome in the second set of models is used as a predictor in the student achievement models. Program eligibility indicates whether or not a student is eligible for Title 1, special education, Section 504 plan, Limited English Proficient, advanced/accelerated, or migrant services. This is recoded into a dichotomous variable that indicates whether or not a student is eligible for any special programs.²⁴

²⁴ The number of students in this file who are eligible for these programs is very small (n=1300). This is partly due to the fact that students who took MI-Access, Michigan's alternative assessment for students with moderate to severe cognitive difficulties, are not included in this analysis, and thus much of the special education and Section 504 population is not included.

CHAPTER 3: DESCRIPTIVE ANALYSES OF MICHIGAN’S TEACHER SUPPLY, TURNOVER, AND UNDERSUPPLY

Teacher supply, retention and undersupply are interrelated aspects of the teacher labor supply in the state of Michigan. The analyses in this chapter present descriptive evidence to highlight these relationships, and to provide an overall empirical picture of teacher supply, school-level teacher retention, and school-level undersupply. The descriptive analyses presented here allow for the investigation of three questions: 1) What is the composition of Michigan’s teaching force (i.e. its supply) in terms of demographics, licenses, and distribution? 2) What is the distribution of school-level teacher retention rates over various types of schools? 3) What is the distribution of school-level undersupply over various types of schools? This chapter begins with findings regarding the characteristics and composition of the teaching workforce in Michigan, focusing in particular on the distribution of teachers across various types of schools, license types, and subjects. The analysis then turns to Michigan’s overall labor supply, looking at the number of teacher entering, leaving, and staying in the profession, and then for those who stayed, the proportion who changed schools and those who did not. The final portion of the chapter focuses on the two key school-level characteristics that are of interest throughout this analysis—school-level teacher retention and school-level undersupply, in order to understand how these characteristics vary across types of schools.

While descriptive analyses cannot provide evidence regarding causal relationships, they provide critically important information regarding the population of teachers and schools under analysis. With these universe data, they are even more

informative than data from representative samples, as they represent the complete population of teachers and schools, and therefore observed differences reflect *real* differences in the education practice in Michigan.

Investigating Michigan’s Teacher Labor Supply: Previous Analyses

In order to understand teacher demand, supply and turnover, and undersupply, it is necessary to first understand the state of Michigan’s teaching force—its demographic composition, the distribution of licenses across schools, and issues related to out-of-field teaching and a “reserve” supply of teachers. This provides background information that is necessary to correctly situate the findings in the remainder of the dissertation in the appropriate context. To provide this information, I will rely on findings from the REL-Midwest collaboration between the Michigan Department of Education, the Center for Educational Performance and Information, and Michigan State University.^{1,2}

*What are the demographic characteristics of Michigan’s teaching force?*³

The REL-Midwest research team investigated the characteristics of Michigan’s teaching force, with particular attention paid to distributional issues as well as the relationship between teacher characteristics and placements and AYP (see Table 3.1). Overall, the state’s teaching workforce is primarily white (89%) and aging. The minority representation among the teaching force appears to increase with age, as younger teachers appear to be predominately white.

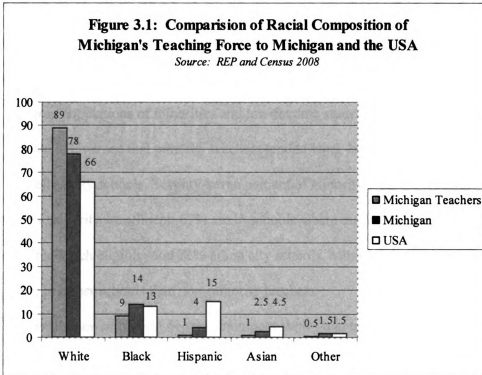
¹ The REL-Midwest collaboration has produced three technical reports. I served as a graduate research associate on this project from its inception and have co-authored (either a first or second author) all three reports. The work of this project, and the findings in these reports, form the foundation for my dissertation.

² Data from all of the reports mentioned here are from the 2007 REP, although analyses from the 2008 REP (the most recent teacher data available) do not show any substantial differences in distribution or findings.

³ Findings from Lynn & Keesler et. al (2007): Technical Report 1: Beyond Compliance: Descriptive Characteristics of Public School Teachers in Michigan.

The data show that the majority of minority teachers in Michigan are in urban schools, and that urban fringe, town, and rural schools employ teachers who are almost exclusively white.

Figure 3.1 below compares Michigan's teaching force to the racial composition of Michigan and of the United States. White teachers are overrepresented in the Michigan teaching force relative to the overall Michigan population (89% in the teaching force, compared with 78% in the population, while black and Hispanic teachers are underrepresented.



One important federal reporting requirement under No Child Left Behind, as well as a focus of the current administration's Race to the Top competition, is the equitable distribution of teachers across types of schools. Table 3.2 shows the distribution of full- and part-time teachers by teacher characteristics and across school types; Table 3.3 shows

the distribution of minority and non-minority teachers by teacher characteristics and across school types; and Table 3.4 shows the distribution of teachers by highest education by teacher characteristics and across school types. Equitable distribution questions often related to the distribution of teachers in terms of quality and effectiveness. With the current data, Michigan does not calculate teacher effectiveness based on student test scores and results from annual educator evaluations, although this activity will begin in the 2010-2011 school years. Until this new system is available, equitable distribution calculations focus on issues like race and education, not student achievement growth.

The distribution of teachers by school types indicates that there does not appear to be a pattern of part-time teaching in high poverty or high minority schools (see Table 3.2). In fact, there appear to be more part-time teachers in urban fringe schools, and schools with low populations of minorities and low-income students.

In Table 3.3, we see that minority teachers tend to work in minority and poor schools, as well as city schools. Seventy-seven percent of minority teachers are in schools with over 50% minority enrollment; 67% are in schools where more than half of the students are free-lunch eligible; and 72% are in city schools. Minority teachers are more likely to work in larger schools than non-minority teachers.

The distribution of highest degree attainment shows that teachers with doctoral and specialist degrees tend to be concentrated at the high school level and in schools in the urban fringe (see Table 3.4). Teachers with master's and bachelor's degrees are employed primarily in elementary and middle school. There do not appear to be noteworthy inequities in the distribution of teachers over schools by highest education degree.

Finally, the distribution of teachers over locale types suggests that teachers who work in cities are more likely to teach in schools where over half of the students receive free and reduced lunch (61%), and in schools with enrollments over 1000, than teachers in rural, urban fringe, or town schools (see Table 3.5). Teachers in rural schools are also more likely to teach in schools with higher numbers of low-income students (30% are in schools with over half of students eligible for free and reduced lunch); however they are also more likely to teach in small schools (under 300 students). Teachers in rural schools also look different with respect to their educational degree attainment; they are less likely to have an advanced degree and are more likely to have a bachelor's degree (60%).

Licensure and Endorsement

Michigan's teaching force is defined not only by their demographic characteristics and placements, but most importantly, by their training and skill, as represented here by their license levels.⁴ The most commonly held licenses are provisional, professional, and 18/30 Hour Continuing licenses. The provisional certificate, Michigan's initial teaching certificate, is issued following successful completion of an approved elementary or secondary teacher preparation program, including student teaching, and after a candidate has passed all components of the Michigan Test for Teacher Certification (MTTC). The certificate is valid for up to six years. During this time, the teacher is expected to complete a minimum of three years of successful teaching experience, and to finish at least eighteen semester hours in a planned course of study as a prerequisite for the next

⁴ As of this writing, Michigan does not collect information regarding teacher quality and effectiveness. License and endorsement status serve as proxies for these qualities, along with the "highly qualified" indicator. Michigan will begin annual educator evaluations in 2011 and the results of these evaluations will become part of the REP. They will also begin to calculate growth for all teachers when the data are available and applicable, and make these measures available as well.

level of certification.⁵ The professional education certificate is Michigan's advanced teaching certificate, and requires completion of 18 semester hours in a planned course of study after the issuance of an approved initial teaching certificate (or an approved master's degree earned at any time), and three years of successful teaching experience. This certificate is valid for up to five years; it must be renewed every five years by completing six semester hours at an approved teacher preparation institution or a state board-approved institution.

The 30 or 18 hour continuing licenses are no longer issued, although teachers with these certificates continue to provide instructional services. They remains valid as long as the holder continues to serve in an educational capacity for 100 days in any given five-year period.⁶

Table 3.6 presents the distribution of licenses by teacher demographics in the state of Michigan. Age has a predictable relationship with license status, with the two age groups most likely to have provisional licenses being those under 25 and those between the ages of 25 and 29. The majority of teachers (76%) who are between the ages of 30 and 39 have professional licenses. Among those teachers who are 40-54 years of age, 47% hold professional and 40% hold 18/30 hour continuing licenses. In the oldest group (those aged 55 and above), the majority (62%) hold 18/30 hour continuing licenses.

5 (The provisional certificate can be renewed if all of the requirements for the Professional Education certificate have not been met.)

⁶ Michigan also issues substitute licenses to those who stand-in for other teachers. The must have completed 90 semester hours of satisfactory (minimum 2.0 grade point average) credit consolidated at one four-year regionally accredited college and university. The substitute permit is valid for teaching a maximum of 150 days during the school year in day-to-day substitute teaching assignments. This permit is not valid for any regular or extended assignment. However, substitute teachers are not part of the instructional universe for this analysis, and therefore are not described in any detail here.

There appear to be no significant differences by gender with respect to the types of licenses held by males and females.

Of the 19,718 teachers with provisional licenses, 16,180 (82%) hold a bachelor's degree. Teachers with bachelor's degrees are most likely to have professional (41%) or provisional (37%) licenses. Teachers with master's degrees are more likely to have professional (58%) or 18/30 hour continuing (35%) licenses. Teachers holding doctoral or specialist degrees are nearly evenly split between those with 18/30 continuing (49%) and professional (43%) licenses.

The distribution of licenses by minority teacher status suggests that non-minority teachers are more likely to hold professional (51%) than 18/30 hour continuing (28%) or provisional (19%) licenses. Minority teachers, like non-minority teachers, are more likely to hold professional licenses (38%). However, more minority teachers hold provisional licenses (26%), compared with 19% of non-minority teachers. There does not appear to be a relationship between full-time status and license type.

*Teacher licensure and endorsement: Are teachers teaching out of field? Is there a reserve supply of teachers to meet the demands of the MMC?*⁷

One of the critical questions with respect to teacher licensure was out of field teaching. Teachers are generally not teaching out of field in the state of Michigan. Only 2% of the total high school teaching population is teaching out-of-field. This includes 123 in mathematics, 164 in English, 96 in science, 122 in social studies, and 64 world language teachers. The match between subject assignment and endorsement seems to be closest when teachers have only one subject assignment, but even with multiple

⁷ Findings from Lynn & Keesler (2008), *Technical Report 3.1: Beyond Compliance: Teacher Licensure and Endorsement in the State of Michigan*.

assignments, there are a very small number of mismatches at the high school level. It is often the case that out-of-field teachers have several assignments, one of which is out-of-field. It is relatively rare to find a teacher who is teaching full-time in a subject without a valid endorsement. Table 3.7 presents those teachers who are assigned to a core subject but do not hold at least one endorsement in that subject area.

Endorsements and Subject Assignments—Potential Reserve

A second question is the extent to which teachers who hold multiple endorsements could be moved from one assignment (such as social studies) to another in which there was greater need in order to meet the increased demands of the MMC. However, there does not appear to be a reserve of teachers with “unused” endorsements who could be reassigned within a school to meet subject area shortages. (See Table 3.8) In general, teachers with a math endorsement are the most likely to actually teach math (83%). Of the remaining 17%, half of those are teaching science. Math appears to be a “dominant” endorsement. Those who hold a math and another endorsement are more likely to actually teach math. There is a relationship between math and science endorsements. For those with science as one of their endorsements, 55% teach in science, and another 30% teach in math. These findings suggest that schools are not likely to have a large “reserve” of people who could teach math or world language. Schools may have a reserve of science teachers, although if reassigned they are likely to be pulled from math courses.

Michigan’s Overall Teacher Labor Supply

Many analyses of teacher labor supply focus at the state or national level on the number of teachers entering, leaving and remaining in the profession. While this analysis

moves to the school-level and studies labor supply at that level, it is nonetheless important to understand Michigan's overall teacher labor supply. Table 3.9 below presents the proportion of Michigan's teaching force that is retained in the profession, enters the profession, and leaves the profession each year. Additionally, for those who are retained, Table 3.9 shows the proportion of those teachers who moved to different schools ("movers") and those who stayed in their previous schools ("stayers").

In the overall teaching workforce, Michigan retains between 92% and 94% of its teachers. Note that the absolute size of the instructional workforce is declining over time, and in 2008, was three percent smaller than in 2006. Each year, a slightly smaller percentage of the overall population leaves the profession, while the size of the entering cohorts of teachers has steadily declined, down 7% in 2008 from the 2006 rate of entry.

Within those teachers who are retained in the profession, Michigan has a steady 11% inter-school mobility rate, which is the overall teacher "churn" rate for the state of Michigan at an aggregate level. As will be seen later in this dissertation, this churn rate is not equally distributed among schools.

School-Level Teacher Retention by School Characteristics

This dissertation conceptualizes teacher retention as a school-level characteristic. Retention rates may be distributed unevenly among schools. As described in Chapter 2, average school-level teacher retention rates were calculated for each high school in the state of Michigan. Table 3.10 presents mean school-level teacher retention rate by school characteristics, in order to understand the distribution of teacher retention rates across

schools.⁸ Does teacher retention rate vary by key school characteristics, such as locale or student poverty? Recall that teacher retention rate is an average over the period 2005-2008, so the mean teacher retention rate presented here is the mean of that average by school characteristics. The mean school-level teacher average retention rate for all high schools in the state of Michigan is 86.20%, which indicates that for all high schools in the state of Michigan, the average four year retention rate was 86.20%. Examining the distribution of mean school-level teacher retention rates over a variety of school characteristics, school teacher retention rates differ significantly by locale, with teacher retention rates in city schools at 79.73%, which is significantly lower than the rates in suburban, town and rural areas. Suburban, rural and town schools did not have significantly different teacher retention rates.

Average teacher retention rates also differ significantly by the composition of the student body, both in terms of percent minority students and percent of students who are eligible for free and reduced lunch. High minority schools have a teacher retention rate of 74.98%, which is significantly lower than other categories. Schools with over 50% of students eligible for free and reduced lunch have significantly lower teacher retention rates (80.27% for schools with 50-69% of students eligible for free and reduced lunch, and 76.09% for schools with over 70% of students eligible).

In terms of school size, there is again significant variation across sizes, with small schools having lower average teacher retention rates (81.81%) than medium (86.40%) and large schools (87.84%). Charter schools have significantly lower retention rates than non-charter schools, with charter schools having retention rates of 71.91%, compared

⁸ This table presents the mean undersupply in each subject by school characteristics using a series of oneway ANOVAs. The reported F-test indicates significant between-group variation on each characteristic.

with 87.06% for non charter schools. However, there are only a small number of these schools available to make this comparison. Magnet schools do not have significantly different teacher retention rates than non-magnet schools.

Schools that have a lower percentage of teachers with professional licenses also have significantly lower teacher retention rates. Schools with less than 80% of teachers with professional licenses have an average retention rate of 79.15%, compared with 87.69% for schools with 80-90% of teachers with professional licenses, and 89.55% for schools with greater than 90% of teachers with professional licenses. The percentage of minority teachers is also related to teacher retention rates. For schools with the highest percentages of minority teachers, the teacher retention rate was 76.51%, which was significantly lower than the teacher retention rates for schools with fewer minority teachers.⁹ Teacher retention rates are not significantly different across schools with different percentages of highly qualified teachers. To summarize, the types of schools that have lower mean teacher retention rates are city schools, schools with high proportions of minority and low-income students, small schools and charter schools. Schools with a lower percentage of teachers with professional licenses, and schools with a higher percentage of minority teachers also appear to have lower mean teacher retention rates.

School-Level Undersupply by School Characteristics

Tables 3.11, 3.12, and 3.13 present mean mathematics, English language arts, and science undersupply, respectively, by school characteristics, to develop a profile of the

⁹ This variable is highly skewed, as many schools have zero or less than one percent of their population with minority teachers. Michigan's instructional workforce is only 11% minority teachers. Therefore, this variable is categorized into less than 1%, 1-10% and greater than 10%.

type of schools that are undersupplied in each subject.¹⁰ The undersupply calculations range from approximately -5 to 5 (see Appendix A for descriptions of all variables used in the study), with positive numbers indicating undersupply.

Mathematics Undersupply

The mean school-level teacher undersupply is .48, which suggests that, on average, schools are not significantly undersupplied, although there is considerable variation in those rates (standard deviation=1.20). This adds further evidence to suggest the importance of a school-level organizational analysis, in order to capture those school-level differences and their relationship with student achievement outcomes.

Suburban schools had higher mean rates of mathematics teacher undersupply, with an average of .83, and a standard deviation of 1.37, while city schools had an average of .64 with a standard deviation of 1.68. This suggests that there were city and suburban schools with both much higher rates of undersupply, as well as much lower. Rural schools had the lowest average rate of mathematics undersupply, with the smallest standard deviation (.18 and .84, respectively). The differences among average mathematics undersupply between locales are statistically significant at the .001 level.

Schools with the lowest populations of minority students have the lowest average rate of mathematics teacher undersupply (.31). Interestingly, the distribution of average mathematics teacher undersupply rates among the school categories of free and reduced lunch is mixed; those schools with the lowest rates of free and reduced lunch have, by far, the lowest average rates of mathematics teacher undersupply (.04), but schools with 10-29% of their students free/reduced lunch eligible have average mathematics undersupply

¹⁰ This table presents the mean undersupply in each subject by school characteristics using a series of oneway ANOVAs. The reported F-test indicates significant between-group variation on each characteristic.

rates of .67, which is the highest of the five categories, followed by schools with 50-69% of their students free/reduced lunch eligible, at .62. However, schools with greater than 70% of their students eligible for free and reduced lunch have average rates of mathematics teacher undersupply of .29, which is the second lowest rate. This suggests that the relationship between mathematics teacher undersupply and school socioeconomic status is mixed, and may not be strongly and linearly related as one would expect.

Mathematics teacher undersupply varies widely among school sizes, with this between-group variation exceeding the threshold for statistical significance. Small schools (those with less than 300 students) have the lowest average rates of mathematics teacher undersupply, with an average of -0.15, which means that small schools, on average, are slightly oversupplied. Large schools (those with more than 1000 students) have an *average* mathematics teacher undersupply of 1.06, which suggests that on average, large schools are significantly undersupplied. Charter and magnet schools are both less likely to exhibit mathematics teacher undersupply than non-charter and non-magnet schools, which may be due in part to their smaller size, although these relationships are not statistically significant.

Finally, when looking at school compositional characteristics, there is not a clear relationship between the percent of the teachers in a school with professional licenses and mathematics teacher undersupply. There is a significant relationship between the proportion of minority teachers in a school and average mathematics teacher undersupply, with schools with higher rates of minority teachers having higher rates of average mathematics undersupply. Schools with less than 75% of teachers highly qualified also have higher average rates of mathematics teacher undersupply.

English Language Arts and Science Undersupply

Table 3.12 presents average English language arts teacher undersupply rates by school characteristics and Table 3.13 presents average science teacher undersupply rates by school characteristics. These tables are included, but are not discussed in detail for two reasons. The first is that there is not an average English language arts or science teacher undersupply; the average English language arts teacher undersupply is $-.78$ and the average science teacher undersupply is $-.77$, which suggests that schools, on average, have an adequate supply of ELA and science teachers. The caveat again is that there are important variations across schools, with some schools experiencing significant undersupply in these areas, but on the average, these are not undersupplied areas. This underscores the importance of investigating this not at the state level, but at a school-by-school level. The second reason these tables are not presented in detail is that the relationships mimic those outlined in the mathematics teacher undersupply. I will highlight key differences here.

Suburban schools have the highest average rates of English language arts teacher oversupply, with the average rate of teacher supply at -1.10 . Using a full FTE of over- or under-supply as a significant threshold as described in Chapter 4, this represents a potential significant oversupply.¹¹ Recall that suburban schools also had the highest average rates of mathematics teacher undersupply. This suggests that suburban schools find it easier to adequately staff their English language arts classes than mathematics classes. Like in mathematics, rural schools have the lowest rates of ELA and science

¹¹ Note that “oversupply” is used somewhat loosely; in this formula, a rate of 0 would indicate a school that was perfectly balanced in terms of supply and demand. Positive numbers indicate an undersupply; negative numbers indicate an “oversupply,” or a supply that is greater than estimated demand. This does not suggest that these schools should reduce their teaching force, but rather that, given the assumptions of the formula, they are not in danger of being undersupplied at this particular time.

oversupply, which suggest that rural schools most accurately match their teacher supply to their demand needs. These relationships are similar in science.

Unlike in mathematics, in both English language arts and science, the relationship between percent free and reduced lunch and undersupply is linear and positive. As the percent of students who are free and reduced lunch eligible increases, the average oversupply of English language arts and science teachers decreases. However, even in the schools with the highest rates of free and reduced lunch, the average undersupply is negative (i.e. is not undersupplied).

Average undersupply rates in English language arts and science are related to school size in a similar manner as in mathematics. Small schools have the smallest rates of oversupply, while large schools have the largest rates. This is the converse of the mathematics relationship, where small schools had the lowest undersupply and large schools had the highest average undersupply. Again, like with rural schools, this suggests that small schools are most able to match their supply and demand needs accurately, while large schools are more likely to be significantly over- or undersupplied.

The relationship between the school compositional variables (percent teachers with professional licenses, percent of teachers who are minority teachers, and percent of teachers who are highly qualified) follow the same pattern as mathematics, with schools with lower proportions of professional teachers, higher proportions of minority teachers, and lower proportions of highly qualified teachers have lower rates of oversupply (whereas in mathematics, these types of schools had higher rates of undersupply). Schools with these compositional characteristics are more likely to trend toward being undersupplied, even if the absolute rate is not undersupplied.

Table 3.1: Characteristics of Michigan public school teachers by full-time equivalency status

	a Full-time		b Part-time		Total	
	N	%	N	%	N	%
Gender						
Male	24,037	26	1,210	18	25,247	26
Female	67,813	74	5,574	82	73,387	74
Missing	35	<1	13	<1	48	<1
Age						
<25	1,425	2	197	3	1,622	2
25-29	11,881	13	821	12	12,702	13
30-39	26,893	29	2,448	36	29,341	30
40-54	32,659	36	2,137	31	34,796	35
>=55	18,992	21	1,181	17	20,173	20
Missing	35	<1	13	<1	48	<1
Race/ethnicity						
White	82,105	89	6,390	94	88,495	90
Black or African American	8,160	9	289	4	8,449	9
Hispanic or Latino	821	1	49	1	870	1
Asian American	472	1	31	<1	503	<1
American Indian or Alaska Native	213	<1	15	<1	228	<1
Native Hawaiian or Other Pacific Islander	55	<1	4	<1	59	<1
Multiple	24	<1	6	<1	30	<1
Missing	35	<1	13	<1	48	<1
Highest educational level						
Doctoral or Specialist's degree ^c	256	1	44	1	300	1
Master's degree	48,333	53	3,471	51	51,804	53
Bachelor's degree	40,422	44	2,966	43	43,388	44
High school	1,496	2	180	2	1,676	2
Other	761	1	98	1	859	1
Function Assignment						
Elementary school	34,279	38	2,348	34	36,627	38
Middle/Junior high school	15,094	16	777	12	15,871	16
High school	20,316	22	1,518	25	21,834	22
Special education	13,076	14	905	13	13,981	14
Career/technical education institutions	1,753	2	126	2	1,879	2
Other ^d	3,827	4	847	12	4,674	5
Multiple	3,540	4	276	4	3,816	4

Table 3.1 (cont'd)

	^a Full-time		^b Part-time		Total	
	N	%	N	%	N	%
Subject assignment						
English Language Arts	6,736	7	628	9	7,364	8
Social Sciences	4,004	4	264	4	4,268	4
Sciences	4,455	5	265	4	4,720	5
Mathematics	5,530	6	316	5	5,846	6
World Language	1,807	2	306	4	2,114	2
Bilingual Education	259	<1	8	<1	267	<1
Business	841	1	64	1	905	1
Technology	1,332	1	136	1	1,468	1
The Arts	4,852	5	603	9	5,455	6
Career and Technical	1,530	2	178	3	1,708	2
Wellness	3,356	4	313	5	3,669	4
Elementary Education	31,539	35	1,666	24	33,205	34
Special Education	12,585	14	887	13	13,472	14
Early Childhood	974	1	216	3	1,190	1
Alternative	222	<1	32	<1	254	<1
Other ^e	1,155	1	391	6	1,546	2
Multiple	10,708	11	527	8	11,235	11
Total Teachers	91,885	93%	6,797	7%	98,682	100%

^a Full-time teachers are those whose FTE \geq .99.

^b Part-time teachers are those whose FTE<1.

^c Includes doctoral, education specialist, law, and medical degrees.

^d Includes adult/continuing education, compensatory education, preschool, and summer school.

^e Includes agricultural science and natural resources, family and consumer education, driver and safety education, and Jr. ROTC.

Table 3.2: Characteristics of Michigan schools by teacher full-time equivalency status

	Full-time equivalency		Part-time		Total	
	N	%	N	%	N	%
ALL PUBLIC SCHOOLS						
Instructional level						
Primary	40,724	44	2,945	43	43,669	44
Middle/Junior High School	18,123	20	971	14	19,094	19
High School	24,674	27	1,700	25	226,374	27
Other	6,465	7	648	10	7,113	7
Missing	2,432	2	533	8	2,432	2
Enrollment size						
Less than 300	14,105	15	1,311	19	15,416	16
300-999	59,059	64	3,869	57	62,928	64
1,000 or more	16,208	17	1,029	15	17,234	17
Missing	2,513	3	591	9	3,104	3
Locale						
City	23,869	26	1,338	19	25,207	26
Urban fringe	37,815	41	3,022	45	40,837	41
Town	18,425	20	1,268	19	19,693	20
Rural	9,877	11	636	9	10,513	11
Missing	1,899	2	501	7	2,432	2
Percent minority enrollment in school						
Less than 5 percent	19,785	22	1,439	21	21,224	22
5 to 19%	34,935	38	2,843	42	37,778	39
20 to 49%	15,634	17	1,094	16	16,728	17
50% or more	18,341	20	711	10	19,052	19
Missing	3,190	3	710	10	3,900	4
Percent of students in school eligible for free or reduced-price school lunch						
Less than 10%	16,267	18	1,548	23	17,815	18
10 to 29%	27,045	30	2,252	33	27,045	30
30 to 49%	20,044	22	1,199	18	21,243	22
50 to 69%	11,887	13	618	9	12,505	13
70% or more	13,439	15	470	7	13,909	14
Missing	3,203	3	710	10	3,913	4
Total Teachers	89,994	93%	6,929	7%	96,923	100%

a

The school characteristic categories employed in this table are those used in Table 2.1 of the National Center for Education Statistics (NCES) Statistical Analysis Report *Teachers' Tools for the 21st Century: A Report on Teachers' Use of Technology*, NCES 2000-102. U.S. Department of Education, Office of Educational Research and Improvement.

Table 3.3: Characteristics of Teachers by Minority and Non-Minority Status				
Teacher Characteristics	Minority Teacher		Non-Minority Teacher	
	n	%	n	%
Gender				
Male	2,222	22	23,025	26
Female	7,917	78	65,470	74
Age				
Less than 25	94	1	1,528	2
25-29	874	9	11,828	13
30-39	2,874	28	26,467	30
40-54	3,647	36	31,149	35
Over 55	2,650	26	17,523	20
Highest education				
Doctoral or Specialist	111	1	844	1
Master's	5,523	54	46,281	52
Bachelor's	4,018	40	39,370	44
High School	366	3	1,310	1
Other	121	1	690	1
Multiple Subjects	332	3	10,901	12
Multiple Schools	173	2	3,642	4
Special Education	1,836	18	11,608	13
School Characteristics				
Instructional Level				
Primary	4,813	47	38,847	44
Middle/Junior high school	1,441	15	17,644	20
High School	2,659	26	23,705	27
Other	745	7	6,354	7
Percent minority enrollment				
Less than 5%	182	2	21,035	24
5-19%	606	6	37,155	42
20-49%	912	9	15,802	18
50% or greater	7,839	77	11,209	13
Percent free/reduced lunch eligible				
Less than 10%	750	7	17,047	19
10-29%	878	9	28,414	32
30-49%	1,139	11	20,098	23
50-69%	1,904	19	10,591	12
70% or greater	4,864	48	9,042	10
Locale				
City	7,295	72	17,902	20
Urban Fringe	1,893	19	38,917	44
Town	375	4	19,313	22
Rural	95	1	10,418	12
Enrollment Size				
Less than 300 students	1,429	14	13,974	16
300-999 students	6,033	60	56,876	64
Greater than 1000 students	2,148	21	15,076	17
Total Number of Teachers	10,139	10	88,495	90
Missing data: race/eth (48) instruct level & locale (2423) min enroll(3900),freelunch(3913) size(3104)				

Table 3.4: Characteristics of Teachers by Education Degree Attainment										
	Doctoral Specialist		Master's		Bachelor's		High School		Other	
	n	%	n	%	n	%	n	%	n	%
Teacher Characteristics										
Gender										
Male	329	34	12,536	24	11,572	27	563	34	247	29
Female	626	66	39,268	76	31,816	33	1,113	67	564	66
Age										
Less than 25	1	<1	22	<1	1,509	3	67	4	23	3
25-29	18	2	2,304	4	10,064	23	228	14	88	10
30-39	143	15	15,684	30	12,818	30	503	30	193	22
40-54	346	36	20,855	40	12,635	29	612	37	348	41
Over 55	447	47	12,939	25	6,362	15	266	16	159	19
Race/Ethnicity										
White	844	88	46,281	89	39,370	91	1,310	78	690	80
Black/African American	84	9	4,693	9	3,256	8	321	19	95	11
Hispanic or Latino	11	1	412	1	404	1	30	1	13	1
Asian American	13	1	262	1	208	<1	11	1	9	1
American Indian/Alaskan native	3	<1	116	<1	102	<1	4	<1	3	<1
Native Hawaiian or Other	-	<1	30	<1	28	<1	-	<1	1	<1
Multiple	-	<1	10	<1	20	<1	-	<1	-	<1
Multiple Subjects	109	11	5,723	11	5,217	12	140	9	46	6
Multiple Buildings	29	3	1,635	3	2,096	5	37	2	19	2
Special Education	147	16	7,353	14	5,556	13	237	16	179	25
School Characteristics										
Instructional Level										
Primary	290	30	23,319	45	19,217	44	574	34	269	31
Middle/Junior high school	210	22	10,661	21	7,845	18	271	16	107	12
High School	364	38	13,974	27	11,292	26	540	32	204	24
Other	62	6	2,811	5	3,872	9	179	11	189	22
Percent minority enrollment										
Less than 5%	133	14	10,299	20	10,378	24	312	19	102	12
5-19%	380	40	21,027	41	15,944	37	255	15	172	20
20-49%	216	23	9,072	18	6,900	16	282	17	258	30
50% or greater	183	19	9,686	19	8,374	19	648	39	161	19
Percent free/reduced lunch eligible										
Less than 10%	214	22	10,204	20	7,051	16	215	13	131	15
10-29%	322	34	16,455	32	12,128	28	283	17	109	13
30-49%	167	17	10,651	21	9,981	23	337	20	107	12
50-69%	92	10	5,787	11	6,037	14	386	23	203	24
70% or greater	116	12	6,979	13	6,395	15	276	16	143	17

Table 3.4 (cont'd)										
	Doctoral		Master's		Bachelor's		High School		Other	
	n	%	n	%	n	%	n	%	n	%
Locale										
City	244	26	13,966	27	10,046	23	643	38	308	36
Urban Fringe	568	59	23,181	45	16,333	38	516	31	239	28
Town	88	9	9,674	19	9,574	22	167	10	190	22
Rural	26	3	3,944	8	6,273	14	238	14	32	4
Enrollment Size										
Less than 300 students	101	11	6,887	13	8,043	19	229	14	156	18
300-999 students	524	55	33,400	64	27,606	64	989	59	409	48
Greater than 1000 students	296	31	10,197	20	6,310	15	291	17	140	16
Total Number of Teachers	955	1	51,804	52	43,388	44	1,676	2	859	1
<i>Note: Missing data = 48 on race/ethnicity, 2,423 on instructional level and locale, 3,900 on minority</i>										

Table 3.5: Characteristics of Teachers by Locale								
	City		Urban Fringe		Town		Rural	
	n	%	n	%	n	%	n	%
Teacher Characteristics								
Gender								
Male	5,978	24	10,157	25	5,414	27	3,127	30
Female	19,219	76	30,653	75	14,274	73	7,386	70
Age								
Less than 25	363	1	747	2	305	2	161	2
25-29	2,863	11	5,789	14	2,522	13	1,228	12
30-39	6,943	28	12,985	32	5,843	30	3,013	29
40-54	8,954	36	13,582	33	7,297	37	4,075	39
Over 55	6,074	24	7,707	19	3,721	19	2,036	19
Race/Ethnicity								
White	17,902	71	38,917	95	19,313	98	10,418	99
Black/African American	6,494	26	1,372	3	160	1	11	<1
Hispanic or Latino	450	2	247	1	105	1	26	<1
Asian American	234	1	196	<1	48	<1	11	<1
American Indian/Alaskan native	92	<1	46	<1	39	<1	40	<1
Native Hawaiian or Other	19	<1	21	<1	17	<1	1	<1
Multiple	6	<1	11	<1	6	<1	6	<1
Highest education								
Doctoral or Specialist	244	1	568	1	88	<1	26	<1
Master's	13,966	55	23,181	57	9,674	49	3,944	38
Bachelor's	10,046	40	16,333	40	9,574	49	6,273	60
High School	643	3	516	1	167	1	238	2
Other	308	1	239	<1	190	1	32	<1
Multiple Subjects	1,534	6	4,614	11	2,949	15	2,023	19
Multiple Schools	527	2	1,130	2	1,081	5	1,033	10
Special Education	3,817	15	5,368	13	2,558	13	1,178	11
School Characteristics								
Instructional Level								
Primary	12,874	51	18,601	46	8,074	41	4,120	39
Middle/Junior high school	4,105	16	8,790	22	4,249	22	1,950	19
High School	6,399	25	11,124	30	5,843	30	3,008	29
Other	1,829	7	2,322	6	1,527	8	1,435	14
Percent minority enrollment								
Less than 5%	684	3	6,818	17	7,635	39	6,087	58
5-19%	4,347	17	21,188	52	8,951	45	3,292	31
20-49%	5,106	20	8,540	21	2,224	11	858	8
50% or greater	14,715	58	3,884	10	306	2	147	1

Table 3.5 (cont'd)								
	City		Urban Fringe		Town		Rural	
	n	%	n	%	n	%	n	%
Percent free/reduced lunch eligible								
Less than 10%	2,764	11	10,611	26	3,274	17	1,166	11
10-29%	3,839	15	16,495	40	7,150	36	1,813	17
30-49%	3,045	12	7,828	19	6,163	31	4,208	40
50-69%	4,934	20	3,291	8	1,916	10	2,364	22
70% or greater	10,270	41	2,198	5	610	3	831	8
Enrollment Size								
Less than 300 students	3,955	16	4,414	11	3,623	18	3,424	33
300-999 students	15,763	63	26,889	66	13,462	68	6,814	65
Greater than 1000 students	5,337	21	9,417	23	2,276	12	204	2
Total Number of Teachers	25,207	26	40,837	42	19,693	20	10,513	11
<i>Note: There are 2,432 teachers with missing data on locale.</i>								

Table 3.6: License Demographics for Michigan's Instructional Workforce

	License Type* (N)						
	Provisional	Professional	18/30 Cont	Vocational	Non-Instr.	Unknown	Total
Age							
<25	945	5	0	3	14	62	1,029
25-29	8,533	3,217	0	40	75	128	11,993
30-39	6,485	22,452	8	158	215	367	29,685
40-54	3,236	16,521	13,908	429	226	647	34,967
>=55	519	7,260	13,445	178	122	245	21,769
[Unknown]	0	0	0	0	0	49	49
Gender							
Male	5,347	12,357	6,666	488	189	316	25,363
Female	14,371	37,098	20,695	320	463	1,133	74,080
[Unknown]	0	0	0	0	0	49	49
Highest Educational Level							
Doctoral/Specialist	44	414	473	8	8	11	958
Master's Degree	2,836	30,267	18,304	155	56	251	51,869
Bachelors Degree	16,180	18,051	8,069	358	475	490	43,623
High School	499	512	310	144	72	492	2,029
Other	159	211	205	143	40	251	1,009
[Unknown]	0	0	0	0	1	3	4
Minority Status	Provisional	Professional	18/30 Cont	Vocational	Non-Instr.	Unknown	Total
Non-Minority Teacher	16,983	45,477	24,782	685	210	971	89,108
Minority Teacher	2,735	3,978	2,579	123	442	478	10,335
[Unknown]	0	0	0	0	0	49	49
Full/Part-Time Status	Provisional	Professional	18/30 Cont	Vocational	Non-Instr.	Unknown	Total
Part-Time	1,595	3,300	1,579	90	127	450	7,141
Full-Time	18,123	46,155	25,782	718	525	1,048	92,351
Total	19,718	49,455	27,361	808	652	1,498	99,492

* For those with multiple licenses, license type refers to the teacher's "highest" license, where Professional>18/30 Hour Continuing>Provisional>Vocational>Other>Missing.
Non-Instructional includes Administrative, Support Staff, and Substitute licenses.

Table 3.6 (cont'd)

	License Type (%)					
	Provisional	Professional	18/30 Cont	Vocational	Non-Instr.	Unknown
Age						
	92%	0%	0%	0%	1%	6%
	71%	27%	0%	0%	1%	1%
	22%	76%	0%	1%	1%	1%
	9%	47%	40%	1%	1%	2%
	2%	33%	62%	1%	1%	1%
	0%	0%	0%	0%	0%	100%
Gender						
	21%	49%	26%	2%	1%	1%
	19%	50%	28%	0%	1%	2%
	0%	0%	0%	0%	0%	100%
Highest Educational Level						
	5%	43%	49%	1%	1%	1%
	5%	58%	35%	0%	0%	0%
	37%	41%	18%	1%	1%	1%
	25%	25%	15%	7%	4%	24%
	16%	21%	20%	14%	4%	25%
	0%	0%	0%	0%	25%	75%
Minority Status						
	19%	51%	28%	1%	0%	1%
	26%	38%	25%	1%	4%	5%
	0%	0%	0%	0%	0%	100%
Full/Part-Time Status						
	22%	46%	22%	1%	2%	6%
	20%	50%	28%	1%	1%	1%
Total	20%	50%	28%	1%	1%	2%

Table 3.7: Out of Field in Core Subjects

*Includes teachers with at least one high school assignment.

Math Assignment	Endorsed in Subject?		<i>Total</i>
	No	Yes	
Yes	3%	97%	100%
	123	3,959	4,082
No	95%	5%	100%
	19,396	1,016	20,412
<i>Total</i>	80%	20%	100%
	19,519	4,975	24,494

ELA Assignment	Endorsed in Subject?		<i>Total</i>
	No	Yes	
Yes	3%	97%	100%
	164	5,017	5,181
No	87%	13%	100%
	16,865	2,448	19,313
<i>Total</i>	70%	30%	100%
	17,029	7,465	24,494

Science Assignment	Endorsed in Subject?		<i>Total</i>
	No	Yes	
Yes	2%	98%	100%
	96	3,875	3,971
No	89%	11%	100%
	18,321	2,202	20,523
<i>Total</i>	75%	25%	100%
	18,417	6,077	24,494

Social Studies Assignment	Endorsed in Subject?		<i>Total</i>
	No	Yes	
Yes	3%	97%	100%
	122	4,163	4,285
No	75%	25%	100%
	15,208	5,001	20,209
<i>Total</i>	63%	37%	100%
	15,330	9,164	24,494

World Language Assignment	Endorsed in Subject?		<i>Total</i>
	No	Yes	
Yes	3%	97%	100%
	64	1,804	1,868
No	97%	3%	100%
	21,999	627	22,626
<i>Total</i>	90%	10%	100%
	22,063	2,431	24,494*

Table 3.8: Teaching Assignment by Endorsement Pattern (Potential "Reserve" of Teachers)

Cells contain row percentages and n's

	Subject Assignment							
	Alt	Arts	Biling Ed	Business	Career Tech	Elem	ELA	Math
Math-X	16 1%	17 1%	0 0%	8 0%	2 0%	3 0%	39 1%	2228 83%
Lang.Arts-X	37 1%	94 3%	1 0%	10 0%	6 0%	2 0%	2292 68%	207 6%
Science-X	24 1%	25 1%	0 0%	4 0%	2 0%	6 0%	95 3%	821 30%
SocialSci-X	37 1%	124 3%	2 0%	18 0%	2 0%	6 0%	1164 29%	540 14%
World Lang-X	2 0%	22 2%	3 0%	1 0%	0 0%	1 0%	142 13%	66 6%
	Subject Assignment							
	Science	Social Sciences	Spec Ed	Tech	Wellness	World Lang	Misc	NA
Math-X	222 8%	17 1%	3 0%	34 1%	25 1%	40 1%	15 1%	2 0%
Lang.Arts-X	52 2%	257 8%	11 0%	15 0%	76 2%	272 8%	54 2%	4 0%
Science-X	1521 55%	75 3%	3 0%	18 1%	104 4%	29 1%	29 1%	1 0%
SocialSci-X	138 3%	1352 34%	14 0%	35 1%	231 6%	239 6%	66 2%	6 0%
World Lang-X	10 1%	35 3%	0 0%	2 0%	7 1%	767 72%	9 1%	1 0%
								<i>Total</i>
								2671 100%
								3390 100%
								2757 100%
								3974 100%
								1068 100%

Table 3.9: Teacher Labor Supply in Michigan, 2006-2008						
	2006	%	2007	%	2008	%
Overall Teaching Workforce						
Retained in the profession	102,365	92%	102,439	93%	101,379	94%
Left the profession (attrition rate)	9,609	8%	8,616	7%	8,536	6%
Entered the profession	8,690		7,476		6,414	
Total number of teachers in summer count	111,055		109,915		107,793	
Within-School Mobility Rate						
Stayers (same school)	91,234	89%	91,471	89%	90,247	89%
Movers (different school)	11,131	11%	10,968	11%	11,132	11%
Total retained teachers	102,365	100%	102,439		101,379	

Table 3.10: Mean Teacher Retention Rate by High School Characteristics

	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>F-test</i>
Locale				
City	79.73	14.91	96	.000
Suburb	87.10	6.57	161	
Town	88.60	4.65	78	
Rural	87.37	5.70	245	
Percent minority				
Less than 5%	87.93	5.14	174	.000
5-10%	88.73	4.96	141	
10-15%	89.33	3.98	56	
15-65%	86.42	5.80	132	
Greater than 65%	74.98	15.21	77	
Percent free/reduced lunch				
Less than 10%	88.30	6.35	43	.000
10-29%	89.14	4.27	202	
30-49%	87.36	5.31	204	
50-69%	80.27	11.87	96	
70% or greater	76.09	16.19	35	
School size				
Less than 300 students	81.81	13.23	85	.000
300-999 students	86.40	8.34	307	
Greater than 1000 students	87.84	4.86	188	
Charter Schools				
Charter	71.91	17.06	33	.000
Non-charter	87.06	6.91	547	
Magnet Schools				
Magnet	87.26	6.86	65	.139
Non-magnet	86.06	8.76	515	
Percent teachers with prof licenses				
Less than 80%	79.15	13.33	134	.000
80-90%	87.69	4.66	297	
Greater than 90%	89.55	4.85	149	
Percent minority teachers				
Less than 1%	88.15	5.45	275	.000
1-10%	87.81	4.91	214	
Greater than 10%	76.51	14.58	91	
Percent highly qualified teachers				
Less than 75%	85.97	7.48	197	
75-85%	86.45	7.33	144	
Greater than 85%	86.22	10.01	239	
<i>Total</i>	<i>86.20</i>	<i>8.57</i>	<i>580</i>	

Teacher retention rate is the average teacher retention rate for each school from 2005-2008

Table 3.11: Mean Mathematics Teacher Undersupply by High School Characteristics

	<i>Mean</i>	<i>SD</i>	<i>N</i>	
Locale				
City	.64	1.68	94	.000
Suburb	.83	1.37	161	
Town	.51	.90	78	
Rural	.18	.84	245	
Percent minority				
Less than 5%	.31	.81	174	.000
5-10%	.46	1.12	141	
10-15%	.64	1.13	56	
15-65%	.58	1.42	132	
Greater than 65%	.63	1.65	75	
Percent free/reduced lunch				
Less than 10%	.04	1.60	43	.000
10-29%	.67	1.13	202	
30-49%	.37	1.02	204	
50-69%	.62	1.44	94	
70% or greater	.29	1.18	35	
School size				
Less than 300 students	-0.15	.73	83	.000
300-999 students	.30	.91	307	
Greater than 1000 students	1.06	1.52	188	
Charter Schools				
Charter	.06	1.08	33	.425
Non-charter	.51	1.21	545	
Magnet Schools				
Magnet	.22	1.38	65	.080
Non-magnet	.52	1.12	513	
Percent teachers with prof licenses				
Less than 80%	.35	1.16	132	.366
80-90%	.52	1.31	297	
Greater than 90%	.42	1.00	149	
Percent minority teachers				
Less than 1%	.28	.98	275	.000
1-10%	.72	1.22	214	
Greater than 10%	.54	1.63	89	
Percent highly qualified teachers				
Less than 75%	.74	1.34	196	.000
75-85%	.46	1.23	144	
Greater than 85%	.28	1.02	238	
<i>Total</i>	<i>.48</i>	<i>1.20</i>	<i>578</i>	

Math undersupply is calculated for SY 2009 based on 2008 REP, student enrollment from 2006-2008

Table 3.12: Mean English Language Arts Teacher Undersupply by High School Characteristics

	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>F-test</i>
Locale				
City	-.81	2.56	94	.003
Suburb	-1.10	1.81	161	
Town	-.89	1.06	78	
Rural	-.52	.88	245	
Percent minority				
Less than 5%	-.61	1.01	174	.062
5-10%	-.69	1.15	141	
10-15%	-1.17	1.79	56	
15-65%	-1.00	1.67	132	
Greater than 65%	-.67	2.64	75	
Percent free/reduced lunch				
Less than 10%	-1.43	1.93	43	.037
10-29%	-.78	1.38	202	
30-49%	-.79	1.31	204	
50-69%	-.58	2.10	94	
70% or greater	-.46	1.87	35	
School size				
Less than 300 students	-.43	.70	83	.006
300-999 students	-.71	1.36	307	
Greater than 1000 students	-1.05	2.09	188	
Charter Schools				
Charter	-.86	2.81	33	.752
Non-charter	-.77	1.48	545	
Magnet Schools				
Magnet	-.86	2.27	65	.659
Non-magnet	-.77	1.48	513	
Percent teachers with prof licenses				
Less than 80%	-.58	1.91	132	.130
80-90%	-.90	1.59	297	
Greater than 90%	-.72	1.18	149	
Percent minority teachers				
Less than 1%	-.58	1.04	275	.008
1-10%	-1.03	1.57	214	
Greater than 10%	-.79	2.62	89	
Percent highly qualified teachers				
Less than 75%	-.59	1.59	196	.086
75-85%	-.79	1.41	144	
Greater than 85%	-.93	1.67	238	
Total	<i>-.78</i>	<i>1.58</i>	<i>578</i>	

ELA undersupply is calculated for SY 2009 based on 2008 REP & student enrollment from 2006-2008

Table 3.13: Mean Science Teacher Undersupply by High School Characteristics

	<i>Mean</i>	<i>SD</i>	<i>N</i>	<i>F-test</i>
Locale				
City	-1.04	1.82	94	.000
Suburb	-1.07	1.36	161	
Town	-.68	.89	78	
Rural	-.50	.71	245	
Percent minority				
Less than 5%	-.47	.75	174	.000
5-10%	-.73	.93	141	
10-15%	-.99	1.02	56	
15-65%	-1.21	1.59	132	
Greater than 65%	-.58	1.54	75	
Percent free/reduced lunch				
Less than 10%	-1.96	1.91	43	.000
10-29%	-.90	1.11	202	
30-49%	-.56	.89	204	
50-69%	-.53	1.23	94	
70% or greater	-.39	1.16	35	
School size				
Less than 300 students	-.46	.64	83	.000
300-999 students	-.50	.88	307	
Greater than 1000 students	-1.34	1.58	188	
Charter Schools				
Charter	-.25	1.09	33	.464
Non-charter	-.80	1.20	545	
Magnet Schools				
Magnet	-.74	1.13	65	.501
Non-magnet	-.77	1.21	513	
Percent teachers with prof licenses				
Less than 80%	-.50	1.13	132	.028
80-90%	-.88	1.28	297	
Greater than 90%	-.77	1.06	149	
Percent minority teachers				
Less than 1%	-.54	.80	275	.000
1-10%	-1.06	1.40	214	
Greater than 10%	-.77	1.53	89	
Percent highly qualified teachers				
Less than 75%	-.80	1.34	196	.000
75-85%	-.80	1.31	144	
Greater than 85%	-.72	.99	238	
<i>Total</i>	<i>-.77</i>	<i>1.19</i>	<i>578</i>	

Science undersupply is calculated for SY 2009 based on 2008 REP & student enroll from 2006-2008

CHAPTER 4: ESTIMATING TEACHER DEMAND USING STATE ADMINISTRATIVE DATA: CHALLENGES AND RESOLUTIONS

The demand formula presented briefly in Chapter 1 provides a powerful tool for use by practitioners and by researchers in order to estimate teacher demand in the context of specific curricular requirements. While easily accessible to practitioners and researchers alike, the formula makes several assumptions that bear further investigation. The purpose of this essay is to revisit the demand formula and its underlying assumptions, test the assumptions where appropriate, and make an informed decision regarding possible modifications to the formula for future use. The revised formula is then used in Chapter 7, Teacher Undersupply.

Background to the Problem: Estimating Teacher Demand

In order to provide students with increased rigor in their high school coursework and prepare students more fully for the demands of a global technological economy, many states have increased graduation requirements. Since 2004, 18 states plus the District of Columbia report having raised graduation requirements to meet the American Diploma Project's college- and work- career-ready curriculum, which includes 4 years of challenging math and English, and an additional 12 states plan to do so in the next few years (Achieve, Inc., 2008). Forty-two states have at least some statewide requirements for high school graduation (American Association of State Colleges and Universities, 2006), and 25 states now offer an optional college-preparatory diploma (Dounay, 2006).

Michigan is among the set of states leading the way in these reforms, adopting one of the most comprehensive sets of high school graduation requirements in the country, known as the Michigan Merit Curriculum in 2006. The Merit Curriculum is

meant to improve high school content and better align the Michigan diploma with business and postsecondary requirements through more rigorous curriculum, standards, and assessment (Cherry Commission, 2004). By establishing a statewide set of required courses, the Merit Curriculum creates a minimum floor that might raise expectations and achievement for students. This reform draws on a small but growing body of research, suggesting that certain core courses, especially those in math and science, can have significant effects on students' long-term labor outcomes (Goodman, 2010; Levine & Zimmerman, 1995; Rose & Betts, 2004). The intervention might also positively influence students' educational aspirations by standardizing high school course-taking around requirements for a postsecondary education (Bryk, Lee, & Holland, 1993; Lee, 2002).

While other states are increasing graduating requirements, Michigan stands out for the rigor and specificity of courses required under the new policy, including Algebra 1, geometry, Algebra 2, Biology 1, chemistry or physics, and at least two years of foreign language. While these courses are likely to prove beneficial for student achievement outcomes,¹ the question remains: will Michigan, and other states implementing similar reforms, have an adequate supply of teachers to meet the increased instructional demands from a curriculum such as this one?

¹ It is important to note that, while increased graduation requirements are generally thought to improve student achievement and graduation rates (Achieve, 2009; Balfanz & West, 2009), there is some evidence that increased graduation requirements are associated with lower high school completion rates (Lillard & DeCicca, 2001), and that mandatory high school graduation exams, a closely related policy reform, increase dropout rates, particularly among low-income students (Dee & Jacob 2007; Jacob 2001; Warren, Jenkins, & Kulick, 2006) and little evidence that they improve student achievement (Grotsky, Warren, & Kalogrides, 2009; Dee & Jacob 2007). Given this mixed literature and the importance of evaluating this statewide reform, the Michigan Merit Curriculum is the topic of an ongoing, IES-funded study, conducted by the Michigan Consortium for Educational Research, to establish the effectiveness of the MMC in increasing student achievement and postsecondary transitions (Michigan Consortium for Educational Research, 2010).

Estimating demand to respond to changes in curricular requirements

Traditional economic-based demand formulas center on the concept of estimating demand as a function of price or cost, and in general assume that demand decreases as price increases (Ehrenberg & Smith, 1997). Economists are also interested at the point at which the supply and demand curves meet, or are in equilibrium (i.e. there is enough supply to meet the demand with no excess). When looking at the predicted demand of teachers required to meet instructional needs arising from a set of specified curricular requirements, a different type of formula is required. The goal is not to predict the number of teachers needed based on general cost or population changes. Instead, the goal is to take a point-in-time change (increase in curricular requirements) and estimate the corresponding increase in teacher time necessary to meet those requirements.

Much current work calculates teacher demand based on how many positions are open at a given level of compensation (Arnold, Choy & Bobbitt, 1993; Boe & Gilford, 1992; Ingersoll, 2001; Ingersoll & Perda, 2009; Guarino, Santibanez, & Daley, 2006). This focuses on funded positions, rather than on the number of sections of a course that will need to be offered to meet curricular requirements. Looking only at job openings does not take into account schools who may need teachers but who are not able to post for jobs because of budgetary constraints, increased class size, or other organizational issues. Demand can also be expressed as the changes due to growth or decline in enrollment, the student/teacher ratio, or staff requirements and the loss of teachers due to attrition (Arnold, Choy, & Bobbitt, 1993).

There are two types of demand: constrained demand (the number of teachers employed) and unrestricted demand (the number of teachers who might be hired without

constraints; Carroll, Reichardt, & Guarino, 2000). The method for estimating demand presented here is a version of unrestricted demand, although it does represent a *necessary* amount of unrestricted demand, in that the course offerings must be made available to students in order to meet the needs of the merit curriculum.

There are several methods available for calculating teacher demand. One demand calculation is provided by OECD (Santiago, 2002). In this calculation, the number of teachers needed is determined as follows: (student population/average class size) x (average number of required learning hours for students/teachers' teaching load). However, the drawback of this formula is that it ignores the level of detail associated with subject area, something that is remedied in the formula presented below. Another common formula for demand, usually estimated at the state or national level, is $\text{demand} = \text{enrollment} \times \text{pupil/teacher ratio}$ (Boe & Gilford, 1992). Both of the above analyses were conducted at either the national or state level, highlighting one of the challenges identified with many of the demand models available (i.e. the one used by NCES, the OECD model, etc.). These methods produce national level projections that are not able to provide adequate and accurate information for policymakers at the school and district level (Carroll, Reichardt, & Guarino, 2000).

While the demand formula presented above is similar in nature to other formulas, this work makes two unique contributions. The first is that this method provides estimates at the school level, and also utilizes local estimates for elements such as enrollment and class size. The second is that this paper tests the validity and accuracy of this method. While there are several widely used ways to estimate teacher demand, far more attention has been paid to teacher supply estimation methods. This paper provides an opportunity

to gather and present evidence regarding the extent to which these methods produce reasonable estimates.

Demand Formula, Assumptions, and Alterations

The developed demand formula is as follows:

$$D_i = \frac{a(x_i)/y}{z} \quad (1)$$

where

D_i = number of teachers needed to meet graduation requirements in a subject area

a = proportion of student body that needs to be enrolled in each subject each year in order to meet graduation requirements

x_i = total student enrollment in each school

y = class size

z = number of periods taught per FTE per day²

Assumptions

Enrollment weight (a): The proportion of the student body expected to take a given subject in each year, in order to meet the graduation requirements. For math and English language arts, since four years of each are required, it is assumed that 100% of the student body takes a math and an English course each year, and thus the value of a is 1.0 for mathematics and English language arts. For science and social studies, $a=.75$ because students are required to take three years of each.

These weights represent a lower-bound estimate of the number of students in a school who are taking a particular subject. They do not take into account factors such as students who take more courses in a given subject than required or students who have to

² See Keesler, Wyse & Jones (2008) on the IES website (http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/techbrief/tr_00508.pdf). This formula has been vetted by IES and is considered as a promising tool for use by other states. While the report was a collaborative effort, the development of the formula was an individual contribution on my part.

re-take courses. The actual number of students taking courses may be higher than would be estimated by this.

Student enrollment (x_i): Student enrollment is calculated by taking the average enrollment over the previous three years. There is no differentiation by grade level; the enrollment is for the entire student body. The implied assumption here is that any factors contributing to the enrollments of the three previous years—i.e. rates of drop outs, cohort sizes, student attrition—will remain constant in the future.

Class size (y): Class size is assumed to be 25, reflecting common practices in many high schools. However, this assumption can be changed to reflect more local conditions or other assumptions (i.e. if a policymaker wanted to test the potential effect of decreasing class size on the necessary teaching workforce, the assumptions could be altered to a smaller class). This number also reflects an “ideal” class size in that it is an agreed-upon size that is not too large but also not unreasonably small given the practices and realities of most schools.

Number of periods per FTE (z): Number of periods taught per FTE per day (z) is assumed to be five.

It is important to note that this formula produces *conservative* estimates of demand, by estimating the minimum demand, given a certain set of curricular requirements. Therefore, schools that have inadequate supply to meet the demand can be considered rather significantly undersupplied; there may be more schools identified if the demand calculation were more liberal. The assumptions reflect common practices in many high schools. However, these assumptions are not reflective of *all* schools, and

particularly are not valid for schools with non-traditional organizational patterns; this should be taken into account when calculating demand for individual schools.³

Testing Each Assumption

The core purpose of this paper is to revisit the assumptions of the current demand formula, evaluating the formula for rigor and for its ability to produce reliable demand estimates. The next section details the steps taken to test each assumption.

Enrollment estimates. In the original formula, enrollment is estimated by averaging three years of enrollment data for each high school in the state. However, Michigan is a state with declining population, and net out-migration. Additionally, it is not clear from looking at the past enrollments how “shocks” to the system, such as significant economic decline, major layoffs and factory closings, and home foreclosures and bankruptcies (all of which have been experienced heavily in Michigan) affect school enrollments. Finally, as mentioned previously, simply taking an average of the previous three years of enrollment data assumes that any trends in enrollment represented by those data (i.e. drop out rate, incoming class size) will remain relatively static. In order to address this, enrollment projections are calculated using single exponential smoothing with a 0.4 smoothing constant to determine state level public school enrollments, as well as with a 0.7 smoothing constant for comparison, following the methods used by NCES (Hussar & Bailey, 2008). Exponential smoothing places more weight on recent observations than on earlier ones (Hussar & Bailey, 2008). By using three years of data,

³ The 580 high schools included in this study are schools that have more “traditional” configurations, as alternative schools and schools that were not coded as “regular” for any reason were not included. Therefore, this formula applies reasonably well to the sample analyzed in this dissertation. See Chapter 2: Data and Methods, for more information on sample selection.

and these smoothing constants, the enrollment rates reflect the realities of enrollment in Michigan more precisely than a simple average.

Districts and states can use a variety of different methods to predict overall cohort enrollment, such as the cohort survival model (California Department of Finance, 2006), the ratio model (Campbell, 1997), the extended demographic model, which uses enrollment data, birth records, student dropouts, and migration and grade retention records (Campbell, 1997), and a multiregional cohort enrollment model that allows for intradistrict mobility and school choice, and requires a shorter time scale (Sweeney & Middleton, 2005). These strategies were all researched and considered for use in this analysis. However, they are all most appropriate for predicting enrollment for larger units of analysis, such as districts, and required data inputs that are not readily available at the school level, such as birth records. The weighted average with a smoothing constant allowed for the calculation of individual school enrollment projections using available data.

Class size and FTEs taught per person. In the original application of the formula with Michigan data, class size was assumed to be 25, and the state was provided with alternate estimates under assumptions of 15, 20, and 30 students per class. The number of FTEs taught per teacher per day was assumed to be five. However, a small value in the denominator can change the estimates of demand considerably, and since both of these numbers are found in the denominator, it is particularly important to produce very accurate estimates.

Class size is a critical and well-researched issue in education, although there is lack of a consensus on the effect of class size. When there are studies of actual class size

(not student teacher ratio), the findings are still mixed. Some meta-analyses show that reduced class sizes do not systematically lead to improved student achievement (Hanushek, 1997), while others (Hedges, Laine, & Greenwald, 1994; Kruger, 2000) the effect may be more positive. Results from the Tennessee STAR experiment, one of the largest randomized field experiments of reduced class sizes, show that reduced class size has a significantly positive effect on test scores, particularly for children from disadvantaged situations (Finn & Achilles, 1990; Kruger, 1999). However, these effects are questioned by others (Hanushek, 1999) who are concerned that bias, implementation issues and design issues may reduce the magnitude of any treatment effect, and that the expensive nature of reduced class sizes is not necessarily merited by the findings. Twenty-five students was assumed to be a reasonable class size—one that was small enough that it did not place unreasonable instructional burdens on the teacher, but large enough that it did not reflect the more costly “class size reduction” reforms.

One important caveat, however, is that using an assumed class size allows a principal, district superintendent, or state policymaker to estimate demand *based on an ideal class size* or based on potential class sizes due to a proposed change in class size as part of an educational intervention. There may be situations in which it is not desirable to know the exact class size because the demand projection may need to reflect desired conditions, not current conditions. For example, if a school is already undersupplied and using large classes in math in order to meet the demand, then calculating the student-teacher ratio in math and using that to estimate demand is not helpful because it will make it appear as though the school is adequately staffed, when in fact they are undersupplied because of large class sizes. For this reason, the class size will stay at an

estimated level (25) for this analysis at this time.⁴ Given the budgetary restrictions and severe economic climate facing schools today, reflecting actual class sizes is not optimal, as many school districts are dealing with teacher and funding shortages by demanding more of their current instructional staff. This is a situation in which reflecting the reality is not desirable; reflecting a reasonable class size goal will produce more realistic estimates of which schools are likely struggling to meet their instructional needs given their current staff.

FTEs per person are assumed to be five, based on common practice and suggestion by the state. This again will be left as an assumption, as it may be another area that is affected negatively by budgetary shortfalls. However, in order to test the possible implications for varying FTEs, an assumed distribution is generated and used for comparative purposes (see below).

Weighting the enrollment. Currently, the (a) value is defined as 1.0 for mathematics and English language arts and .75 for science and social studies. This is based on the assumption that, if students need to take four years of mathematics and English language arts, they will take mathematics and English language arts each year they are in school, and thus all students in a given school will be taking a mathematics and English language arts class each year—which means that 100% of the total enrollment will need to have courses available to them. In reality, this is likely to be a

⁴ In many traditional demand calculations, class size is approximated by student-teacher ratio (Santiago, 2002). This precludes a subject-specific student/teacher ratio. It was possible to calculate student/teacher ratio for each subject for this analysis, but this strategy was rejected as it has the same limitations as identifying the actual class size; it may reflect reality and therefore not allow areas of undersupply to be addressed. Additionally, that does not address the fact that certain classes may be small while others large within the same subject.

lower bound estimate, as students may either elect to take more than the minimum number of courses or may need to retake courses in order to obtain credit.⁵

Michigan does not currently have student transcript data available, making it virtually impossible to construct individual course taking histories from the data system in order to understand the number of courses taken in each subject by each student in a given year. However, beginning in the winter of 2011, Michigan will begin collecting bi-annual student transcript data from every high school in the entire state, via the Michigan eTranscript initiative. At this time, the data will be available to understand coursetaking patterns for Michigan students. Using these data when they become available, this weight on the enrollment will be adjusted. However, given the lack of reasonable options for these data at the time of this writing, the previously defined weights will continue to be used. As stated above, they produce conservative estimates of demand, as it is unlikely that students would take any fewer courses per year than the weights reflect, and may in fact take more, which would increase demand.

Challenges with ratios. The demand formula outlined above is a ratio of enrollment to class size and FTEs per teacher. Ratios can be very sensitive to small changes in the denominator (Rice, 2007). There are statistical methods for constructing confidence intervals around ratio estimates (the Delta method or the Fieller Method); however, these methods require that variances of both the numerator and the denominator can be produced. In the current equation, variance estimates could be constructed around

⁵ Interestingly, in May 2010, the Michigan legislature passed a bill repealing the requirement that students pass Algebra II. This requirement was likely to have caused more students to have to retake mathematics courses (Detroit Free Press, May 7, 2010). Anecdotal evidence also suggests that students who failed Algebra II in school are obtaining that credit via credit recovery through programs like Michigan Virtual University (MCER internal meeting, May 7, 2010). These trends suggest that a weight of 1.0 may not be a lower bound, if students are only taking Algebra II once in high school, or not required to take it at all.

enrollment, but not around class size or FTEs because those are constants. I can assume a sampling distribution around both class size and FTE and use that assumed distribution to test the sensitivity of the ratio to changes in class size or FTE. For example, I can assume that class size has a standard deviation of 0.5, which would mean that 64% of all class sizes fall within one student (from 24.5 to 25.5). I can then calculate the variance of the ratio and see how variable the demand calculation is under that assumed standard deviation, then repeat the procedure with other standard deviations (Rice, 2007).

Testing and Validating the Formula

In the next section, differences in the distribution of demand and undersupply estimates under the original formula and the distribution of demand and undersupply estimates under the tested formula are examined. The complicating aspect here is that one method is not, by definition, “better” than the other one, so the comparison is difficult. It can be assumed that the second method is more rigorous because of the increased computational rigor introduced, but this may be a tenuous assumption. The computational rigor of the second method may produce estimates that are more realistic to actual conditions, but that are less valuable in terms of highlighting schools that could use additional staffing resources.

Therefore, the key questions addressed when testing the formula are: Do these two formulas produce similar estimates? If not, how are they different? What is the range of demand under the original formula, versus under the new formula? Do the two formulas produce distributions with different qualities? More importantly, for which schools are the differences in estimates most pronounced? In other words, do the two

methods work equally well for traditional schools but one has an advantage in less traditional schools?

Rather than having an absolute test or standard by which these methods are compared, the distributional aspects produced under each method will be examined for patterns and evidence, and the strengths and weaknesses of each method will be discussed, with evidence arrayed for the use of one over the other based on a variety of criteria. The end goal is to make a recommendation regarding which formula should be utilized by policymakers. Is there evidence to suggest that the more computationally rigorous formula produces “better” estimates? Is there evidence that one of the formulas works more efficiently than the other for a certain type of school? These conclusions are necessary addendums to the method, in order to provide better guidance to end users regarding how to use the formula and how to understand its limitations.

Analytic Strategy

To implement and test the modifications described above, this chapter first tests various enrollment specifications and identifies one that seems optimal by comparing the distributions of demand and undersupply under each specification. Next, distributional assumptions are added to class size and classes taught per FTE, and these are used in conjunction with the enrollment estimate identified by the first step. Distributions of demand and undersupply are again produced and compared. The final stage investigates the differential functioning of various demand scenarios by looking at changes in the distribution of undersupply by school characteristics, and classifications of significant undersupply under different demand specifications. The chapter concludes by identifying

the most optimal demand formula, based on all accumulated evidence, and this formula is then used to conduct the undersupply analyses in Chapter 7.

As stated above, there is no one test that conclusively proves which demand formula is the “best.” The approach taken, therefore, is one of accumulating evidence to support or reject modifications to assumptions, and seeking to iteratively implement a modification, assess the distributional changes, and make an appropriate decision based on evidence.

Results

Part I: Choosing an optimal enrollment estimate

Enrollment rates were calculated using three methods: a simple three-year average, a projected enrollment using three years of data and a 0.7 smoothing constant; and a projected enrollment using three years of data and a 0.4 smoothing constant. The formula for projected enrollment with a smoothing constant is:

$$P = \alpha X_t + \alpha(1 - \alpha)X_{t-1} + \alpha(1 - \alpha)^2 X_{t-2}$$

where

P = projected value

α = smoothing constant (either 0.4 or 0.7)

X_t = observation for time t

Below, Figure 4.1 presents the histograms for mathematics demand calculations under the three enrollment estimates, followed by the undersupply calculations. For the sake of parsimony, only mathematics is presented here, although similar calculations and comparisons were conducted for English language arts and science. These histograms are included in the Technical Appendix B.

Figure 4.1: Comparison of Mathematics Demand With Varying Enrollment

Estimates

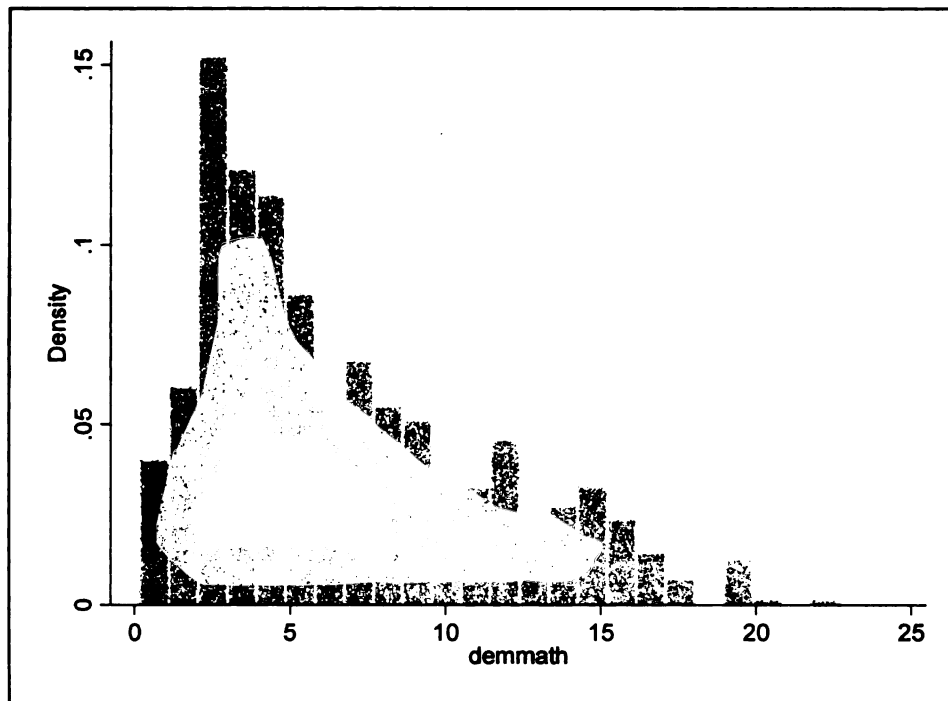


Figure 4.1.1: Original formula

Figure 4.1 (cont'd)

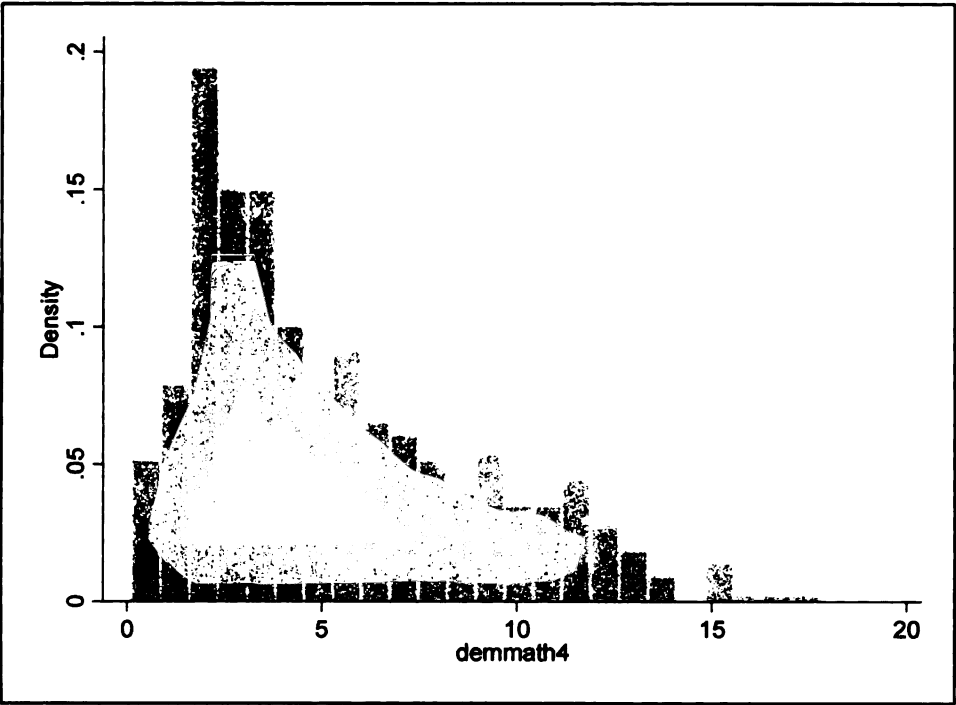


Figure 4.1.2: Smoothing constant 0.4

Figure 4.1 (cont'd)

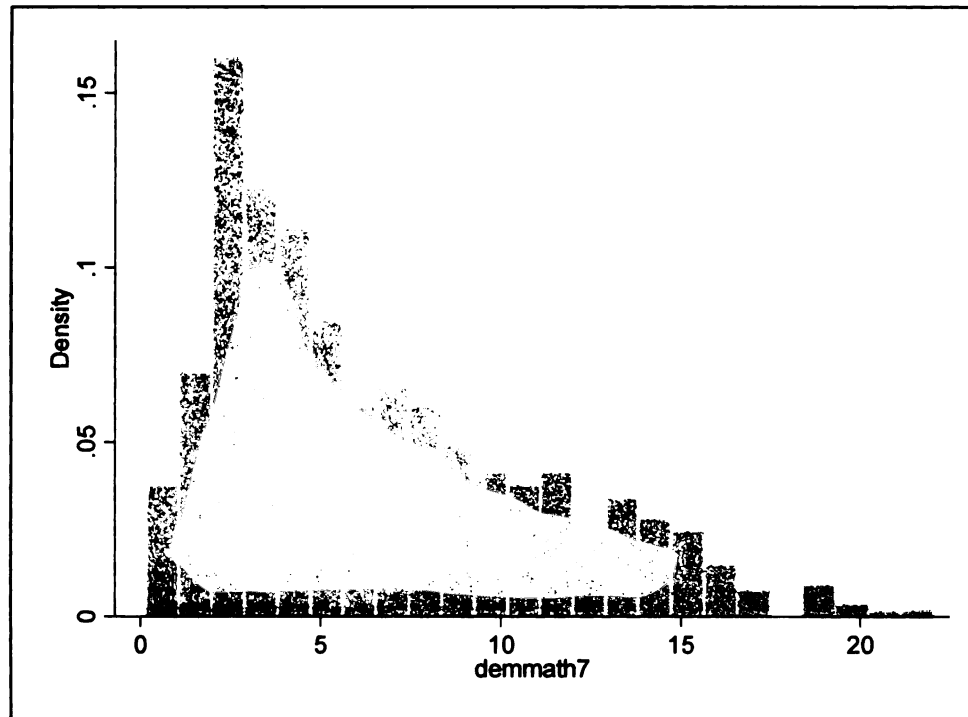


Figure 4.1.3: Smoothing constant 0.7

Viewing the histograms, they are shaped similarly, which suggests that there are not dramatically different distributions of demand regardless of enrollment estimation methods. The distribution is slightly more spread out in the original and in the 0.7 smoothing constant demand estimations, while the 0.4 smoothing constant produces a slightly more truncated distribution.

Turning to the undersupply calculations, the three distributions are presented below:

Figure 4.2: Comparison of Mathematics Undersupply with Different Enrollment

Estimates

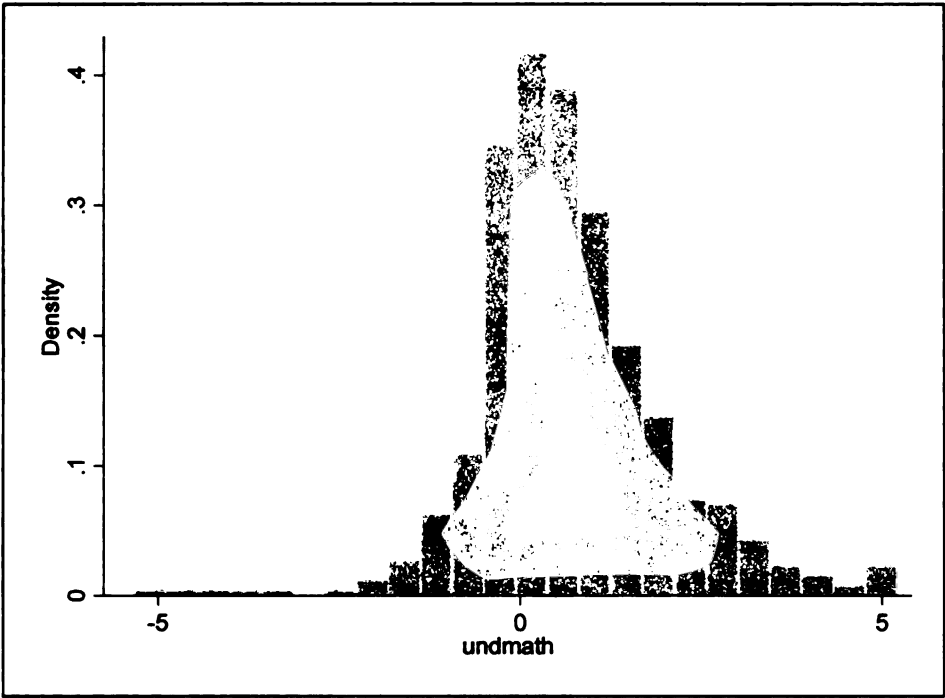


Figure 4.2.1: Original formula

Figure 4.2 (cont'd)

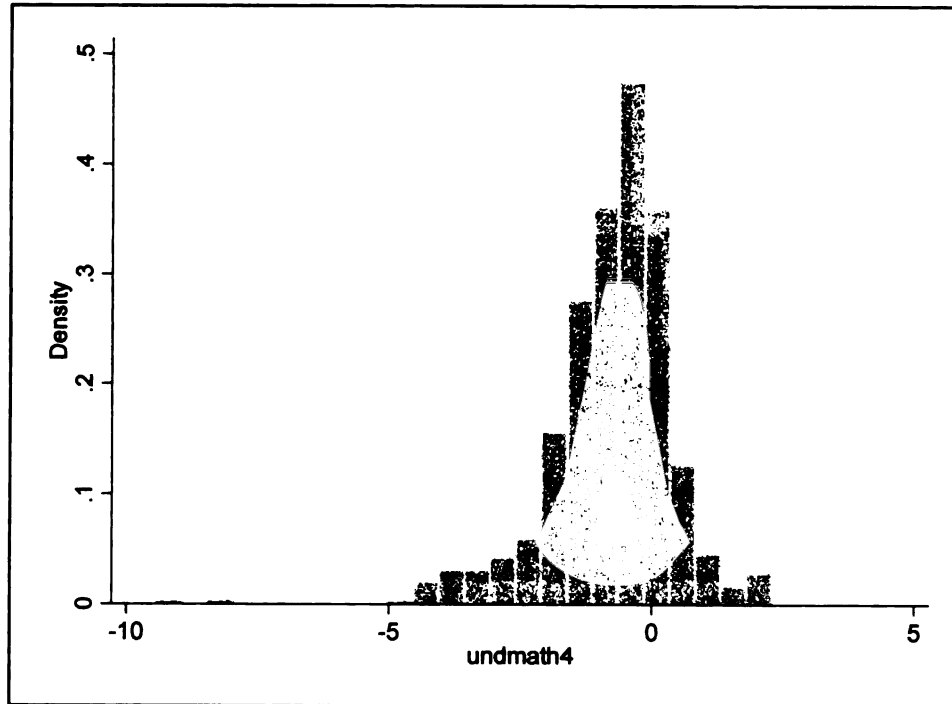


Figure 4.2.2: Smoothing constant 0.4

Figure 4.2 (cont'd)

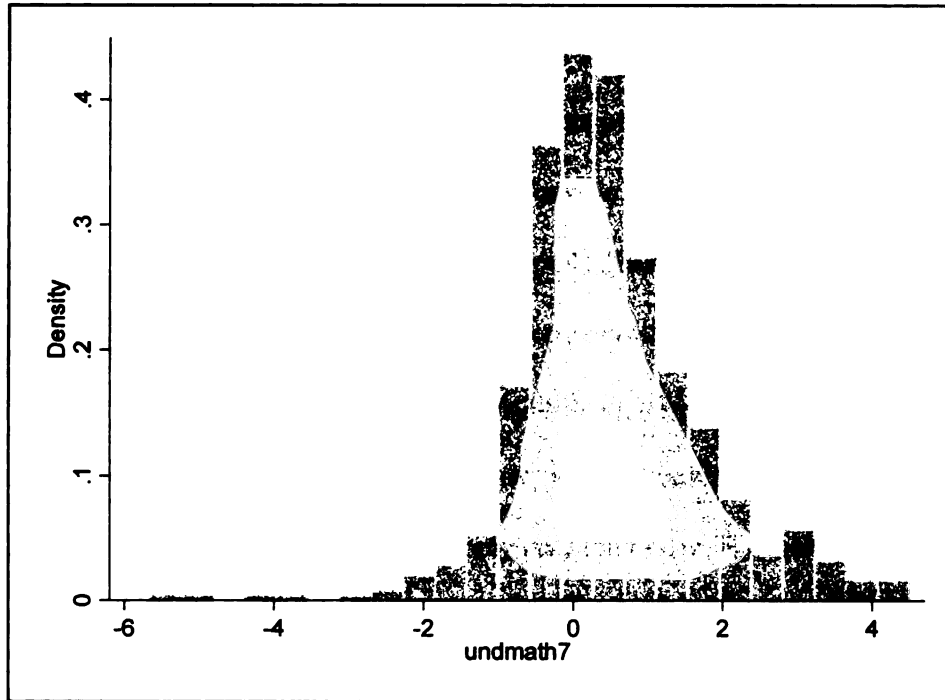


Figure 4.2.3: Smoothing constant 0.7

Here, there are clear differences in the distributions of undersupply. The original formula produces an undersupply distribution that is weighted toward positive numbers, where positive numbers indicate greater amounts of undersupply. The 0.7 smoothing constant enrollment estimates produce a distribution shaped like that of the original enrollment estimates, but one that is less spread out and more peaked. Finally, the 0.4 smoothing constant produces a markedly different distribution, one that is more truncated and with more of the distribution in the negative numbers, which suggests adequate supply. This reflects the implications of the demand estimates—when the demand estimates have longer tails, there are more schools identified as undersupplied.

Which enrollment estimate is the most appropriate to use for this analysis? Using a smoothing constant to generate a projected enrollment is conceptually an important

modification, given Michigan's shifting enrollments and declining population. Which smoothing constant is most appropriate? Based on the need to appropriately identify undersupply while taking changes in enrollments into account, the 0.7 smoothing constant will be used for enrollment projections, as the use of this constant causes the weight of earlier observations to decrease rapidly (Hussar & Bailey, 2008). This is more appropriate for the situation in Michigan, where enrollments can decline quickly in a given school. Additionally, NCES suggests utilizing a higher smoothing constant when the data are population estimates as opposed to sample estimates. The distributions for ELA and science demand and undersupply are included in Technical Appendix B.

Part II: Distributional assumptions for class size and courses per FTE: Addressing the issues with ratios

The demand formula is a ratio—the ratio of enrollment to class size, and then that entire quantity as a ratio to periods taught per FTE. However, the denominators—class size and courses taught per FTE—are assumed values, and therefore little is known about their possible range of values.

In a ratio $Z=Y/X$, Y and X are measured with some error (indicated by their standard deviations). The ratio (Z) will have a standard deviation, and the size of that standard deviation (and thus, the precision with which that ratio is measured) will depend on several factors, including; 1) the precision with which X and Y are measured, and 2) the correlation between X and Y . If X and Y are measured with great precision, then the difference between the $E(Z)$, or the expectation of Z , and actual Z is small. However, if X and Y are not measured with precision, the difference between the ratio and the expected

value is large. Furthermore, ratios are more variable when the denominator is small, and correlation between X and Y decreases the variance of the ratio (Rice, 2007).

In this case, the numerator (Y) is the enrollment. The denominator, X, is class size, and the number of courses taught per FTE. Both class size and courses per FTE are assumed, with class size assumed at 25 and courses per FTE at 5. As assumptions, they have no standard deviation. In practice, however, there is a distribution around class size and number of courses taught. The wider this distribution (i.e. the less precisely they are measured), the more variable the demand calculation will be.

Given that they are assumptions, and that they are designed to be adjusted by the end user to reflect changing conditions or to test the demand under planned circumstances, assuming a distribution is somewhat of an academic exercise. However, when estimating this for every school in the whole state, it is important to quantify how variable the estimates of undersupply could be with a range of values for the assumed values. This is akin to a sensitivity analysis—quantifying how sensitive the demand formula is to changes in the assumptions, and how sensitive this estimate would have to be in order to invalidate inferences around undersupply. This is evaluated empirically, as described below.

In order to quantify the sensitivity of the ratio to changes in the denominator, I generated two new variables for class size; one with a mean of 25 and a standard deviation of 1 [$\sim N(25, 1)$], and one with a mean of 25 and a standard deviation of .5 [$\sim N(25, .5)$]. I also generated a variable for courses taught per FTE, with a mean of 5 and a standard deviation of .25 [$\sim N(5, .25)$]. This was because the range of the number of courses taught is bounded by realities of a teaching day—no teacher is teaching 8 courses

in a day, or as few as two. A standard deviation of .25 produces a lower bound of 4.28 courses and an upper bound of 5.9 courses.

For mathematics only (again, for parsimony, with ELA and science included in the Technical Appendix C), Figure 4.3 presents the distributions of **demand** compared under four situations: (1) Original demand formula, (2) enrollment with .7 smoothing constant (selected as the optimal enrollment estimate from above analyses), (3) enrollment with .7 smoothing, class size \sim N(25, 1) and courses \sim N(5, .25), and (4) enrollment with .7 smoothing, class size \sim N(25, .5) and courses \sim N(5, .25)

Figure 4.3: Comparison of Demand Estimates Under Four Estimation Scenarios

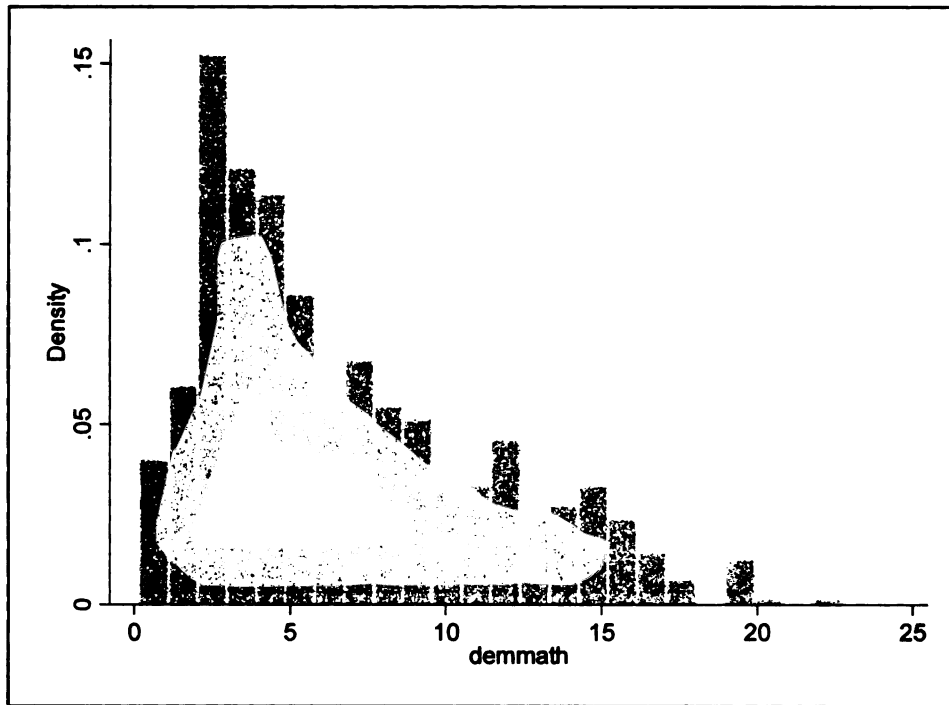


Figure 4.3.1: Scenario 1: Original formula

Figure 4.3. (cont'd)

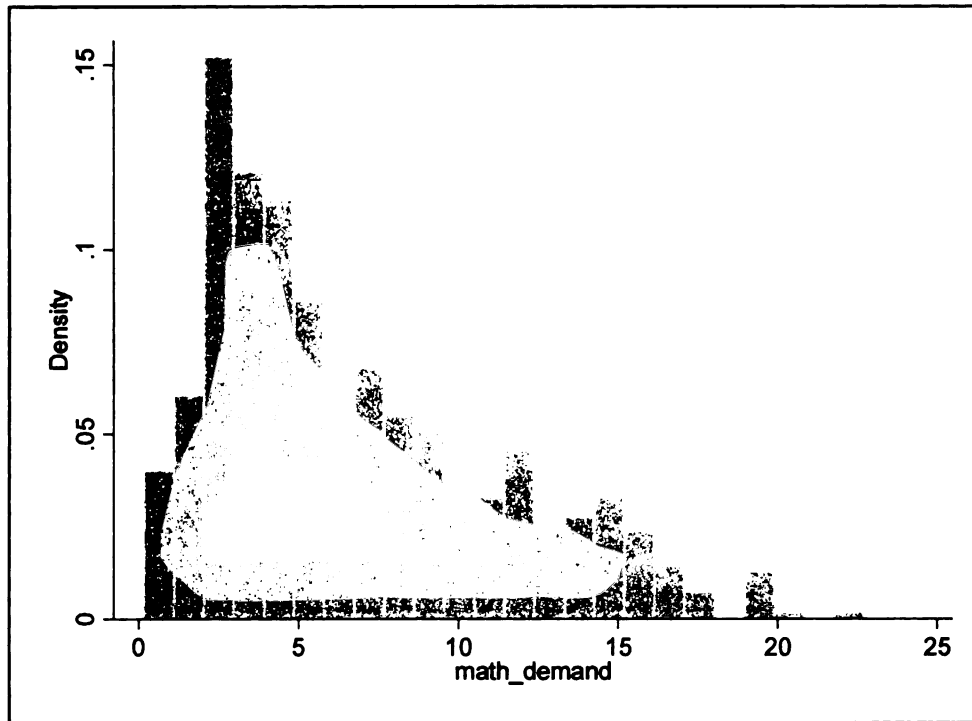


Figure 4.3.2: Scenario 2: Enrollment 0.7 smoothing constant

Figure 4.3 (cont'd)

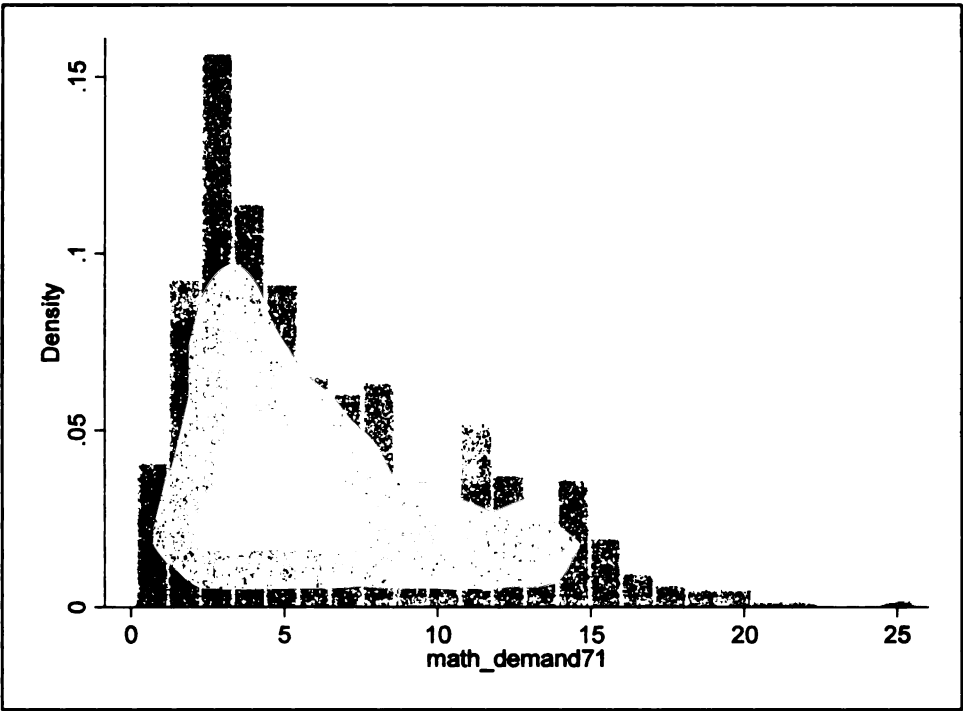


Figure 4.3.3: Scenario 3: 0.7 smooth, class size~N(25, 1), courses~N(5, .25)

Figure 4.3 (con'td)

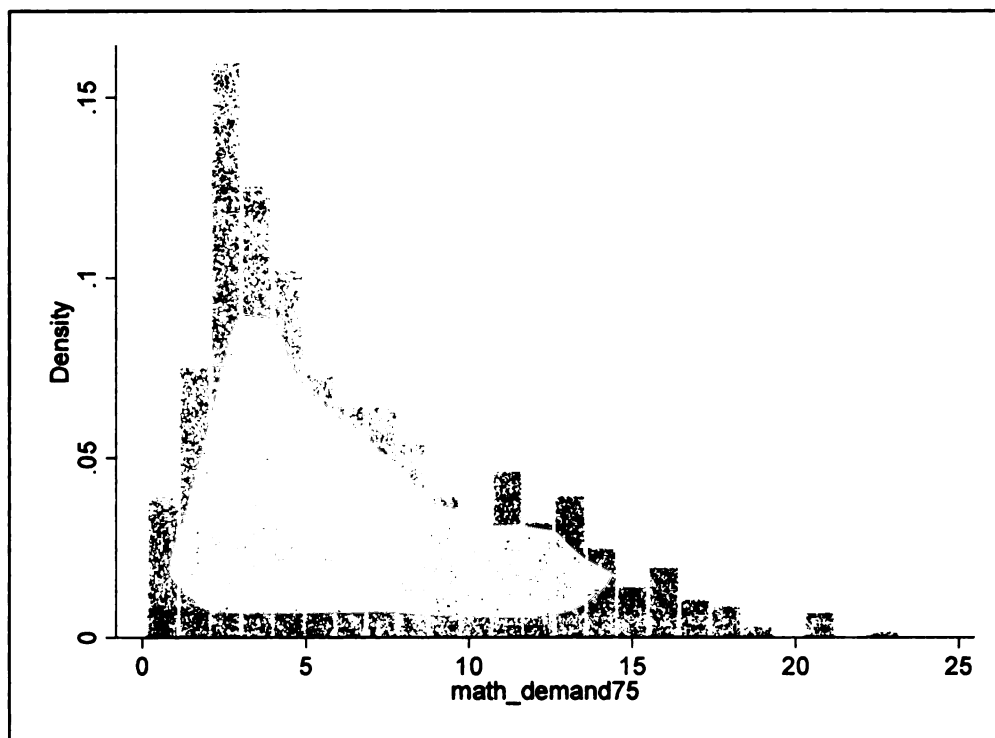


Figure 4.3.4: Scenario 4: 0.7 smooth, class size~N(25, .5) courses~N(5, .25)

Observing the distributions, while changing enrollment estimates from the former (simple average) to a weighted average using a smoothing constant does not appear to change the distributions of demand greatly, there do appear to be changes in the distribution in situation (3) and (4). The tail of the distribution is more elongated and more positively skewed, which suggests a greater amount of observed demand when including distributional assumptions around the class size and number of courses taught per FTE.

Figure 4.4: Comparison of Undersupply Under Four Estimation Scenarios

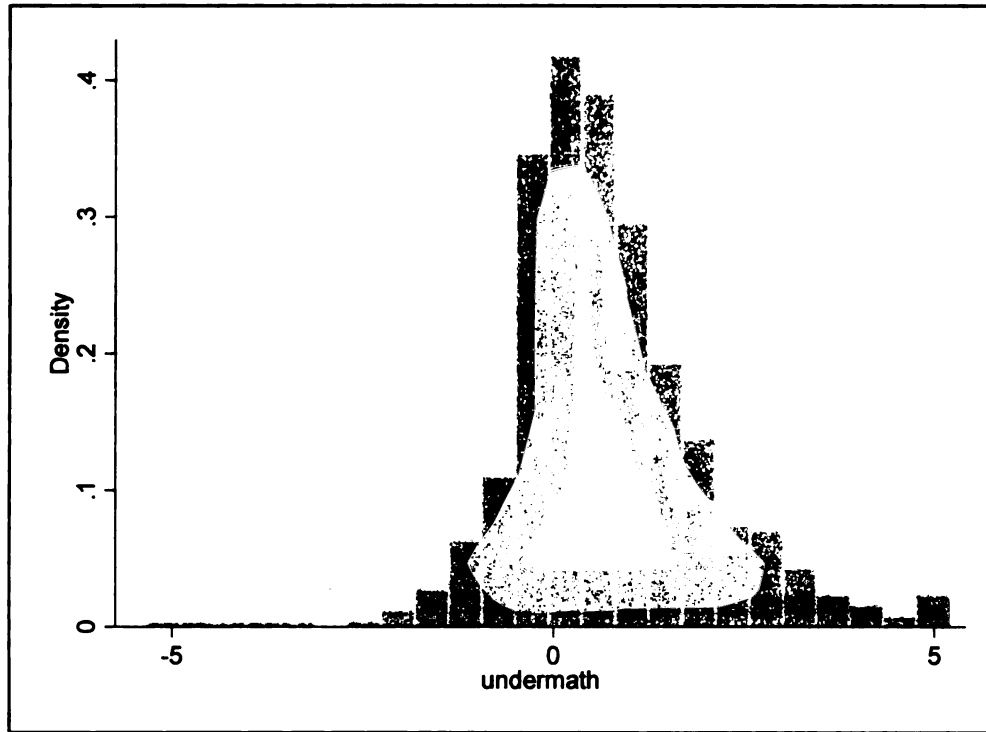


Figure 4.4.1: Scenario 1: Original formula

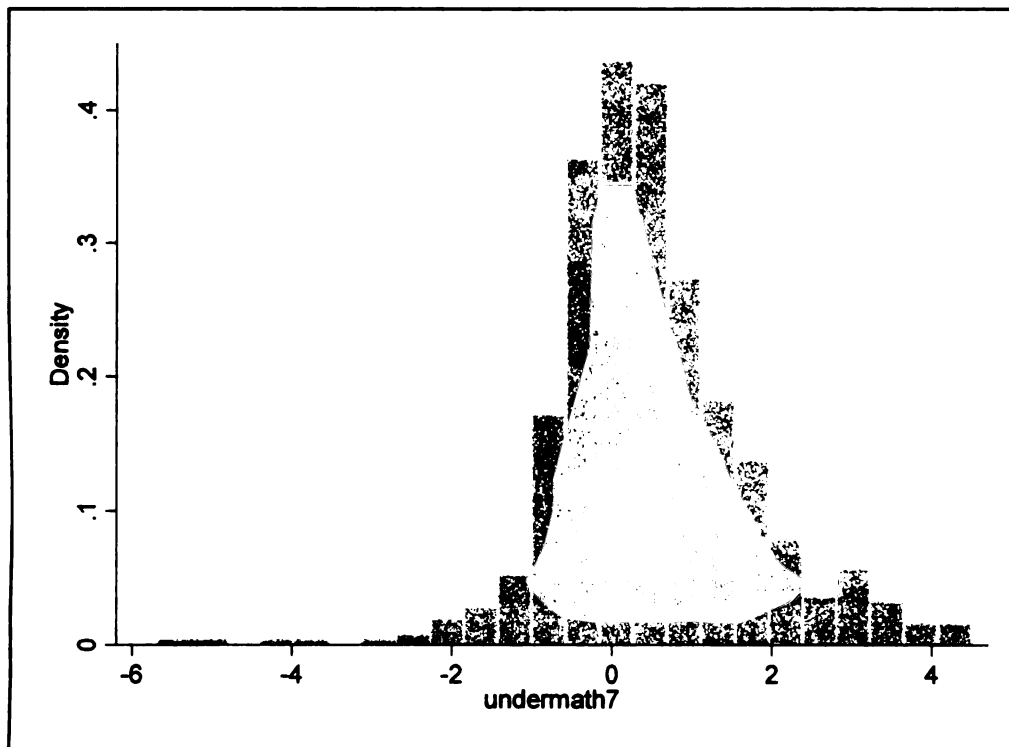


Figure 4.4.2: Scenario 2: Enrollment 0.7 smoothing constant

Figure 4.4 (cont'd)

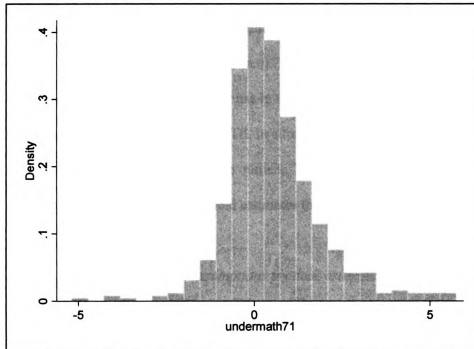


Figure 4.4.3: Scenario 3: 0.7 smooth, class size~ $N(25, 1)$, courses~ $N(5, .25)$

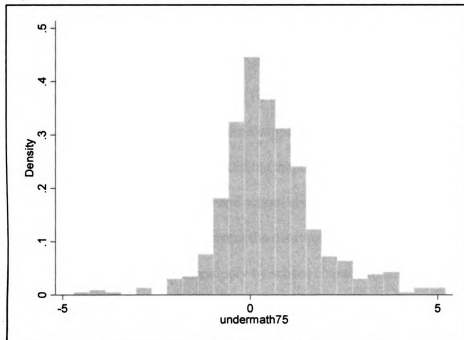


Figure 4.4.4: Scenario 4: 0.7 smooth, class size~ $N(25, .5)$ courses~ $N(5, .25)$

Figure 4.4 shows the distribution of undersupply under the four scenarios. As mentioned above, the undersupply distributions are more conservative than those in the original demand formula when using the 0.7 smoothing constant, which is why this enrollment estimate was selected for use in the final analyses. When using distributional assumptions around class size and FTE, in situation (3) and (4), the distributions become more peaked, but there are also more outliers. This reflects the fact that these small changes in the denominator can yield estimates that are more varied than those generated when using only assumed values.

Do different demand formula assumptions produce estimates that vary by school characteristics?

The analyses above demonstrate the shape and characteristics of the distributions of demand and undersupply under differing sets of assumptions. The second important question is: do the demand estimates lead to an identification of undersupplied schools that varies with school characteristics? In other words, is there differential formula functioning by school characteristics? To investigate this, the mean mathematics, English language arts, and science undersupply was calculated by a set of school characteristics. This table is introduced in Chapter 3, to demonstrate the descriptive relationship between undersupply and school characteristics.

Tables 4.1, 4.2, and 4.3 present the distribution of mean mathematics, English language arts, and science undersupply by school characteristics under three different demand formula assumptions: 1) the original demand formula, 2) using a 0.7 smoothing constant on the enrollment, and 3) a 0.7 smoothing constant on the enrollment and distributional assumptions on class size [$(\sim N(25, .5))$] and courses per FTE [$(\sim N(5, .25))$].

If the distribution of mean levels of undersupply by school characteristics varies when using different demand specifications, this is evidence that certain formulas function differentially.

The comparative analyses presented in Tables 4.1, 4.2 and 4.3 demonstrate that, in each of the demand formula specifications, the relationships between undersupply and different school characteristics do not differ. The magnitude of the relationships is different, reflecting the fact that the original demand formula produced more generous estimates of undersupply, but the order of relationships is the same. For example, in Table 4.1, the distribution of mean mathematics undersupply over the four locale types using a demand formula with a smoothing constant on enrollment is .64 for city schools, .81 for suburban schools, .51 for town schools and .18 for rural schools. Using the original formula, the distribution is .98 for city schools, 1.07 for suburban schools, .74 for town schools, and .32 for rural schools—the same ordering as under the first formula. Finally, in the most complex formula, the one with a smoothing constant on the enrollment and distributional assumptions on class size and FTE, the distribution of math undersupply by locales is .69 for city schools, .90 for suburban schools, .52 for town schools, and .19 for rural schools. In all three situations, rural schools have the lowest mean mathematics undersupply, while suburban schools have the highest mean mathematics undersupply. This is true for each subject and for each set of characteristics, and this provides strong evidence to conclude that, regardless of demand formula specification, the formulas will work similarly for all school types.

How does the categorization of undersupply differ in each of the four situations?

A second critical question, and possibly one that is more policy-relevant, is the extent to which determinations of significant undersupply (i.e. more than one FTE needed) change based on estimations. Although the continuous undersupply calculations are helpful in understanding changes in teacher supply more broadly, schools, districts, and states will likely want to know, “Do I have a large enough undersupply that I am going to potentially have a staffing problem?” Therefore, comparing how schools are classified as undersupplied or not undersupplied under each formula is important. In Table 4.4, the categorization of undersupply are compared in each of the situations.

The classification results show that all modifications to the original formula lead to fewer schools being classified as significantly undersupplied. This suggests that it is important to use these types of modifications because this helps keep this demand formula conservative in terms of the estimates it produces. One reason for the importance of using these types of modification is the fact that many schools have declining enrollments. With a three year average, unweighted, each year contributes the same to the final enrollment estimate. Using a smoothing constant, older observations count for less than newer observations. Therefore, predicted enrollments under the smoothing constant are lower than average three year enrollments; which in turn produces lower demand estimates. Estimates under the two distributions assumptions of the standard deviation of class size are very similar. For this reason, Situation (2), which is the enrollment projection method only, seems to be an optimal method.

Table 4.5 presents the cross-classifications of undersupply between Scenario 2 and the other scenarios, in order to understand variations in how schools are classified. There were 34 schools classified as undersupplied in the original formula who were not

when using the estimated enrollments in Situation 2 (see A in Table 4.2). Are those schools likely to experience an undersupply, or were they falsely tagged by the original estimate that did not take into account potential declining enrollments? Similarly, there were 34 schools were not undersupplied under situation (2) but were undersupplied under situation (3) and 16 schools were undersupplied in (2) but not in (3) (see B in Table 4.2). This means that assuming a distribution around class size of $N(25, 1)$ leads to 50 schools being classified differently. For situations 3 and 4, the number is of schools classified differently is 47, and they are arrayed similarly to the situations in 2 and 3 (see C in Table 4.3). This suggests that if a distribution is going to be assumed around class size and FTEs, it is better to use the more conservative estimate of class size $\sim N(25, .5)$

Finally, for A and C, I identified the schools who were “misclassified” (i.e. those who were classified as undersupplied on one analysis and not the other, and vice versa) and did simple cross tabulations with chi-square tests for significant differences by four key structural categories of schools: locale, percent minority, percent free and reduced lunch, and school size. The purpose of this is to ascertain whether or not the misclassifications are more or less likely to happen to certain types of schools. Table 4.6 shows the crosstabulations for A (Situation 1 and Situation 2) and Table 4.7 shows them for C (Situation 4 and Situation 2).

The difference between these two demand estimates was the inclusion of the 0.7 smoothing constant on the enrollment projections. This led to more conservative estimates of demand, and fewer schools being identified as significantly undersupplied. There are no statistically significant differences between those categorized differently by

the two formulas and those categorized the same on locale, school minority composition, or free and reduced lunch

Small schools were significantly less likely to be classified differently under the two demand formulas than large schools. This suggests that larger schools are more susceptible to differences in how the demand formula is specified. When the enrollment was projected using the 0.7 smoothing constant (i.e. in Situation 2) rather than averaged over three years in the original formula (Situation 1), these schools had lower enrollments and thus less undersupply. This is likely a positive modification, as the previous enrollment estimates were over-estimating the number of students in the school, and thus overestimating demand.

Table 4.7 presents the crosstabulations for C (from Table 4.5), looking at misclassifications under Situation 2 (0.7 smoothing constant on enrollment) and Situation 4 (smoothing constant and distributional assumptions on class size and FTE). The assumed distribution around class size and course per FTE is the more conservative estimate than that presented in Situation 3. Again, as seen in Table 4.6, there is no statistically significant differences between schools classified the same way under each formula and those classified differently on locale, percent minority, and percent free or reduced lunch. Fewer small schools were misclassified than expected, and more large schools were misclassified.

To summarize, there is not evidence to suggest that any of these specifications of demand function differently based on school locale, percent minority students, or percent free and reduced lunch students. As seen in the comparison of mean undersupply by school characteristics, the distributions are the same, although the absolute quantity in

each category differs. Even when using the “significant” threshold, this formula is not functioning differentially.

The original formula, without the enrollment projection methodology, was overestimating undersupply for large schools. This provides firm evidence to suggest that using a better enrollment measure was a positive improvement for the formula.

Conclusions

For the enrollment estimates, using a weighted average seems to be critical in order to keep the analysis on the conservative side and avoid Type I errors. If this formula is being utilized by a district or school individually, they can decide whether to use an average, a weighted average, or their own prediction of what they think their enrollment is going to be, based on their own experiences.

The distribution of undersupply over types of schools does not differ regardless of the demand formula specification. Changes in classification of significant undersupply are largely due to the shape of the distributions, and the fact that imposing a 1.0 FTE cut point as the threshold for “significant” undersupply places that cut point at different locations on the varying distributions. In the original formula, more schools fell above that cut point, because the formula was more generous, and also, because enrollment was likely overestimated by using a simple average, particularly in large schools. Therefore schools were classified as undersupplied who may not actually have experienced undersupply.⁶

⁶ It is possible to verify the accuracy of this formula by using the next year’s data (in this case, REP 2009) and assessing the extent to which the schools predicted to be undersupplied are actually undersupplied. This was not done for this analysis as the focus was on improving the demand formula as an estimation strategy, but will be done in the future as a final verification step, when 2009 data is made available by MDE.

Assuming a distribution around class size and courses is problematic. It is instructive to know the degree to which these assumptions can change classification errors. However, for use in the formula and analyses, using the assumed distributions does not make sense, as the point of the formula is to estimate which schools MAY be undersupplied under certain conditions. The goal of the formula is important here. It is true that changes in these distributional assumptions could change inferences, as schools are categorized differently under different situations. However, since they are not empirically based, I hesitate to continue to use them in analysis.

Given this evidence, in future analyses, the smoothing constant of 0.7 will be used with no distributional assumptions on class size and courses per FTE as the main analysis. For comparison, the final undersupply models will be run using the smoothing constant and distributional analyses on class size $\sim N(25, .5)$ and courses per FTE $\sim N(5, .25)$. This will show how the results may change in a situation where the assumptions are not fixed, but with conservative estimates of distribution.

A final question related to this formula is the extent to which it is valid. Measurement literature outlines various dimensions of validity (content, criterion, construct) but the core concern in a formula of this nature is construct validity—the extent to which this formula is measuring what it is intended to measure (teacher demand)—and face validity, the “believability” of this formula. Validity, even in the pure measurement sense, does not have one “test” that establishes the validity of an instrument or a construct. Rather, validity is demonstrated via the accumulation of evidence.

Evidence for the validity of this formula and method includes the computations and comparisons undertaken in this analysis above, particularly the comparisons of the

distribution of estimates obtained under each method. Face validity is also established by the extent to which this formula is seen as viable and believable to a policymaker audience. As the original formula has already been vetted by the state of Michigan, the Institute for Education Sciences, the Regional Educational Laboratory-Midwest, representatives of other regional educational laboratories, and by a number of academic audiences, there is substantial evidence that this formula passes the face validity test. The rigor to which the formula is submitted here suggests that it has construct validity as well.

The purpose of this formula is to provide a useful tool to practitioners and researchers alike to allow for demand calculations that are sensitive to school-specific variations in demand and supply; that takes into account the context of curricular requirements; and that can be used as a planning tool by practitioners in which they can adjust the assumptions to reflect the current or future conditions in their school in order to inform decision making.

Table 4.1: Mean Mathematics Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions

	<i>Enrollment with Smoothing**</i>			<i>Original Demand Formula</i>			<i>Enrollment w/Smooth & Dist'n on Class</i>			
	<i>Mean</i>	<i>SD</i>	<i>F-test</i>	<i>Mean</i>	<i>SD</i>	<i>F-test</i>	<i>Mean</i>	<i>SD</i>	<i>F-test</i>	<i>N</i>
Locale										
City	.64	1.68	.000	0.98	1.87	.000	.69	1.80	.000	94
Suburb	.83	1.37		1.07	1.38		.90	1.46		161
Town	.51	.90		0.74	0.99		.52	1.01		78
Rural	.18	.84		0.32	0.85		.19	.81		245
Percent minority										
Less than 5%	.31	.81	.168	0.46	0.87	.038	.29	.86	.039	174
5-10%	.46	1.12		0.69	1.17		.48	1.10		141
10-15%	.64	1.13		0.82	1.2		.71	1.20		56
15-65%	.58	1.42		0.82	1.49		.66	1.50		132
Greater than 65%	.63	1.65		0.92	1.8		.70	1.84		75
Percent free/reduced lunch										
Less than 10%	.04	1.60	.006	0.24	1.61	.008	-.003	1.51	.003	43
10-29%	.67	1.13		0.88	1.2		.73	1.16		202
30-49%	.37	1.02		0.56	1.08		.40	1.09		204
50-69%	.62	1.44		0.83	1.62		.63	1.64		94
70% or greater	.29	1.18		0.5	1.19		.30	1.25		35
School size										
Less than 300 students	-0.15	.73	.000	-0.08	0.7	.000	-.13	.75	.000	83
300-999 students	.30	.91		0.45	0.92		.30	.95		307
Greater than 1000 students	1.06	1.52		1.42	1.6		1.15	1.62		188
Charter Schools										
Charter	.06	1.08	.039	0.73	1.28	.003	.54	1.28	.054	33
Non-charter	.51	1.21		0.05	1.06		.10	1.13		545
Magnet Schools										
Magnet	.22	1.38	.061	0.72	1.27	.122	.57	1.26	.004	65
Non-magnet	.52	1.12		0.46	1.45		.09	1.28		513

Table 4.1 (cont'd)										
	Enrollment with Smoothing**			Original Demand Formula			Enrollment w/Smooth & Dist'n on Class Size and FTE			
	Mean	SD	F-test	Mean	SD	F-test	Mean	SD	F-test	N
Percent teachers with prof licenses										
Less than 80%	.35	1.16	.367	0.5	1.26	.135	.39	1.29	.468	132
80-90%	.52	1.31		0.76	1.38		.55	1.35		297
Greater than 90%	.42	1.00		0.74	1.07		.55	1.08		149
Percent minority teachers										
Less than 1%	.28	.98	.000	0.44	1.01	.000	.28	.98	.001	275
1-10%	.72	1.22		0.96	1.29		.77	1.29		214
Greater than 10%	.54	1.63		0.83	1.79		.61	1.81		89
Percent highly qualified teachers										
Less than 75%	.74	1.34	.000	1	1.45	.000	.75	1.44	.003	196
75-85%	.46	1.23		0.68	1.26		.49	1.22		144
Greater than 85%	.28	1.02		0.44	1.08		.33	1.13		238
Total	.48	1.20		0.69	1.28		.578	.51	1.27	578

Math undersupply is calculated for the SY 2009 based on 2008 REP and student enrollment

******This calculation is identified as the optimal calculation and is used primarily in the Chapter 7 undersupply analyses

Table 4.2: Mean English Language Arts Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions

	Enrollment with Smoothing**			Original Formula			Enrollment w/Smooth & Dist'n on Class Size/ FTE			N
	Mean	SD	F-test	Mean	SD	F-test	Mean	SD	F-test	
Locale										
City	-.81	2.56	.003	-0.48	2.69	0.028	-0.76	2.64	0.012	94
Suburb	-1.10	1.81		-0.86	1.83		-1.03	1.85		161
Town	-.89	1.06		-0.68	1.05		-0.9	1.04		78
Rural	-.52	.88		-0.38	0.89		-0.51	0.91		245
Percent minority										
Less than 5%	-.61	1.01	.062	-0.46	1.01	0.083	-0.63	1.07	0.181	174
5-10%	-.69	1.15		-0.46	1.12		-0.67	1.12		141
10-15%	-1.17	1.79		-0.99	1.86		-1.1	1.79		56
15-65%	-1.00	1.67		-0.77	1.66		-0.93	1.71		132
Greater than 65%	-.67	2.64		-0.38	2.81		-0.6	2.74		75
Percent free/reduced lunch										
Less than 10%	-1.43	1.93	.037	-1.23	1.91	0.036	-1.47	1.96	0.024	43
10-29%	-.78	1.38		-0.56	1.38		-0.72	1.36		202
30-49%	-.79	1.31		-0.59	1.29		-0.76	1.33		204
50-69%	-.58	2.10		-0.36	2.29		-0.56	2.21		94
70% or greater	-.46	1.87		-0.24	1.89		-0.45	1.92		35
School size										
Less than 300 students	-.43	.70	.006	-0.37	0.69	0.301	-0.41	0.7	0.027	83
300-999 students	-.71	1.36		-0.55	1.38		-0.7	1.38		307
Greater than 1000 students	-1.05	2.09		-0.69	2.17		-0.97	2.15		188
Charter Schools										
Charter	-.86	2.81	.752	-0.55	1.52	0.268	-0.74	1.52	0.768	545
Non-charter	-.77	1.48		-0.87	2.81		-0.83	2.83		33
Magnet Schools										
Magnet	-.86	2.27	.658	-0.56	1.5	0.795	-0.72	1.52	0.198	513
Non-magnet	-.77	1.48		-0.62	2.33		-0.99	2.24		65

Table 4.2 (cont'd)									
	Enrollment with Smoothing**			Original Formula			Enrollment w/Smooth & Dist'n on Class Size/ FTE		
	<i>Mean</i>	<i>SD</i>	<i>F-test</i>	<i>Mean</i>	<i>SD</i>	<i>F-test</i>	<i>Mean</i>	<i>SD</i>	<i>F-test</i>
Percent teachers with prof licenses									
Less than 80%	-.58	1.91	.130	-0.43	2	0.312	-0.53	1.95	0.109
80-90%	-.90	1.59		-0.67	1.62		-0.88	1.63	
Greater than 90%	-.72	1.18		-0.5	1.17		-0.68	1.18	
Percent minority teachers									
Less than 1%	-.58	1.04	.008	-0.42	1.03	0.042	-0.58	1.02	0.026
1-10%	-1.03	1.57		-0.79	1.57		-0.98	1.6	
Greater than 10%	-.79	2.62		-0.5	2.77		-0.72	2.74	
Percent highly qualified teachers									
Less than 75%	-.59	1.59	.086	-0.33	1.68	0.02	-0.58	1.62	0.164
75-85%	-.79	1.41		-0.57	1.38		-0.75	1.38	
Greater than 85%	-.93	1.67		-0.77	1.68		-0.88	1.73	
<i>Total</i>	-.78	1.58		-0.6	1.62		-0.75	1.62	

ELA undersupply is calculated for the 2009 school year based on 2008 REP and student

****This calculation is identified as the optimal calculation and is used primarily in the Chapter 7 undersupply analyses**

Table 4.3: Mean Science Teacher Undersupply by School Characteristics Under Different Demand Formula Assumptions

	Enrollment with Smoothing**			Original Demand Formula			Enrollment w/Smoothing & Dist'n on Class Size/ FTE			N
	Mean	SD	F-test	Mean	SD	F-test	Mean	SD	F-test	
Locale										
City	-1.04	1.82	.000	-.79	1.86	.000	-1.00	1.86	.000	94
Suburb	-1.07	1.36		-.88	1.33		-1.01	1.41		161
Town	-.68	.89		-.52	.90		-.69	.93		78
Rural	-.50	.71		-.39	.72		-.49	.73		245
Percent minority										
Less than 5%	-.47	.75	.000	-.37	.76	.000	-.49	.76	.000	174
5-10%	-.73	.93		-.56	.91		-.71	.98		141
10-15%	-.99	1.02		-.85	1.05		-.94	1.20		56
15-65%	-1.21	1.59		-1.03	1.56		-1.15	1.59		132
Greater than 65%	-.58	1.54		-.37	1.56		-.53	1.59		75
Percent free/reduced lunch										
Less than 10%	-1.96	1.91	.000	-1.81	1.87	.000	-1.99	1.95	.000	43
10-29%	-.90	1.11		-.74	1.09		-.86	1.13		202
30-49%	-.56	.89		-.41	.86		-.54	.90		204
50-69%	-.53	1.23		-.37	1.27		-.51	1.30		94
70% or greater	-.39	1.16		-.23	1.19		-.38	1.14		35
School size										
Less than 300 students	-.46	.64	.000	-.41	.62	.000	-.45	.63	.000	83
300-999 students	-.50	.88		-.39	.88		-.50	.92		307
Greater than 1000 students	-1.34	1.58		-1.06	1.63		-1.27	1.64		188
Charter Schools										
Charter	-.25	1.09	.010	-.62	1.20	.077	-.78	1.23	.012	33
Non-charter	-.80	1.20		-.25	1.07		-.22	1.09		545
Magnet Schools										
Magnet	-.74	1.13	.819	-.62	1.20	.645	-.73	1.23	.536	65
Non-magnet	-.77	1.21		-.55	1.12		-.83	1.18		513

Table 4.3 (cont'd)										
	Enrollment with Smoothing**			Original Demand Formula			Enrollment w/Smoothing & Dist'n on Class Size/ FTE			
	Mean	SD	F-test	Mean	SD	F-test	Mean	SD	F-test	N
Less than 80%	-.50	1.13	.009	-.39	1.12	.038	-.47	1.13	.008	132
80-90%	-.88	1.28		-.71	.127		-.87	1.32		297
Greater than 90%	-.77	1.06		-.60	1.06		-.74	1.07		149
Percent minority teachers										
Less than 1%	-.54	.80	.000	-.42	.81	.000	-.54	.81	.001	275
1-10%	-1.06	1.40		-.88	1.38		-1.02	1.44		214
Greater than 10%	-.77	1.53		-.56	1.56		-.72	1.58		89
Percent highly qualified teachers										
Less than 75%	-.80	1.34	.756	-.61	1.35	.971	-.79	1.39	.631	196
75-85%	-.80	1.31		-.63	1.27		-.77	1.30		144
Greater than 85%	-.72	.99		-.60	.99		-.69	1.02		238
Total	-.77	1.19		-.61	1.19		-.74	1.23		578

Science undersupply is calculated for SY 2009 based on 2008 REP and student enrollment from

****This calculation is identified as the optimal calculation and is used primarily in the Chapter 7 undersupply analyses**

Table 4.4: Categorization of Undersupply Under Four Sets of Assumptions

	<i>Situation 1</i>	<i>Situation 2</i>	<i>Situation 3</i>	<i>Situation 4</i>
<i>Not Undersupplied</i>	393	426	408	409
<i>Undersupplied</i>	185	152	170	169

Situation 1: Original demand formula

Situation 2: 0.7 Smoothing constant only

Situation 3: 0.7 smoothing constant, class size~N(25, 1) and courses~N(5, .25)

Situation 4: 0.7 smoothing, class size~N(25, .5) and courses~N(5, .25)

Table 4.5: Comparison of Cross-Categorizations of Undersupply Under Four Sets of Assumptions

		<i>Situation 2 US</i>	<i>Situation 2 Not US</i>	<i>Total</i>
A	<i>Situation 1: Undersupplied</i>	151	34	185
	<i>Situation 1: Not US</i>	1	392	393
	<i>Total</i>	152	426	578
		<i>Situation 2 US</i>	<i>Situation 2 Not US</i>	<i>Total</i>
B	<i>Situation 3: Undersupplied</i>	136	34	170
	<i>Situation 3: Not US</i>	16	392	408
	<i>Total</i>	152	426	578
		<i>Situation 2 US</i>	<i>Situation 2 Not US</i>	<i>Total</i>
C	<i>Situation 4: Undersupplied</i>	137	32	169
	<i>Situation 4: Not US</i>	15	394	409
	<i>Total</i>	152	426	578

Situation 1: Original demand formula

Situation 2: 0.7 Smoothing constant only

Situation 3: 0.7 smoothing constant, class size~N(25, 1) and courses~N(5, .25)

Situation 4: 0.7 smoothing, class size~N(25, .5) and courses~N(5, .25)

**Table 4.6: "Misclassification" Between Situation 1 and Situation 2
("A" from Table 4.4)**

	<i>Same Classification</i>	<i>Different Classification</i>	<i>Expected Dist'n</i>	<i>Chi-Square</i>
Locale				
City	17%	14%	17%	0.439
Suburb	27%	37%	28%	
Town	13%	17%	13%	
Rural	43%	31%	42%	
Percent minority				
Less than 5%	30%	29%	30%	.415
5-10%	24%	29%	24%	
10-15%	10%	3%	10%	
15-65%	22%	31%	23%	
Greater than 65%	14%	9%	13%	
Percent free/reduced lunch				
Less than 10%	8%	3%	7%	.281
10-29%	35%	37%	35%	
30-49%	34%	49%	35%	
50-69%	17%	9%	17%	
70% or greater	6%	3%	6%	
School size				
Less than 300 students	15%	3%	15%	.008
300-999 students	54%	43%	53%	
Greater than 1000 students	31%	54%	32%	

Situation 1: Original demand formula

Situation 2: 0.7 Smoothing constant only

**Table 4.7: "Misclassification" Between Situation 2 and Situation 4
("C" from Table 4.4)**

	<i>Same Classification</i>	<i>Different Classification</i>	<i>Expected Dist'n</i>	<i>Chi- Square</i>
Locale				
City	17%	9%	17%	0.173
Suburb	27%	36%	28%	
Town	13%	19%	13%	
Rural	43%	36%	42%	
Percent minority				
Less than 5%	30%	28%	30%	.161
5-10%	24%	30%	24%	
10-15%	9%	13%	10%	
15-65%	22%	28%	23%	
Greater than 65%	14%	2%	13%	
Percent free/reduced lunch				
Less than 10%	7%	15%	7%	.149
10-29%	35%	32%	35%	
30-49%	35%	40%	35%	
50-69%	17%	11%	17%	
70% or greater	6%	2%	6%	
School size				
Less than 300 students	16%	0%	15%	.005
300-999 students	53%	53%	53%	
Greater than 1000 students	31%	47%	32%	

Situation 2: 0.7 Smoothing constant only

Situation 4: 0.7 smoothing, class size~N(25, .5) and courses~N(5, .25)

CHAPTER 5: TEACHER CHURN RATE: THE RELATIONSHIP BETWEEN TEACHER RETENTION AND STUDENT ACHIEVEMENT

Introduction

High schools face the challenge of recruiting and retaining an adequate number of qualified and effective instructional staff. An elevated rate of teacher turnover within high schools may contribute to a disorganized school culture that is lacking in a sense of community, which in turn may have negative implications for student achievement and school effectiveness. Prior research has provided insights regarding the types of schools that are likely to have higher rates of teacher turnover, but there is little research regarding the *impact* of high levels of teacher turnover on student achievement outcomes (Carroll, Reichardt, & Guarino, 2000; Guarino et al., 2006; Hanushek, Kain, & Rivkin, 2004; Ingersoll, 2001; Lankford, Loeb, & Wyckoff, 2002; Shen, 1997; Smith & Ingersoll, 2004; Stockard & Lehman, 2004; Whitener et al., 1998). The purpose of this study is to investigate the relationship between high school teacher “churn rates,” as measured by teacher retention, or the average proportion of teachers who remain the same in a given school over a four year time period, and student achievement outcomes, as measured by mathematics achievement test scores on the state assessments.¹ This study also tests the relationship between teacher retention and student mobility, with student mobility as both a covariate with teacher retention, and as an intermediate outcome through which teacher retention affects student achievement.

¹ This study is limited to high schools because it is part of a larger project that focuses on the impact of high school curricular changes and teacher supply and demand on high school student achievement. However, preliminary analyses of elementary and middle school teacher retention rates were conducted, and there did not appear to be a significant relationship between teacher retention and achievement in the elementary and middle school levels. Therefore, this study focuses on high schools for both substantive and methodological reasons.

In this study, teacher turnover refers to teacher movement into and out of schools (which may or may not be movement into or out of the profession). In other literature, teacher attrition refers to teachers leaving the profession entirely. For the school as an organization, however, attrition and turnover have similar effects on the schools in that both represent a decrease of staff that must be replaced (Ingersoll & Perda, 2009). This study takes an organizational perspective, considering teacher retention at the school level as an organizational characteristic of each high school, and investigates the impact of this organizational characteristic on the effectiveness of the organization itself (i.e. student achievement). The phenomenon of teacher turnover has been studied primarily at the state or national level (Ingersoll, 2001; Ingersoll & Perda, 2009; Loeb, Darling-Hammond, & Luczak, 2005). This approach, while powerful, can mask important differences that occur on a school-by-school basis. By contrast, this is a school-level analysis that utilizes state administrative data from Michigan to take into account distinct organizational factors related to teacher retention and student achievement.

Additionally, this analysis investigates introduces student mobility into the relationship between school-level teacher retention and student achievement outcomes. Student mobility, like teacher retention, can be considered an organizational characteristic of schools and a potential indicator of school climate. However, the interaction between teacher retention and student mobility is not clear. This study estimates the relationship between teacher retention and student mobility, both as a covariate with teacher retention in predicting student achievement outcomes, as well as an intermediate outcome predicted by teacher retention.

Background to the Problem

Teacher Turnover as an Organizational Characteristic of Schools

Ingersoll (2001) has argued that teacher turnover is an organizational feature of a school, and as such, contributes to the community and culture of the school. Teacher turnover, as a characteristic of the school, can contribute to the ability of a school to build trust among members, and to develop a sense of community. This draws on the sociology of education, which has shown that the presence of a sense of community and cohesion among families, teachers, and students is important for the success of schools (e.g., Bryk & Schneider, 2002; Durkheim, 1961; Grant, 1988; Parsons, 1959; Rosenholtz, 1989, Waller, 1932). A body of evidence suggests that the community of the school has important implications for school performance and effectiveness (Coleman & Hoffer, 1987; Rosenholtz, 1989; Bryk, Lee, & Smith, 1990). Specifically, the communal nature of certain schools, such as private schools, creates an environment that reinforces shared values and that leads to higher levels of social capital (Coleman & Hoffer, 1987). This sense of community is created by shared values, trust, and reciprocity between individuals in the school (Bryk & Schneider, 2002). Schools in which there is a stronger sense of community and cooperation between students and teachers have been shown to have higher achievement levels, as well as a more equitable distribution of achievement (Lee, Bryk, & Smith, 1993; Lee & Smith, 1997)

Prior research on teacher turnover has demonstrated that teacher turnover is not evenly distributed among schools. Schools with higher proportions of minority students, low-income students, and low-performing students have higher attrition rates, and urban school districts are likely to have higher attrition rates (Guarino et al., 2006). Specifically, the type of schools that tend to have high turnover are: those with high rates of student

poverty (Hanushek et al., 2004; Shen, 1997; Smith & Ingersoll, 2004); small schools (Ingersoll, 2001; Stockard & Lehman, 2004); schools with high numbers of minority students (Carroll et al., 2000; Hanushek et al., 2004); charter schools (Smith & Ingersoll, 2004); those with a high proportion of inexperienced teachers (Shen, 1997); private schools (Smith & Ingersoll, 2004; Ingersoll, 2001; Whitener et al., 1997; Arnold, Choy & Bobbitt, 1993) and urban schools (Lankford et al., 2002). Several analyses found that working conditions, particularly large class size, facilities problems, multi-track schools, and a lack of textbooks, are key factors in predicting teacher attrition (Loeb, Darling-Hammond, & Luczak, 2005; Futerick, 2007; Hanushek et al., 2004; Guarino et al., 2006).²

High rates of teacher turnover may be disruptive to the quality of the school and school performance and effectiveness (Ingersoll, 2001; Ingersoll & Perda, 2009), but the mechanism by which teacher turnover impacts achievement or other outcomes is unclear. Some evidence suggests that when qualified teachers leave, they are replaced by less qualified teachers (Reichardt, 2008), which in turn is inferred to have a negative impact on school quality due to younger, more inexperienced teaching staff rotating in to fill these positions (National Commission on Teaching and America's Future, 2002). This

² The focus of this paper is on school-level teacher retention, and does not address the predictors of and reasons for individual-level mobility decisions. However, other research addresses these individual reasons for leaving teaching, which include childbearing (Stinebrickner, 1998; 2002); financial considerations (Dolton & van der Klaauw, 1999; Shen, 1997; Loeb, Darling-Hammond, & Luczak, 2005). For a more thorough review, see Guarino et al. (2006). There is debate about whether salary or organizational factors are more important: Stinebrickner (1998) finds that salary considerations are more important than organizational characteristics, while Hanushek, Kain, and Rivkin (2004) find that teacher attrition is related more to being in schools with lower-achieving and minority students than to salary considerations. One of the most important reasons for teacher turnover was job dissatisfaction (Ingersoll, 2001). This suggests that if teachers do not like their jobs, they are more likely to leave, but that this decision is not based solely on monetary considerations. Rather, teachers choose to leave due to issues that relate to the climate of the school—lack of support from the school administration, or student discipline problems (Ingersoll, 2001).

does not necessarily address the possibility that turnover, as a “school climate” variable, could have an impact on student outcomes. The majority of the current research on teacher turnover makes this theoretical leap—teacher turnover means “new” teachers in the classroom, which means they are less likely to be experienced and effective. Given that evidence from other research shows that teacher experience and effectiveness matter for student outcomes, teacher turnover is presumed to affect student outcomes via this introduction of “new” teachers into the classroom.³ This hypothesis, however, is largely untested.

While there is substantial evidence regarding the characteristics of schools that are related to high teacher turnover, there is a lack of research that investigates teacher turnover as factor in student achievement outcomes. There is an underlying assumption that high levels of organizational turnover are related to decreased performance (Ingersoll, 2001). This is due to the fact that organizational research has shown that organizations with unclear or nonroutine processes or systems that require higher levels of interaction among members of the organization are more likely to experience higher turnover and thus decreased performance (Burns & Stalker, 1961; Ingersoll, 2001; Kanter, 1977; Likert, 1967; Porter, Lawler, & Hackman, 1975; Turner & Lawrence, 1964; Walton, 1980). Schools are seen as these types of organizations (Bidwell, 1965, Ingersoll, 2003, 2001; Lortie, 1975).⁴ However, at this juncture, this is primarily a hypothesized link that has not been tested extensively.

³ While this paper focuses on the organizational and communal implications of teacher turnover, turnover is also costly in terms of monetary outlays for schools, with \$10,000 being the estimated cost of hiring a new teacher (Barnes, Crowe, & Schaefer, 2007; Milanowski & Odden, 2007; Reichardt, 2006).

⁴ Some teacher supply research suggests that the teaching profession is plagued by abnormally high rates of turnover within schools, as well as high rates of attrition from the profession entirely (Ingersoll, 2001).

Student Retention: The Other Level of Retention

Student retention and mobility is another critical component of the organizational culture of a school. Mobility in high school has been shown to be related to diminished odds of high school graduation and decreased academic achievement, even after controlling for prior achievement, student and family background, and residential mobility (Haveman & Wolfe, 1995, Rumberger & Larson, 1998; Temple & Reynolds, 1997). Other studies have linked higher rates of student mobility in grades one through eight to an increased risk of dropping out at the high school level (Rumberger & Larson, 1998; Swanson & Schneider, 1999; Teachman, Paasch, & Carver, 1996). Student mobility rates were higher than dropout rates among high schools, and averaged 19% for the typical high school (Rumberger & Thomas, 2000).

Student mobility disrupts the formation of positive relationships between students and teachers, which are important for success in school (Newman, Lohman, Newman, Myers, & Smith, 2000; Stanton-Salazar & Dornbusch, 1995). Studies have shown that middle school teacher bonding is positively related to academic achievement (Johnson, Crosnoe, & Elder, 2001; Muller, 2001). When students move between schools, these are the critical social ties that are broken (Midgley, Feldlaufer, & Eccles, 1989; Roeser, Eccles, & Sameroff, 1998). These student/teacher relationships are important because teachers can help develop a significant attachment to learning through their

However, others have found that when comparing teachers to comparable fields, such as nurses, social workers and accountants, teacher turnover is not significantly higher (Harris & Adams, 2007). Stinebrickner (2002) found that exit rates are not lower in other professions, and that non-teachers change professions more but non-teachers also return to the workforce more quickly after an exit than teachers do. The crux of this argument appears to revolve primarily around pre-retirement attrition. Harris & Adams (2007) suggest that teacher attrition is lower, but that there are larger numbers of early retirements than in other professions. There is also some evidence to suggest that the distribution of teacher attrition is U-shaped, with high attrition among older and very young teachers (Grissmer & Kirby, 1997; Harris & Adams, 2007)

encouragement (Croninger & Lee, 2001; Rosenfeld, Richman, & Bowen, 2000). These types of bonds provide students with access to an informal network of knowledge in their schools (Stanton-Salazar & Dornbusch, 1995).

Importantly, however, student mobility is not merely an important factor in predicting achievement or in the formation and cessation of individual social ties. While experienced by individuals, student mobility can also be considered a factor in the school organizational culture much like teacher mobility. Student mobility affects not only the individual, but also the organization itself. If both students and teachers leave a school in response to a disorganized school culture and climate, these two factors are likely interrelated, but the relationship between these two levels of “churn”—teacher and student—as not been studied. More importantly, the relationship between teacher retention, student mobility, and student achievement outcomes is unclear. While the evidence shows that student mobility has a negative effect on achievement outcomes, it stands to reason that teacher retention either predicts or interacts with student mobility. As another indicator of school organizational culture, student mobility and teacher retention may exist in a feedback loop, whereby increased levels of one predict increased levels of another, and where this disorganization affects student achievement.⁵

Teacher Turnover in High Schools and Its Relationship to Student Achievement

It has been nearly a decade since the initial passage of the No Child Left Behind Act of 2001. In that time, increasing and measuring student achievement has become an established element in every state’s educational system. In Michigan, one of the key responses to No Child Left Behind, as well as to demands for increased rigor at the high

⁵ Michigan is also a school choice state, which means that parents and students can choose to attend districts and schools other than their home district, provided the school has sufficient room.

school level and stronger college and career preparedness for Michigan students was the enactment of the Michigan Merit Curriculum. Established in 2006, the Michigan Merit Curriculum increased high school graduation requirements to four years of mathematics, four years of English Language Arts, three years of science, three years of social studies, and two years of a world language.⁶ This reform put Michigan in a class of 18 states who have increased their graduation requirements to include four years of challenging mathematics (Achieve, Inc., 2008).

This study focuses on mathematics achievement scores as the main outcome of student achievement. This outcome was chosen for several reasons. Mathematics consists of a set of sequential courses, in that certain skills are required before new skills can be learned. Mathematics is hierarchically structured, with specific courses required as prerequisites to other courses. Hierarchically or concurrently structured courses are likely to be implemented more successful in schools with a more stable instructional workforce, as teachers can work together to ensure that all of the required skills are taught in the classes that students are likely to take, and that there are no “gaps” in instruction at a given school. This sort of curriculum monitoring and alignment may be more difficult to attain in schools with higher rates of teacher turnover.

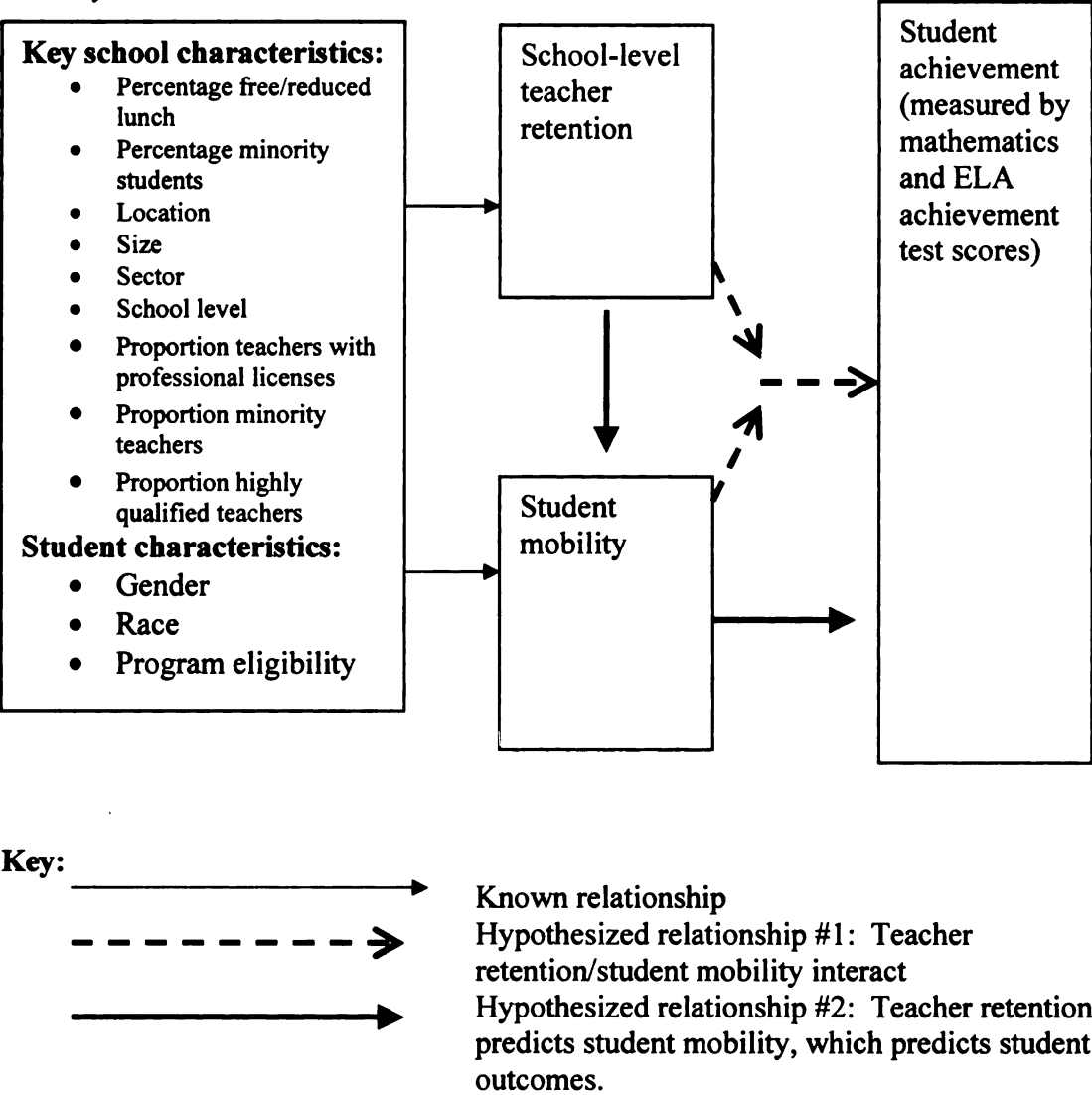
Mathematics is a gatekeeping course for successful high school completion and college entry. Successful completion of math courses has been associated with more short-term positive academic and social outcomes (Frank et al., 2008), and increasing the likelihood of attending college (Adelman, 1999; Sadler & Tai, 2007; Sells, 1973;

⁶ The Michigan Merit Curriculum also specifies specific courses in mathematics, science and social studies, as well as credits in the arts, physical education and an online learning experience. For more information, see http://www.michigan.gov/documents/mde/New_MMC_one_pager_11.15.06_183755_7.pdf.

Simpkins, Davis-Kean, & Eccles, 2006), particularly at four-year institutions (Schneider, Swanson, & Riegle-Crumb 1998; Riegle-Crumb, 2006). Additionally, math has been considered to be the primary key to the social organization of the school, due to its utility in defining academic tracks (Gamoran & Hannigan 2000; Lucas & Good 2001; Stevenson, Schiller, & Schneider, 1994).

This study also tests the relationship between teacher retention and student achievement in the context of student mobility. Therefore, student mobility is treated both as an *outcome*, as well as a covariate. Figure 5.1 below demonstrates the hypothesized relationships between teacher retention, student mobility and student achievement outcomes. The solid arrows indicate relationships that have been established or investigated by previous research, such as predictors of teacher retention and student mobility. However, how these factors relate to student achievement is unclear. The first hypothesized mechanism, illustrated by the red arrows, suggests that teacher retention affects student achievement through an intermediate outcome, student mobility. The second suggested mechanism is that teacher retention affects student achievement directly, and that student mobility interacts with teacher retention in predicting this relationship.

Figure 5.1: Conceptual Model for the Relationship Between Teacher Retention, Student Mobility and Student Achievement Outcomes



The Relationship Between School Characteristics, Teacher Turnover, and Student Outcomes: Hypotheses and Research Questions

School characteristics and the organizational nature and community of a school have been shown to be related to outcomes such as social capital. Additionally, organizational theory has suggested that the strength of an organization is related to its performance. Teacher retention is a phenomenon experienced at the school level. Schools

with higher rates of teacher retention may be able to increase organizational performance by increasing coordination between subjects and grades over time, aligning expectations across subjects and grades, and reinforcing instructional messages. However, these types of functions are likely to occur more frequently in schools where instructional staff have taught together over a period of time and develop a stable school culture that defines and reinforces these messages (Nonaka, 1994). Therefore, higher rates of teacher retention can be hypothesized to create a more positive and stable culture in high schools, which in turn can translate into increased student achievement in mathematics.

A related component of the school culture is student mobility. Increased student mobility has been shown to be detrimental, both to the individual student and to schools with high rates of mobility. One of the reasons for this is that mobility at the student level can lead to a more disjointed, disorganized school culture, which in turn has negative impacts on achievement. Increased teacher retention may be able to mediate that impact by providing a stable teacher workforce, even if the student workforce is more mobile. There are two levels of retention—teacher and student—and it is hypothesized that these interact, with higher levels of teacher retention mediating the negative effect of low student retention and amplifying the effect of high student retention.

Alternatively, teacher retention may impact student achievement outcomes via the mechanism of student mobility. Low levels of teacher retention, by creating the type of disorganized school culture described above, may cause students to move between schools at higher-than-average rates, as students respond to the lack of community and continuity at the school. Parents may also choose to remove students from schools where the culture is less organized and supportive. Students who are more likely to change

schools or drop out due to a lack of attachment to school may not receive the same kind of structured encouragement from teachers and from their school culture to stay in school when teacher retention rates are low and the culture is disjointed. It is hypothesized that teacher retention may act indirectly on student achievement by causing higher rates of student mobility. This question can be tested by utilizing the longitudinal nature of the data, with predictors that are causally prior to the outcomes.

ANALYTIC METHODS

This study utilizes a multilevel modeling approach to estimate the contextual effects of the key predictors of interest, school-specific rates teacher retention and student mobility. Although fixed effects models can be used to control for school factors without making parametric assumptions, I choose to express these models in a multilevel modeling framework, with students at level-1 and schools at level-2, in order to estimate effects simultaneously at both levels as well as estimate a crosslevel interaction. To estimate the relationship between school-level teacher retention, student mobility, and other school contextual factors and student achievement outcomes, I controlled for student level factors such as prior ability, gender, race, and free or reduced lunch eligibility. The models are random intercept models, which model each school mean as a function of the key predictor of interest, school-level teacher retention, and other school covariates. There are no random slopes estimated; however, crosslevel interaction terms are used to explore the relationship between school-level teacher retention and specific student characteristics. It is important to remember that these models consist of students nested within schools; although teacher data is available and used to calculate school-level teacher retention and other instructional workforce characteristic variables, teachers

are not currently linked to students in the Michigan data, and thus, estimating the impact of a given teacher's retention or mobility on a given student's achievement outcomes is not possible. Moreover, this study takes an organizational approach, in which it is hypothesized that it is the composition of the instructional workforce and the organizational culture of a school that impact student achievement outcomes, and therefore modeling individual outcomes as a function of school-level predictors is appropriate.

DATA AND MEASURES

Data Source

The data source utilized for this paper is one of its strengths. As Ingersoll (2001) points out, one of the main challenges to obtaining more precise estimates of both the causes and effects of teacher turnover has been a lack of data, with many studies conducted as single-city studies or with nationally representative data sets that are subject to issues of selection bias and generalizability. This study utilizes a set of rich longitudinal administrative data from the state of Michigan, including data on *all* teachers, students and schools in the entire state, collected longitudinally over a period of four years.^{7,8} This allows for the study of teacher retention for all teachers in all schools. One common issue with teacher retention studies is that when teachers change schools,

⁷ Although this can be considered universe data, with all observations accounted for, it still represents a sample of a larger theoretical population, albeit a highly reliable one. We would expect the error estimates to be smaller, given the fact that this is essentially a well-specified, very large sample, but it is still treated like a sample and inferential statistics are employed for analytic purposes.

⁸ Michigan has collected administrative data for more years than four years. The current system began to be implemented in 2002. However, the older data is less reliable and less complete. Michigan has undertaken significant efforts with regards to data quality, data integrity, and data coordination across the many agencies that collect, report, and utilize administrative data, and have made great improvements in their data. For this reason, the last four years are utilized, as these are all highly reliable and complete years of data.

they may leave the sampling frame of the study. In these data, while it is not possible to account for teachers who leave the sampling frame by leaving the state, it is possible to account for all within-state migration, for all teachers. The data sets are described in full in Chapter 2: Data and Methods, and will not be repeated here.

Measures

Main independent variable: The key predictor of interest is school-level teacher churn rate. Teacher churn rate is calculated by the teacher retention rate, the number of teachers in a given school who remain the same from one year to the next.⁹ Teacher retention rate is calculated over four years, three retention time points: school years 2004-2005, 2005-2006, 2006-2007 and 2007-2008. These rates are averaged to generate an average retention rate for each high school in the sample. Retention is based on a teacher remaining in the same school from one year to the next. If a teacher is in multiple schools, they are counted as either a stayer or mover from all schools in which they teach. This calculation is at the person-level.¹⁰ Figure 5.2 shows the distribution of retention rates over the three retention time points. This variable is dichotomized into two categories for analysis: schools with teacher retention rates of 85% and lower, and schools with retention rates of greater than 85%.¹¹

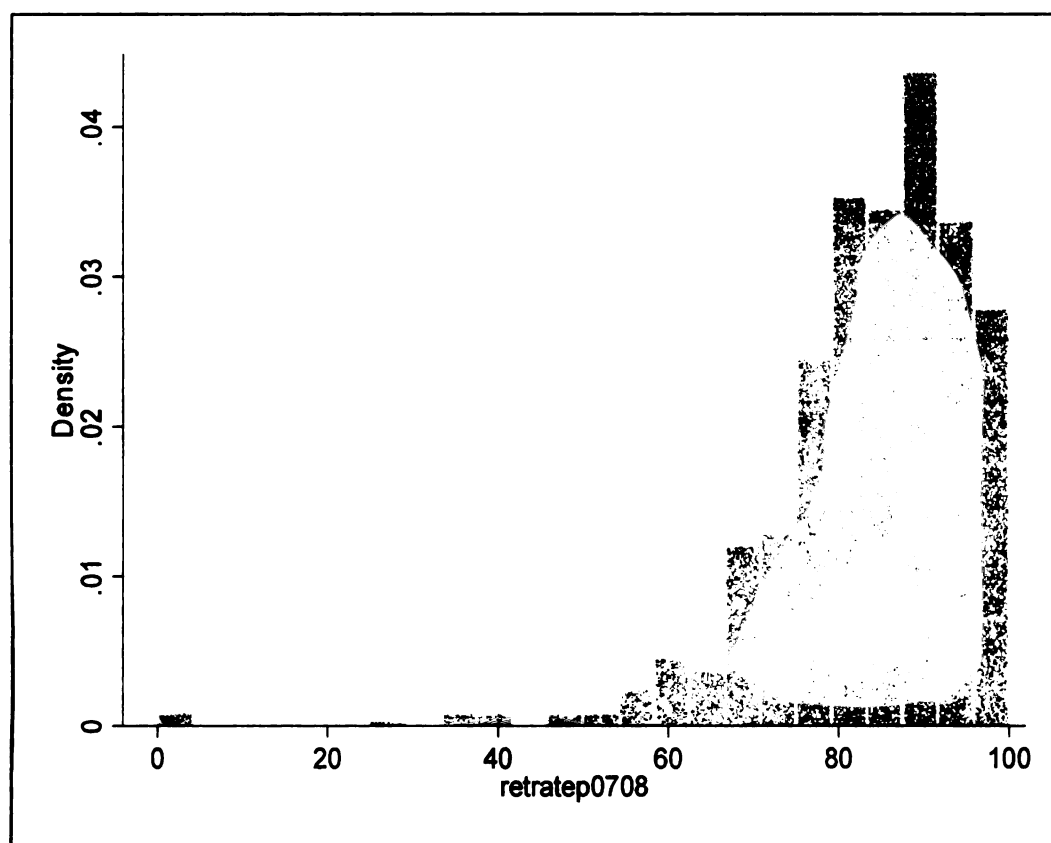
⁹ This method is used rather than calculating the number of “new” teachers in a school in a given year because of the difficulty of defining the denominator for that calculation.

¹⁰ Like Ingersoll (2001), I analyze all turnovers or departures, and do not distinguish between teacher attrition (from the profession) and teacher mobility (between schools). If a teacher is in the same school from one year to the next, they are retained; if they are not, they have “turned over.” This focuses on the organizational aspect of teacher retention, as the consequences to the organization are the same regardless of whether a teacher leaves the field entirely, or simply moves to another school (Ingersoll, 2001).

¹¹ The models were also run with teacher retention rate as a continuous variable. However, the 85% cut represents an important threshold and it is more useful for policy implications to consider schools in two categories, rather than on an incremental continuum.

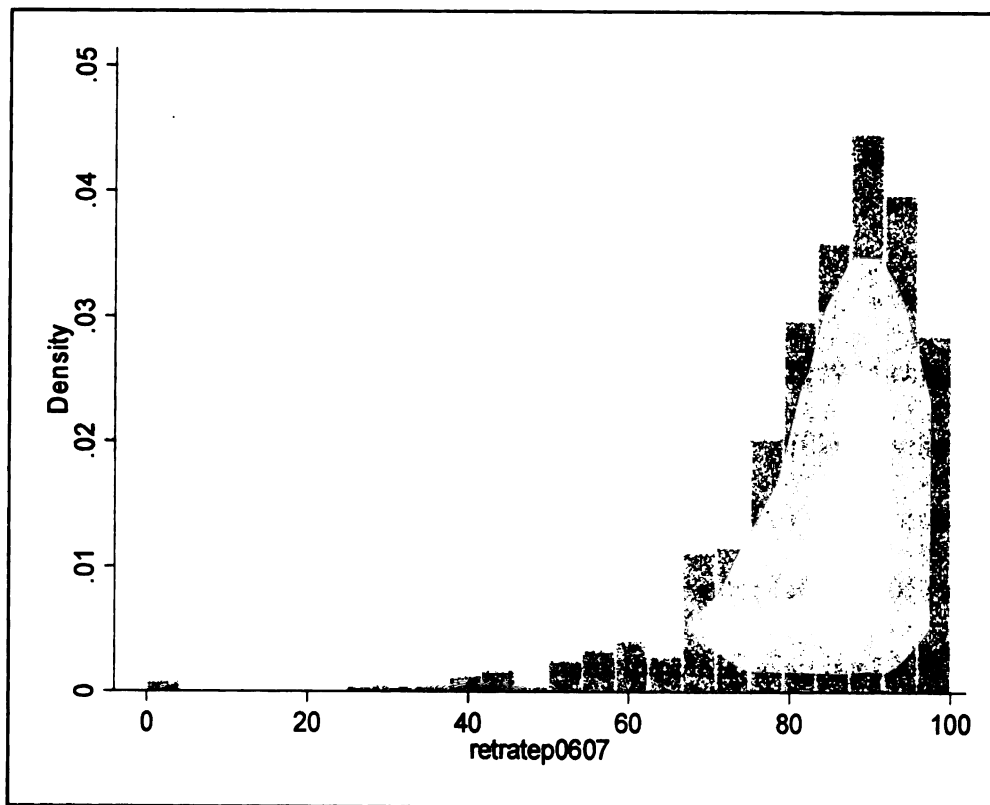
It is important to note that declining teacher retention rates could be a function of economic factors, such as layoffs or early retirement initiatives. However, by considering teacher churn rate as an organizational characteristic, the question of why churn rates are increasing is less relevant. The goal of the study is to interrogate the relationship between school-level teacher churn rates, regardless of their reason, and student achievement outcomes.

Figure 5.2: Distribution of School-Level Teacher Retention Rates

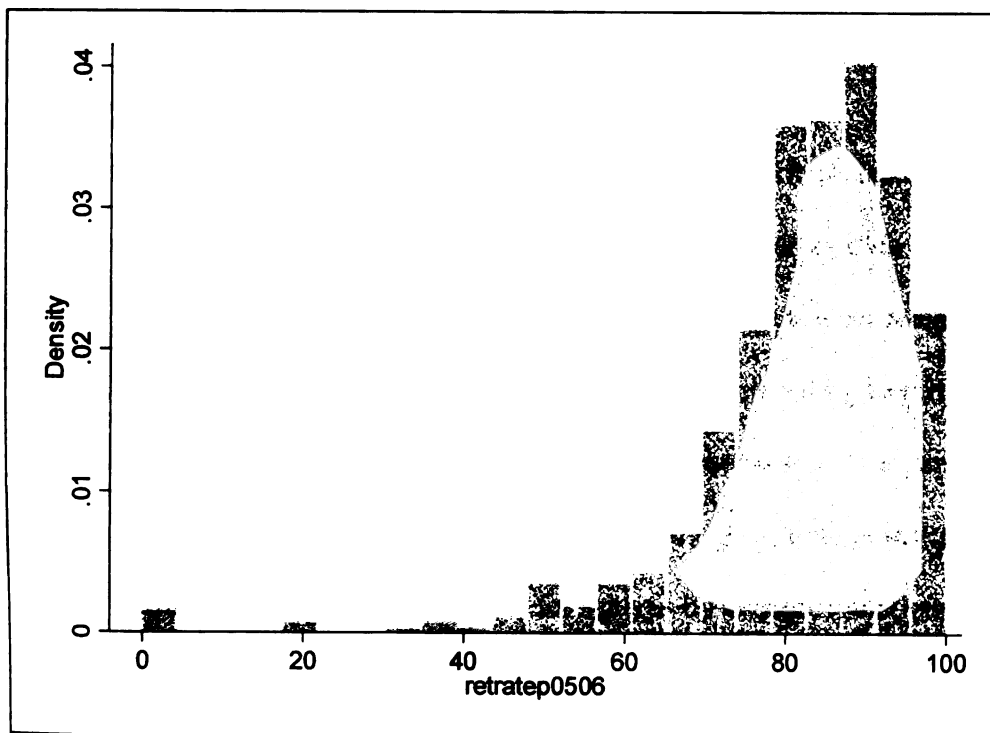


Retention Rate SY 2006-07 to SY 2007-08

Figure 5.2 (cont'd)



Retention Rate SY 2005-06 to SY 2006-07



Retention Rate SY 2004-05 to SY 2005-06

Using the longitudinal nature of the data, the retention variable is calculated over the years 2005-2008; the outcome measures are from the 2009 achievement tests. Therefore, the teacher retention occurs prior to the outcome. While this does not establish causality, it helps suggest the causal ordering of events.

Dependent variables: The outcome variable for the student achievement models is the scale score on mathematics achievement tests from the 2009 Michigan Merit Examination. For the student mobility models, the variable “different school” is used. This variable is constructed for all students who took the MME in the spring of 2009 and indicates whether or not they were in the same school in the fall of 2006 (when they were entering freshmen) as they were when they took the MME in spring 2009.¹²

School-level covariates and student covariates were described in Chapter 2: Data and Methods, and will not be repeated in detail here. For reference, Table 5.1 describes the variables used only for this analysis and presents their summary statistics.

ANALYTIC APPROACH

This study posits school-level teacher retention as a critical factor in explaining variations in student achievement in mathematics, either directly or via student mobility. Therefore, the first step in the analysis is to calculate average teacher retention rates across four years, three retention time points. This was achieved by identifying the proportion of the instructional staff of a given school that remained the same from one year to the next, and then averaging those rates over the three time points. Simple descriptive statistics were then calculated for schools by average retention rate, and for all

¹² Drop out rates present a potential concern. However, this file includes all students who took the MME in 2009 and who had a valid MEAP pretest from 2005. Therefore, these are not drop outs. Students who moved between schools did not drop out, but moved to another school. Drop outs are not included in this analysis.

student- and school-level variables that are used in the multilevel models. Table 3.9 presents the mean school-level teacher retention rates (averaged over four years) by school characteristics (in Chapter 3, referenced here).

The Multilevel Models to Be Estimated

To test the effects of teacher retention on both student mobility and student achievement outcomes, multilevel models (i.e., models with random effects) of student mobility and student achievement in mathematics with high school students nested within schools are estimated. Two sets of models are estimated. These models are discussed in greater detail in Chapter 2: Data and Methods, and only key elements are repeated here. The first set of models are hierarchical linear models (HLMs) to test the effects of teacher retention on student achievement outcomes, and the second are hierarchical generalized linear models (HGLMs) to predict student mobility as a function of teacher retention. The specifications are included below for reference.

In the first set of models, teacher retention predicts student achievement both directly and as a crosslevel interaction with student mobility. The purpose of these models is to test whether or not teacher retention has a direct effect on student achievement, as well as to test whether student mobility interacts with teacher retention in this relationship.

The general model specification for the HLM models is as follows:

Level 1 model: (1)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_j Q' + u_{0j}$$

where

Y_{ij} = outcome (mathematics scale score) for each student i in school j
 β_{0j} = each school mean, represented as a function of the grand mean, student mobility, the matrix of student-level predictors, the school teacher retention rate, and the matrix of school level predictors
 β_{1j} = coefficient for student mobility
 β_j = vector of coeffs for school j
 Z' = vector of student covariates for school j
 γ_{00} = grand mean (intercept)
 γ_{01} = effect of teacher retention on B_{0j} (each school mean)
 γ_j = vector of school-level predictors)
 Q' = vector of school covariates
 u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$
 r_{ij} = level 1 variance (student error term), r_{ij} distributed iid $N(0, \sigma^2)$

In the final HLM model, a crosslevel interaction term is added. The specification for this model is the same as above, with the “same school” slope predicted by teacher retention rate. The slope is not allowed to vary (i.e. is fixed).

Level 1 model: (2)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_j Q' + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{teacher retention})$$

Student mobility is a binary outcome variable; therefore the use of a standard level 1 multilevel model is inappropriate (Raudenbush & Bryk, 2002). Student mobility is indicated by whether or not a student remained in the same school from the fall of 2006, which was their freshmen year, until the spring of 2009, when they took the high school achievement test. To attempt to establish a causal ordering, the longitudinal nature of the data is utilized. Figure 5.3 shows the timing of the data.

One challenge to this analysis might be that the causal path could be in the opposite direction—that student mobility predicts teacher retention, or that student achievement predicts both student mobility and teacher retention. In order to attempt to address that criticism, the longitudinal nature of the data is utilized. Prior individual student academic achievement from their 8th grade pretest, as well as a prior cohort-level mean mathematics achievement from 8th grade are included, to control for both prior individual and group achievement. Teacher retention is calculated over the years 2005-2008, which ends one year prior to the student mobility outcome measure. A prior measure of school-level student mobility—the mobility rate of the 2007 cohort of students for each school—is also included.

The general structure of the HGLM models follows below.

Level 1 structural model:¹³ (3)

$$\eta_{ij} = \beta_{0j} + \beta_j Z'$$

Level 2 model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{teacher retention}) + \gamma_{02}(\text{student mobility 2007 cohort}) + \gamma_j Q' + u_{0j}$$

$$\beta_{pj} = \gamma_{po} \text{ for } p > 0$$

η_{ij} = the log odds of remaining in the same school for each student i in school j

¹³ The level 1 model in HGLM consists of three parts: a sampling model, a link function, and a structural model. The sampling model assumes that Y_{ij} , given the predicted value μ_{ij} , is distributed NID (μ_{ij}, σ^2). The level-1 predicted value, μ_{ij} , can be transformed so that the predictions remain within the given interval, which produces the transformed predicted value η_{ij} . This transformed predicted value is now related to the predictors of the model through the linear structure model. Combining the sampling model, link function, and level 1 structural model reproduces the familiar level-1 HLM model (Raudenbush & Bryk, 2002). The level 1 variance is now heteroskedastic

β_{0j} = each school mean, represented as a function of the grand mean, the matrix of student-level predictors, the school teacher retention rate, and the matrix of school level predictors

β_j = vector of coeffs for school j

Z' = vector of student covariates for school j

γ_{00} = grand mean (intercept)

γ_{01} = effect of teacher retention on B_{0j} (each school mean)

γ_{02} = effect of 2007 cohort student mobility on B_{0j} (each school mean)

γ_j = vector of school-level predictors)

Q' = vector of school covariates

u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$

A baseline for the HLM models with mathematics as the outcome is established by estimating an unconditional random effects ANOVA (output not reported). This allows for the calculation of the intraclass correlation, or the proportion of variance that is between schools. Following the estimation of this baseline model, a series of multilevel models were estimated. Four different multilevel models are presented in Table 3 (mathematics). The first, a bivariate model, estimates the bivariate relationship between school-level teacher retention and student achievement in mathematics. The second (Model 2) includes a second school-level predictor shown in the preliminary analyses to be highly correlated with the student achievement, percent free lunch, in order to control for other key school-level factors. This model also introduces a pretest measure of mathematics achievement at the student level, in order to account for student prior ability. The final two models are fully specified models with all predictors at level 1 and level 2, with the final model containing a crosslevel interaction between teacher retention and

students who remain in the same school, to test the hypothesis that there is a multiplicative effect of teacher retention and student retention on student achievement.¹⁴

A similar modeling scheme is utilized in the models predicting student mobility, beginning with a bivariate model with teacher retention predicting student mobility. The second model includes only the school-level measure of prior student mobility for the 2007 cohort. The final model is a full model, including all level 1 predictors (gender, race and program eligibility) as well as level-2 predictors.

Sensitivity Analyses

When utilizing data from observational studies, there is a concern regarding the impact of an unobservable characteristic on the outcome, one that might invalidate the inferences drawn from the study. When using state administrative data, this is a concern as well, as state data is rich in observations but does not include a large number of variables. Sensitivity analyses are conducted to test the robustness of the inferences to the influence of other unobserved characteristics. Therefore I will characterize the robustness of these inferences to the potential impact of confounding variables (Frank, 2000).

Chapter 2: Data and Methods provides further detail regarding these analyses.

RESULTS

Descriptive Statistics of Key School Variables by Teacher Retention Rate

Table 3.9 (included in Chapter 3) presents mean school-level teacher retention rate by school characteristics, in order to develop a profile of the type of schools that have low teacher retention rates. This table was described in detail in Chapter 3, and

¹⁴ All Level 1 predictors, with the exception of pretest score and same school, are grand mean centered, in order to control for their effect, rather than partial out the impact attributable to student and school. A model with all Level 1 predictors group mean centered was run (output not reported); the effects for gender and race are largely at the student level, not at the school-mean level. Therefore, the decision was made to grand mean center race and gender in the reported models.

therefore only key details are repeated here. In general, the types of schools that have lower mean teacher retention rates are city schools, schools with high proportions of minority and low-income students, small schools and charter schools. Schools with a lower percentage of teachers with professional licenses, and schools with a higher percentage of minority teachers also appear to have lower mean teacher retention rates.

MULTILEVEL MODEL RESULTS

*Mathematics Achievement Outcomes*¹⁵

To assess the relationship between student mathematics achievement and school-level teacher retention, a series of multilevel models are estimated. Using the series of multilevel models outlined previously, I find that school-level teacher retention is positively and significantly associated with school mean mathematics achievement scores (see Table 5.2). In the bivariate regression (Model 1), schools with high teacher retention rates have mean mathematics achievement scores that are 13.28 scale score points higher than students in schools with lower teacher retention rates ($p \leq .000$). The proportion of variance explained between the unconditional model and the bivariate model is 21%. This effect for teacher retention remains in Model 2, with the addition of the measure of student prior ability, the math pretest score, as well as another important school-level covariate, percent free and reduced lunch. Schools with high retention rates have mean mathematics achievement scores that are 2.46 scale score points higher than schools with low retention rates ($p \leq .000$). Importantly, this model now explains 88% of the variance at level 2, and 52% of the variance at Level 1, which suggests that these two school measures, teacher retention and percent free and reduced lunch, account for much of the

¹⁵ Results from the baseline model, a oneway ANOVA with random effects, yield an intraclass correlation of 18% for the mathematics achievement outcome.

variation in school mean mathematics achievement. It could be the case that these measures both serve as proxies—teacher retention for a “school climate” set of predictors, and percent free and reduced lunch for a set of structural characteristics.

Model 3 and Model 4 introduce both school-level and individual predictors in addition to the measures included in Model 2. We see that in Model 3, the relationship between high school-level teacher retention and mathematics achievement is no longer statistically significant ($\gamma_{01}=0.27$, $p \leq .543$). However, other workforce composition variables are significant. Schools with a higher percentage of teachers with professional licenses experience an increase in mean mathematics scale score of .08 for each one percent increase in the percentage of teachers with professional licenses in that school ($\gamma_{02}=0.08$, $p \leq .010$). Schools with a higher percentage of minority teachers have significantly lower mean mathematics achievement, with each one percent increase in the percentage of minority teachers in a school associated with a corresponding -0.07 point decrease in mean mathematics scale score ($\gamma_{03}=-.07$, $p \leq .001$). The percentage of highly qualified teachers is also related to mathematics achievement, with schools with a greater percentage of highly qualified teachers demonstrated higher mean mathematics achievement ($\gamma_{04}=.04$, $p \leq .008$). As all of these variables relate to the organizational characteristics of the instructional workforce of a school, the evidence suggests that this composition does matter with regards to student achievement. From a school structural perspective, the percentage of students eligible for free and reduced lunch was again related to lower mean mathematics achievement, with each one percent increase in the percent of students eligible for free or reduced lunch associated with a .05 decrease in mean mathematics achievement for that school ($\gamma_{06}=-.05$, $p \leq .004$).

With regards to the student-level predictors included in Model 3, the key predictor of interest for students is student mobility. In Model 3, students who are not in the same school in 11th grade as they were in ninth grade have significantly lower mathematics achievement than those students who are in the same school ($\gamma_{10} = -3.07$, $p \leq .000$). The group mean for the same school variable, entered at Level 2, is not significant; this suggests that the relationship between student mobility and teacher retention exists primarily at the individual level; the negative effect of changing schools is experienced by the student. This effect is apparent even after controlling for student race and gender, as well as student prior ability.

Student mobility and teacher retention are both variables that are school organizational characteristics, and therefore, both are of interest in this analysis. To estimate a potential interaction between these two important predictors, a final model (Model 4) was estimated, and included a cross level interaction between teacher retention and the same school indicator variable. Again, students who change schools have lower mathematics achievement scores ($\gamma_{11} = -3.14$, $p \leq .000$). Most importantly, however, the crosslevel interaction term is significant ($\gamma_{12} = -2.14$, $p \leq .019$), which indicates that students who change schools when that school has a high teacher retention rate have mathematics scale scores that are, on average, 2.14 points lower than students who change schools when those schools have lower rates of teacher retention. In other words, there is a magnifying effect of teacher retention for students who remain in the same school. This suggests that students who do not stay in the same school and who are in schools with low teacher retention are at a serious disadvantage in terms of their mathematics achievement outcomes.

The final model explains 92% of the variance at the school level and 53% of the variance at the student level, indicating that this school-level model is able to capture much of the variability in mean mathematics achievement over schools. However, given the restricted range of student predictors, there is still somewhat substantial variation to be explained in student performance.¹⁶

The student achievement outcomes models presented provide two important pieces of evidence with regards to how teacher retention interacts with student achievement and student mobility. First, in the final model, the teacher retention rate is not a significant when predicting mathematics outcomes. However, student mobility is highly significant, with students who remain in the same school performing much better on mathematics achievement tests. This suggests that student mobility is a stronger predictor of student achievement than teacher retention. However, the second hypothesis still needs to be investigated: does teacher retention predict student mobility and thus relate to student achievement via the mechanism of student mobility? The models below investigate that relationship.

Student Mobility As an Outcome

In the bivariate model, school level teacher retention is strongly related to student mobility. Students in schools with high levels of teacher retention (defined as above 85% of teachers retained) are 0.35 times as likely to change schools as students in schools with low levels of teacher retention. Model 2 adds a prior measure of school student mobility, the aggregate mobility for the 2007 cohort. Students in schools with high teacher retention rates are 0.55 times as likely to change schools as students in schools with low

¹⁶ These models were estimated with ELA scale scores as an outcome in order to test these findings. The results were nearly identical, and are not presented here.

teacher retention rates ($p \leq .000$). Prior school-level student mobility is also significantly associated with students changing schools. Students in schools with higher prior student mobility are 1.06 times more likely to change schools than students in schools with lower levels of prior student mobility ($p \leq .000$).

In the full model (Model 3), the relationship between school-level teacher retention and student mobility remains. Students in schools with high levels of teacher retention are 0.62 times as likely as students in schools with low levels of teacher retention to change schools ($p \leq .014$), controlling for prior school-level cohort mobility, and school- and student-level covariates. Given the fact that the previous models demonstrated the strong relationship between student mobility and achievement outcomes, this suggests that low levels of teacher retention are related to high levels of student mobility, which in turn can lead to decreased student achievement outcomes.

Students in schools with higher levels of 2007 cohort student mobility are 1.04 times more likely to change schools ($p \leq .008$). This again suggests that these “churn” rates are on a feedback loop. Students in schools with high levels of prior student mobility are more likely to leave those schools; students in schools with high rates of teacher retention are less likely to leave those schools; and when students remain in those schools, this create a more stable school culture that leads both to more stability in population as well as increased achievement.

Turning to the other school-level covariates, the average mathematics pre-test score for the school was negatively and significantly related to student mobility, with students in schools with higher levels of prior student mathematics ability 0.97 times as likely to change schools ($p \leq .009$). At the student level, a student’s prior mathematics

achievement was negatively related to his likeliness to change schools (odds ratio=0.99; $p \leq .000$). Students who are eligible for free and reduced lunch are 1.66 times more likely to change schools than their non-free/reduced lunch eligible counterparts ($p \leq .000$). Black students and multiethnic students were all significantly more likely to change schools than their white counterparts (odds ratios 3.30 and 3.41, respectively), as were Hispanic students (odds ratio 1.26). Interestingly, female students were 1.09 times more likely to change schools as males ($p \leq .004$). Students in special programs were 2.02 times more likely to change schools as students not in these programs.

Sensitivity Analyses

The models presented above are only able to account for measurable confounding variables. One of the well-known issues in using state administrative data is that these data are rich in observations, but often “poor” in variables (Figlio, 2010). Therefore, robustness indices are used here as a complement to address problems related to unmeasured confounds.¹⁷ While the Impact Threshold for Confounding Variables (ITCV) does not control for the impact of these confounds, it does quantify how powerful they need to be in order to negate the inferences drawn (Crosnoe, 2009; Frank, 2000).¹⁸

The ITCV is the minimum product of the correlation between the predictor and confound and the correlation between the outcome and confound necessary to reduce an estimate below threshold for statistical significance of the key association of interest; in this case, the association between teacher retention and student mobility. If the actual,

¹⁷ As developed by Frank et al (2010), there are three major “threats” to inference in observational studies: a biased sample, a small sample, and unobserved covariates. Given the fact that this sample nearly approximates the population, as it is the entire state of Michigan, and given its sheer size, I consider here the greatest threat to inference to be unmeasured confounds. This is likely true of most studies utilizing state administrative data.

¹⁸ Further details regarding sensitivity analyses are provided in Technical Appendix C.

albeit unknown, product of these two correlations is greater than this threshold, than the inclusion of that unobserved covariate would invalidate the inference. Therefore, I calculate the ITCV for the teacher retention predictor.

The impact of an unmeasured confound (recall impact $= r_{v,y} \times r_{v,x}$, see Frank, 2000) would have to have magnitude greater than .04 to invalidate the inference. Thus to invalidate the inference that low levels of teacher retention predict student mobility, a confounding variable would have to be correlated with teacher retention at 0.19 *and* with student mobility at 0.19, accounting for covariates, z (Frank, 2000). These are moderate correlations.¹⁹ It is important to note that this assumes that the unmeasured confound is uncorrelated with the measured covariates (Frank, 2000). However, the relevant partial correlations that would create the impact of an unobserved confound would be smaller than these zero-order correlations because they would correlate with existing covariates (Frank, 2000; Frank & Sykes et al, 2008).

It is also useful to compare the threshold for an unmeasured variable to the impacts for measured covariates. At level 1, the strongest impact of the measured covariates, aside from the pretest of student ability, is for free and reduced lunch.²⁰ The impact of free and reduced lunch eligibility on the coefficient for school-level teacher

¹⁹ These calculations are obtained by regressing all of the predictors on the treatment (school-level teacher retention) and obtaining the R^2 value, and then regressing all predictors on the outcome without the treatment and obtaining the R^2 value. As the outcome was a student-level dichotomous outcome, which produces pseudo- R^2 s, the variables were aggregated to average school mobility and the continuous teacher retention predictor was used. Seltzer, Kim & Frank (2006) suggest that when performing a cluster-level regression with unbalanced data, precision weights should be used. These weights are calculated by $1/\tau + V_j$, where $V_j = \sigma^2/n_j$ and $\sigma^2 = np(1-p)$. See Technical Appendix D for more detail.

²⁰ The impact of the pretest measures of student ability and school-level student mobility are not compared to the impact of an unobserved covariate because they are obviously clear predictors of the outcomes. The impact of the pretest is well-known to be the most significant predictor and thus not a relevant comparison.

retention on student mobility is .011, which is the product of the correlation with teacher retention $-(.15)$ and the correlation with student mobility $(-.07)$. At level 2, the strongest impact of the measured level 2 covariates, aside from the pre-measure of school-level student mobility, is school average math achievement. The impact of school math achievement on student mobility is .06, which is the product of the correlation with teacher retention $(.30)$ and the correlation with school mean mobility $(.20)$. Thus, the impact of an unmeasured covariate necessary to invalidate the inference of .04 would have to be greater than the impact of the strongest level 1 covariate, although less than the impact of the strongest level 2 covariate. This suggests that the size of the impact of an unmeasured covariate would have to be similar in size to the other measured covariates.

Limitations

The most significant limitation of this study is that, like many studies conducted with state administrative data, there are a limited number of covariates available to use in the models, which raises the question of unobserved covariates that might invalidate the inferences made here. While I have attempted to identify and utilize all available data, to use effective modeling strategies, and to quantify the robustness of the inference to the influence of unobserved covariates, this does not solve the fundamental problem. It is possible that teacher mobility is caused by factors such as school leadership or neighborhood characteristics, and these are not available for the models. In the future, this relationship between teacher retention, student mobility, and student outcomes can be further investigated using fixed effects models to attempt to address these unobserved characteristics and their potential impact on the relationships studied here. Additionally,

at this point in time, assessment data that matches with a pre-test are only available for one cohort, although more data will become available each year. Finally, Michigan's assessment system only allows for one measurement of achievement at the high school level, which is not optimal from a measurement perspective.

Conclusions

Although teacher retention appears to have a direct effect on student achievement outcomes via an interaction with student mobility, this relationship can be explained with other covariates. The results from the first analysis demonstrate that remaining in the same school has a positive impact on student achievement outcomes. The second analysis in turn demonstrates that students in schools with higher levels of teacher retention are more likely to stay in the same school, which in turn has a positive impact on achievement. Therefore, teacher retention appears to affect student achievement via the mechanism of student mobility.

This suggests the importance of organizational culture on student achievement outcomes. Although I do not directly measure or test "school culture," the most positive impacts for student achievement are evident when there is a stable teaching workforce and a stable student body. This stability allows for the organization to function effectively as a unit, and also allows for cohesion, community, trust, and reciprocity to develop and exist between teachers and students, and within the teachers themselves.

It is important to consider what factors might be lead to increased teacher mobility, given the importance of schools maintaining a stable, qualified instructional workforce. These factors may include school leadership, alterations in state policies, economic changes in the state as a whole, or inhospitable work conditions. It is critical to

identify which factors may be increasing teacher mobility, as this analysis shows that when teacher mobility increases, there can be negative consequences for student achievement.

Table 5.1: Variable Description and Summary Statistics					
Variable	Description	Mean	SD	Min	Max
<i>Student Level</i> (n=96,556)					
Math achievement score	Math scale score on 2009 Michigan Merit Examination, Michigan's high school assessment	1096.7	31.1	950	1250
Math pretest achievement score	Math scale score on 2006 MEAP test, last test administered prior to high	813.59	24.1	470	952
Student mobility (different school)	Dummy, 1=different school in 9th grade than 11th grade.	0.14	0.35	0	1
Gender	Dummy, 1 = female	0.51	0.5	0	1
Racial/ethnic code					
American Indian/Pacific Islander	Dummy, 1 = yes. Combined American Indian and Pacific Islander due to small sample sizes	0.01	0.1	0	1
Asian	Dummy, 1 = yes	0.02	0.15	0	1
Black	Dummy, 1 = yes	0.15	0.35	0	1
White	Reference	0.79	0.41	0	1
Hispanic	Dummy, 1 = yes	0.03	0.17	0	1
Multiple race/ethnicity	Dummy, 1 = yes	0.01	0.08	0	1
Free/reduced lunch eligible	Dummy, 1=yes. Indicates whether a student is eligible for free or reduced lunch.	0.27	0.44	0	1
Special programs	Dummy, 1 = eligible for Title I, special education, Section 504, Limited English Proficient, or migrant programs	0.01	0.11	0	1

Table 5.1 (cont'd)					
<i>School Level (n=580 high schools)</i>					
Variable	Description	Mean	SD	Min	Max
Average teacher retention rate	Percentage of teachers who were retained in a school from one year to the next. Averaged over four years (2005-2008), three retention time points	86.2	8.57	0	100
Percent of teachers w/professional licenses	Percentage of teachers who have professional licenses in a school. Averaged over four years (2005-2008)	83.29	12.1	25.2	100
Percent minority teachers	Percentage of teachers in a school who are minority teachers. Averaged over four years	7.97	17.5	0	100
Percent highly qualified teachers	Percentage of teachers in a school who are highly qualified. Averaged over four years (2005-2008)	80.61	12.5	0	100
Percent free/reduced lunch	Percentage of student body eligible for free/reduced lunch	35.88	19.9	0	99.63
Percent minority	Percentage of the student body that is minority students	23.27	30	0	100
Locale					
City	Dummy, 1=yes	0.17	0.37	0	1
Suburb	Reference	0.28	0.45	0	1
Town	Dummy, 1=yes	0.13	0.34	0	1
Rural	Dummy, 1=yes	0.42	0.49	0	1
Charter	Dummy, 1 = school is a charter school	0.06	0.23	0	1
Magnet	Dummy, 1 = school is a magnet school	0.11	0.32	0	1
School size					
Small school	School has less than 300 students	0.15	0.35	0	1
Medium school	School has 300-999 students	0.53	0.5	0	1
Large school	School has over 1000 students	0.32	0.47	0	1
School-level student mobility (2007 cohort)	School-aggregate of the proportion of the 2007 cohort of MME testers who changed schools since the fall of 2005	13.02	15.9	0	100

Table 5.2: Hierarchical Linear Model Predicting Student Mathematics Achievement as a Function of School-level Teacher Retention

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>
	(<i>se</i>)		(<i>se</i>)		(<i>se</i>)		(<i>se</i>)	
Intercept γ_{00}	1093.46	.000	1093.39	.000	1093.47	.000	1093.47	.000
	(.51)		(.22)		(.19)		(.19)	
<i>Level 2 Variables</i>								
High teacher retention rate	13.28	.000	2.46	.000	0.27	.544	0.27	.543
(greater than 85%) γ_{01}	(1.31)		(.53)		(.45)		(.45)	
<i>(Less than 85% retention rate=reference)</i>								
<i>Instructional Workforce Characteristics</i>								
Proportion teachers w/prof licenses γ_{02}					0.08	.010	0.08	.010
					(.03)		(.03)	
Proportion minority teachers γ_{03}					-0.07	.001	-0.07	.001
					(.02)		(.02)	
Proportion HQ teachers γ_{04}					0.04	.008	0.04	.008
					(.02)		(.02)	
<i>School Structural Characteristics</i>								
Percent minority students γ_{05}					0.03	.074	0.03	.077
					(.02)		(.02)	
Percent free/reduced lunch students γ_{06}					-0.05	.004	-0.05	.004
					(.02)		(.02)	
City γ_{07}					0.13	.841	0.13	.840
					(.65)		(.65)	
Town γ_{08}					0.79	.202	0.79	.201
					(.62)		(.62)	
Rural γ_{09}					1.48	.007	1.48	.007
					(.54)		(.54)	
<i>(suburb=reference)</i>								
Small school γ_{010}					-0.86	.247	-0.86	.248
					(.74)		(.74)	
Large school γ_{011}					0.36	.426	0.36	.426
					(.45)		(.45)	
<i>(medium school=reference)</i>								
Charter γ_{012}					1.18	.500	1.18	.501
					(1.75)		(1.75)	
Magnet γ_{013}					0.17	.787	0.17	.787
					(.63)		(.63)	
Retention difference (Y3ret-Y1ret) γ_{014}					5.35	.003	5.35	.003
					(1.73)		(1.73)	

Table 5.2 (cont'd)*Group Means for Level 1 Variables*

Mean math pretest γ_{015}	1.20 (.03)	.000	0.97 (.03)	.000	0.97 (.03)	.000
Mean student mobility γ_{016}			-.01 (.007)	.318	-.01 (.007)	.318

Student Characteristics

Student mobility γ_{10}			-3.07 (.41)	.000	-3.14 (.41)	.000
Math pretest γ_{20}	0.91 (.008)	.000	0.90 (.008)	.000	0.90 (.01)	.000
Free/reduced lunch eligible γ_{30} (non-eligible=reference)	-3.81 (.20)	.000	-3.16 (.19)	.000	-3.15 (.19)	.000
Student race (white=reference)						
American Indian, γ_{40}			-2.18 (.74)	.004	-2.17 (.74)	.004
Asian, γ_{50}			1.52 (.51)	.004	1.54 (.51)	.003
Black, γ_{60}			-7.11 (.39)	.000	-7.07 (.39)	.000
Hispanic, γ_{70}			-2.10 (.51)	.000	-2.08 (.51)	.000
Multiethnic, γ_{80}			-2.84 (.96)	.003	-2.83 (.96)	.004
Female γ_{90}			-.21 (.14)	.135	-.021 (.14)	.139
Program eligible γ_{100}			-5.43 (.72)	.000	-5.42 (.72)	.000

Crosslevel Interactions

Diff school * teacher retention

Intercept (diff school) γ_{110}					-3.14 (.41)	.000
High school-level teacher retention rate, γ_{120}					-2.14 (.92)	.019

<i>Random Effects</i>		df		df		df		df
Level 2 variance component, τ_{00}	138.54	578	21.47	577	13.74	563	13.74	563
Level 1 variance component σ^2	805.56		381.94		378.04		377.99	
Proportion variance explained, Level 1	n/a		52%		53%		53%	
Proportion variance explained, Level 2	21%		88%		92%		92%	

****Pretest & student mobility group mean centered, group means at level 2; other L1 predictors grand mean centered. All level 2 variables grand mean centered.**

**Table 5.3: Hierarchical Generalized Linear Model Predicting Student Mobility
(1=different school from freshmen-junior year)**

Odds Ratios Reported

	Bivariate		Model 2		Model 3	
	<i>O.R.</i>	<i>p-val</i>	<i>O.R.</i>	<i>p-val</i>	<i>O.R.</i>	<i>p-val</i>
	<i>(se)</i>		<i>(se)</i>		<i>(se)</i>	
Intercept γ_{00}	0.10 (.08)	.000	0.10 (.07)	.000	0.10 (.07)	.000
<i>Level 2 Variables</i>						
High teacher retention rate γ_{01} (Greater than 85%)	0.35 (.17)	.000	0.55 (.17)	.001	0.62 (.20)	.014
<i>Low teacher retention rate=reference</i>						
Prior school-level student mobility (2007 cohort) γ_{02}			1.06 (.01)	.000	1.04 (.008)	.000
<i>Instructional Workforce Characteristics</i>						
Proportion teachers w/prof licenses γ_{03}					1.02 (.01)	.041
Proportion minority teachers γ_{04}					1.00 (.01)	.407
Proportion HQ teachers γ_{05}					0.99 (.01)	.034
<i>School Structural Characteristics</i>						
Percent minority students γ_{06}					1.00 (.01)	.415
Percent free/reduced lunch students γ_{07}					0.99 (.01)	.039
City γ_{08}					0.53 (.25)	.013
Town γ_{09}					0.65 (.27)	.110
Rural γ_{010}					0.61 (.20)	.012
<i>(suburb=reference)</i>						
Small school γ_{011}					1.69 (.20)	.010
Large school γ_{012}					1.22 (.21)	.345
<i>(medium school=reference)</i>						
Charter γ_{013}					1.91 (.48)	.182
Magnet γ_{014}					1.16 (.22)	.483

Table 5.3 (cont'd)*Group Means for Level 1 Variables*

Mean math pretest γ_{015}	0.97 (.01)	.009
----------------------------------	---------------	------

Student Characteristics

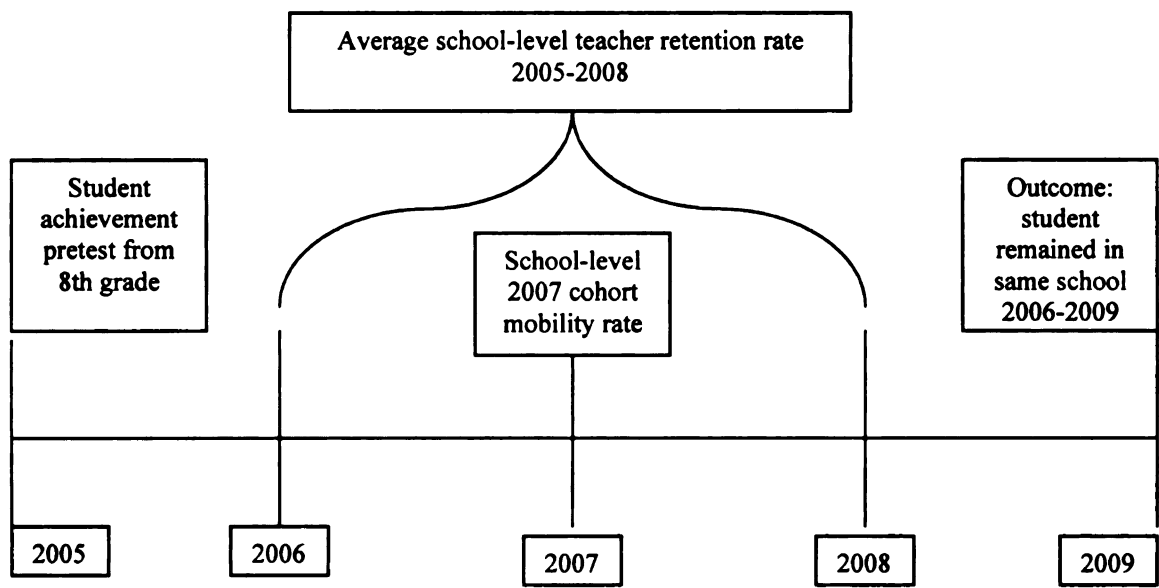
Math pretest γ_{10}	0.99 (.001)	.000
Free/reduced lunch eligible (1=yes) γ_{20}	1.66 (.04)	.000
American Indian, γ_{30}	1.34 (.14)	.064
Asian, γ_{40}	1.00 (.14)	.987
Black, γ_{50}	3.30 (.07)	.000
Hispanic, γ_{60}	1.26 (.12)	.052
Multiethnic, γ_{70}	3.41 (.19)	.000
Female γ_{80}	1.09 (.03)	.004
Program eligible γ_{90}	2.02 (.17)	.000

<i>Random Effects</i>	df		df		df
Level 2 variance component, τ_{00}	3.02	578	2.58	577	2.57
Proportion variance explained, Level 2					563

* All level 2 variables are grand mean centered.

**Pretest and same school are group mean centered, with group means included at level 2.
All other level 1 predictors are grand mean centered.

Figure 5.3: Timing of the Longitudinal Data for Establishing a Causal Path



CHAPTER 6: PROPENSITY ANALYSES OF THE EFFECTS OF TEACHER TURNOVER

The evidence provided in Chapter 5 strongly suggests the importance of high school-level teacher retention in decreasing student mobility and increasing student mathematics achievement. However, the multilevel methods and sensitivity analyses presented in Chapter 4, while rigorous, do not offer a clear framework for understanding the potential *effects* of increasing school-level teacher retention on student mobility. Therefore, this next section of the teacher retention analysis shifts focus to the *effects* of a school having high teacher retention. Using a quasi-experimental design, propensity score matching, the impact of high teacher retention on student mobility will be estimated by comparing similar schools that have either high or low rates of teacher retention. Both the treatment (teacher retention) and the outcome (student mobility) will be analyzed as school-level variables—the proportion of teachers retained, and the proportion of students who leave a school. The analysis presented in the prior chapter uses longitudinal data and multilevel modeling to attempt to establish the relationship between school-level teacher retention, student achievement and student mobility. However, the question remains: for schools that are similar on all characteristics with the only difference being their teacher retention rate, what is the effect of school-level teacher retention on school-level student mobility? In other words, if a school has low teacher retention, what is the potential effect on student mobility of increasing the teacher retention rate so that the school has a high teacher retention rate?

Estimating the Effect of High School-Level Teacher Retention¹

¹ Chapter 2: Data and Methods provides a detailed description of propensity score matching as an analytic technique, as well as develops the methods utilized here. Technical Appendix D also provides additional information regarding model specification and other details.

In order to estimate the effect of school level teacher retention on school-level student mobility, high school-level teacher retention is considered as a “treatment” experienced by some schools. This analysis uses a set of propensity-score weighting schemes to investigate the impact of high school-level teacher retention on student mobility for those schools in the margin of indifference (i.e. those most likely to respond to a treatment offer); those in the control (i.e. those with low teacher retention) and those in the treatment (i.e. those with high teacher retention). As a fourth strategy, this analyses estimates the average treatment on the treated using a series of matching methods, including stratification, nearest neighbor, kernel matching, and radius matching. The purpose is to test whether or not increasing school-level teacher retention may be a useful “treatment” to help to stabilize the student population and decrease student mobility.

In order to investigate this question, a school-level propensity score is estimated, and then a series of propensity score analyses are conducted to compare the average student mobility rates of schools based on their probability of having high school-level teacher retention.²

Propensity Scores: Weighting Method

Propensity scores allow for the comparison of those in a “treatment” group (schools with high teacher retention) and those in a control group (schools with low teacher retention) who have similar propensities to receive the treatment. The treatment,

² Originally, this analysis was conceived as a multilevel analysis that used propensity scores at level 2 with the outcome at level 1. However, due to the fact that this is not the general framework by which multilevel propensity score models are estimated, there are methodological and analytic implications that are beyond the bounds of this analysis. Technical Appendix C outlines general questions and issues, and presents evidence for results. This will be an area for further study, with the goal of developing a clear method by which level-2 propensity scores can be used to estimate level 1 effects.

high teacher retention, is defined as a school with over 85% of the teachers retained from year to year.³ The propensity score is the probability of a school receiving the treatment (high teacher retention), given a set of covariates. The weighting approach allows all information to be retained in the analysis and not lost as can be the case when doing a case by case matching approach to propensity scores (Hirano & Imbens, 2001; Morgan & Harding, 2006; Robins, Hernan, & Brumback, 2000; Robins & Rotnitzky, 1995; Robins, Rotnitzky, & Scharfstein, 2000) Complete details of the weighting method are provided in the data and methods chapter.⁴ After propensity weights are constructed, each school is weighted by its propensity to receive treatment in school-level regressions.

There are some critiques of this method. When the weighting procedure is used to estimate the average treatment effect, Freedman and Berk (2008) found that the weighting method was optimal if 1) study participants were independent and identically distributed (here, schools); 2) selection was exogenous, and 3) the selection equation was properly specified with correct predictor variables and functional forms. When these are not met, the weighting method may increase random error, and may bias the standard errors downward. Kang and Schafer (2007) demonstrated that using inverse probabilities as weights is sensitive to misspecification of the propensity model. For these reasons, the weighting method is used as one method, but is verified with other methods.

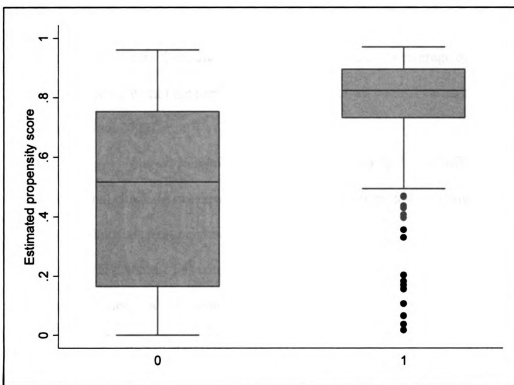
Checking the Weights/Trimming the Weights

³ Tests were done to estimate the best threshold for defining the high teacher retention treatment level. Several thresholds were included in a series of logistic regressions to see which were significant. Additionally, propensity scores were estimated for several thresholds and the distribution of treatment and control cases in each of the blocks was examined. An 85% threshold yielded the most reasonable results, and therefore is used here as the treatment.

⁴ These weights constitute sampling weights, and are applied in this analysis using the pweight command in Stata.

One concern with utilizing the weighting strategy is that extreme values may exert undue influence on the outcome (i.e. one school counting for too many schools). However, this can be addressed by examining a box plot (see Figure 6.1) and trimming the weights. In this analysis, rather than dropping out the observations with extreme weights, observations with weights greater than 18 are assigned a value of 18.

Figure 6.1: Box and Whisker Plot to Demonstrate Distribution of Weights



Checking for Balance

After the propensity score is estimated, checks are performed to ensure that the balancing property is satisfied. If balance is achieved, the treatment assignment and the observed conditions are conditionally independent, given the propensity score. This satisfies the strongly ignorable treatment assignment assumption (Guo & Fraser, 2010). This is done by a series of two-sample t-tests between the treatment and control group, divided into “blocks” by propensity score. Results from these checks for balance are

shown in Table 6.1. There are no significant differences in the means between treatment and control on any of the covariates.⁵

Propensity Scores: Stratification Method

I also investigate this relationship by looking at the average treatment effect within four strata.⁶ This estimates the average treatment effect on the treated (ATT) using stratification matching. The ATT is computed only on the region of common support, and is computed using a weighted (by the number of treated) average of the block-specific treatment effects. In turn these are computed as the difference in average outcomes of treated and controls within the same block for which the all control variables are balanced (Becker & Ichino, 2002).

Propensity Scores: Nearest Neighbor, Kernel Matching and Radius Matching

As a final check on my results, I conduct propensity matching using three additional methods: nearest neighbor matching, kernel matching, and radius matching. In radius matching, the ATT is computed by averaging over the unit-level treatment effects of the treated where the control(s) matched to a treated observation is/are those observations in the control group that lie within a radius of .25; if there are multiple best controls, the average outcome of those controls is used. Kernel matching calculates the ATT by averaging over the unit-level treatment effects of the treated where the control unit outcome matched to a treated observation is obtained as kernel-weighted average of control unit outcomes. Finally, nearest neighbor matching computes the ATT by averaging over the unit-level treatment effects of the treated where the control(s) matched

⁵ This is a function of the pscore estimation method in Stata. The propensity score cannot be estimated if the **balancing** property is not satisfied.

⁶ This is computed using Stata's `atts` command within the `pscore` program. See Becker & Ichino, 2002.

to a treated observation are those observations in the control group that have the closest propensity score; if there are multiple nearest neighbors, such as here, where three neighbors were specified, the average outcome of those controls is used.^{7,8,9}

The outcome variable is a count variable of the number of students in a given school who changed schools between their ninth grade and eleventh grade year. This variable is positively skewed, suggesting the use of a Poisson or negative binomial regression model, but these models do not fit the data well. Therefore, this school-level student mobility variable is dichotomized into high student mobility (greater than 10%) and non-high student mobility, and analyzed using a logistic model for the bivariate and multivariate regressions, and for the weighted regressions. However, logistic models are not a specific option for specification in the other propensity score models; the authors of the program suggest simply using the dichotomous outcome variable as if it were continuous (Becker, 2010; Angrist, 2001). Given that there are known weaknesses in both the weighting method and in these matching methods, I present the coefficients and standard errors from all models, to provide as much evidence as possible regarding the potential effect of high school-level teacher retention on student mobility.

Results

⁷ For a more extensive development of each of these three methods and their strengths and weaknesses, please see Technical Appendix C.

⁸ Nearest neighbor, radius and kernel matching were implemented using Stata's `psmatch2` program. We would expect all estimators to be very similar, as with growing sample size, they come closer to comparing only exact matches (Smith, 2000). In this analysis, however, there are 580 schools; while this is a reasonable sample size, it is by no means very large, and in smaller sample sizes, the choice of matching estimator is important (Heckman, Ichimura, & Todd, 1997).

⁹ To avoid bad matches, I imposed a caliper, which is a tolerance level on the maximum propensity score distance. I chose .25. It is important to note that these findings are very sensitive to the caliper size set. This is a limitation in the caliper method (Smith & Todd, 2005). Also, various calipers were tested, as suggested by Guo & Fraser (2010).

The first step in a propensity score matching analysis is to estimate the propensity score. Only variables that simultaneously influence the participation decision and the outcome and only variables that are unaffected by participation in treatment should be included in the propensity score matching equation (Caliendo & Kopeinig, 2005). Therefore, only variables that are either fixed over time or measured before participation in treatment (i.e high school-level teacher retention) are included. The teacher retention rate is a three year average from 2006-2008, which precludes the use of some variables, like the average percent of professional teachers in each building, from being included in the propensity score estimation. The percent minority students, percent free and reduced lunch students, locale, percent of female students, and charter and magnet status are considered to be time-invariant characteristics. School size category is taken from the 2007 Common Core of Data. To account for student prior performance, the average student pretest scores in math and English language arts are included.¹⁰ The results of the logistic regression to estimate the school-level propensity to have high teacher retention are presented in Table 6.2.¹¹

Schools with a greater proportion of minority students were less likely to have high teacher retention. Large schools (those over 1000 students) were more likely to have high teacher retention. All of the other predictor variables do not meet the threshold for statistical significance.

Table 6.3 presents the results of the propensity score matching analyses under nine different conditions: bivariate, regression with all covariates, weighted by propensity (EOTM, treatment on the treated, treatment on the control), stratification, nearest

¹⁰ These are from the 2005 MEAP, which is prior to the assignment to treatment.

¹¹ All propensity scores are calculated using the pscore program in Stata.

neighbor, kernel matching and radius matching.¹² All of these methods are used to predict the effect of high teacher retention on school-level student mobility by comparing the student mobility in schools with similar propensities to have the “treatment” (high teacher retention) where some schools actually have high teacher retention and other schools do not.

When weighted by propensity to have high teacher retention using the EOTM, schools with high teacher retention are .46 times as likely to have high student mobility, and this effect meets the threshold for statistical significance. Similarly, the treatment effect for the treated shows that schools who got the treatment were 0.49 times as likely to have high student mobility as those of similar propensity who did not get the treatment. Finally, the strongest effect is where we might have expected it—treatment on the control, which suggests how strong the effect of this treatment might have been for those who did not receive it. The treatment effect for the control is 0.42, which implies that schools in the control condition might see a reduction in their likelihood to have high student mobility if they were to have high teacher retention instead of lower teacher retention. For comparison, the unweighted regression with covariates and the unweighted bivariate regression are included. The four additional matching methods all demonstrate the same relationship.

Sensitivity Analyses

If there are unobserved variables that affect assignment to treatment and the outcome variable, there might be ‘hidden bias’ that will not be accounted for by these methods. Given the relatively limited data that we have available, this may be the case

¹² The stratification, nearest neighbor, radius, and kernel matching methods all include bootstrapped standard errors on the average treatment effect.

here. Will the inferences be altered by unobserved factors? The question of interest in sensitivity analyses for propensity scores is slightly different than that asked by sensitivity analyses like those reported in Chapter 5. The researcher wants to investigate how sensitive the assignment to treatment is to hidden bias. Is there an unobserved covariate that is correlated with the likelihood of receiving treatment and with the likelihood of having an increased outcome?

There are two types of bias: overt bias, one that can be seen in the data at hand, and hidden bias, one that cannot be seen because the required information is not available (Rosenbaum, 2002). Hidden bias is unobserved selection. The fact that this bias is unmeasured is crucial, because if it were measured, it could be accounted for and avoided (Kennedy, 2003). Overt and hidden biases that affect causal inference can hypothetically be controlled for in well-implemented randomized experiments, through the mechanism of randomization (Guo & Fraser, 2010). Although researchers control for selection bias through matching, they can adjust only on the observed or measured covariates; thus, selection bias due to unmeasured covariates remains a problem. Therefore, in most observational studies, what remains unknown is the extent to which matching or other adjustments adequately control for bias and yield estimates of treatment effects that are trustworthy. To deal with this problem, Rosenbaum developed sensitivity analysis methods designed to gauge the level of sensitivity of study findings to hidden bias (Guo and Fraser, 2010). Rosenbaum & Rubin (1983) and Rosenbaum (2002, 2005) recommend that researchers regularly conduct sensitivity analyses with observational studies.

The fundamental idea in sensitivity analyses in the propensity score matching context is to manipulate the estimated odds of receiving the treatment to investigate the

extent to which the estimated treatment effects may vary. What the researcher hopes to find is that the estimated treatment effects are robust to a plausible range of selection biases. This idea is developed below.¹³

The participation probability is given by $P_i = P(x_i, u_i) = P(D_i=1|x_i, u_i) = F(\beta x_i + \gamma u_i)$, where x_i are the observed characteristics, u_i is the unobserved variable, and γ is the effect of the unobserved variable on participation decision. If the study is free of hidden bias, γ will be zero and the participation probability will be determined solely by the observed covariates.

If there is hidden bias, two schools with the same observed covariates will have different chances of receiving the treatment. If both units have identical observed covariates, which is what is implied by the matching procedure, then the units only differ in their odds of receiving treatment by a factor that includes the parameter γ and the difference in their unobserved covariates u . If there are no differences in the unobserved variables between the units, or if the unobserved variables have no relationship with the participation in treatment decision, then the odds ratio will be one, which suggests that there is no hidden or unobserved bias.

These sensitivity analyses attempt to evaluate how inferences about the effect of treatment (i.e. high school-level teacher retention) is altered by changing the values of γ (the effect of the unobserved variables on treatment decision) and by differences in the values of the unobserved variables. This is tested using the `mhbounds` program, available in Stata (Becker & Caliendo, 2007). The `mhbounds` program computes Mantel-Haenszel

¹³ Through this section, I rely on the work of Guo & Fraser (2010), as well as that of Becker & Caliendo (2007), and Caliendo and Kopeinig (2005). An application of this type of sensitivity analysis is found in Haviland, Nagin, & Rosenbaum (2007), and is used as a model for this analysis.

bounds to check sensitivity of the estimated average treatment effects on the treated. The Mantel-Haenszel test statistic is used to compare the successful number of schools in the treatment group against the same expected number given the treatment effect is zero. To use this statistic, it is necessary to make individuals in the treatment and control groups as similar as possible, which is accomplished with the matching.

To review: I calculated the propensity score as described above, and then used that propensity score both as a weight and as a method by which to match observations within strata and with each other for the propensity score analysis, as demonstrated in Table 6.3. I then turned to a sensitivity analysis. The `mhbounds` method is appropriate for checking the sensitivity of nearest neighbor matches and the stratification methods. There is not an appropriate sensitivity analysis available for every method of matching.¹⁴

This test provides two test statistics— Q_{MH+} , which is the Mantel-Haenszel statistic with the assumption of an overestimation of the treatment effect. This would occur if we have positive unobserved selection—those schools most likely to have high teacher retention are also most likely to have high student mobility. The Q_{MH-} statistic adjusts the Mantel-Haenszel statistic downward for negative (unobserved) selection, which is where schools with high teacher retention are less likely to have high student mobility. Please note that the second, negative unobserved selection, is the direction we would expect unobserved selection to act in—i.e. those schools likely to have high

¹⁴ Propensity scores are estimated by the `pscore` command in Stata. Stata offers several matching programs, including the suite in `psmatch2`, as well as a series of average treatment on the treated estimators. In Table 6.3, the estimates are drawn from the average treatment on the treated estimators. However, those do not have sensitivity analysis programs associated with them, whereas `psmatch2` does. I used the `psmatch2` matching to implement 1:1 matching, using the propensity score from `pscore`, and also used the `strata` from the `pscore` program with `psmatch2` to generate average treatment effects and then to assess them using the sensitivity analyses described here.

teacher retention are also likely to have low student mobility. For this reason, this statistic is the one of key interest in the sensitivity analyses.

When checking the sensitivity of the estimated average treatment on the treated under the nearest neighbor matching strategy, there is a significant negative treatment effect on the treated, which indicates that schools that receive the treatment (i.e. have high teacher retention) are significantly less likely to have high student mobility. Under the assumption of no hidden bias (which is at $\gamma = 1$), the Mantel-Haenszel test statistic suggests that this study is insensitive to hidden bias, as this result is repeated. However, looking at the bounds under the assumption that we have negative (unobserved) selection (i.e. Q_{MH-}), we see that the result becomes insignificant at 1.4, which means that the confidence interval for the effect would include zero if an unobserved variable caused an odds ratio of treatment assignment to differ between the treatment and comparison groups by 1.4. This shows that the results are somewhat sensitive to deviations in the unconfoundedness assumption. The results of the positive (unobserved) selection show that the effect is significant at $\gamma=1$, which is the assumption of no hidden bias, and become more significant for increasing values of γ (see Table 6.4).

When evaluating the sensitivity of the average treatment on the treated estimate yielded by the stratification method, there is a similar result. There is a significant negative treatment effect on the treated (see Table 6.5). The bounds under the assumption of negative (unobserved) selection show that the result becomes insignificant at a γ level of 1.2, which implies that the effect would likely disappear if an unobserved variable caused an odds ratio of treatment assignment to differ between the treatment and

comparison groups by 1.2. This implies that these results should be interpreted with caution, as this effect is sensitive to deviations in the unconfoundedness assumption.

It is important to note, however, that this does not conclusively prove that hidden bias exists, but rather that the results appear to be moderately sensitive to possible deviations from the identifying unconfoundedness assumption. This provides evidence regarding the robustness of all the matching methods used here, as each method yields very similar results (as demonstrated in Table 6.3), and specifically in defense of the nearest neighbor and kernel matching. However, this does not address the weighted methods.

To quantify the sensitivity of the inference in the weighted regressions, I return to the Frank (2000) method. This asks the question: how large would the impact of an unmeasured confound need to be to invalidate the inference regarding high school-level teacher retention and school-level student mobility, when schools are weighted by their propensity to receive treatment? Rather than testing the sensitivity of the assignment to treatment, we return to focusing on the sensitivity of the effect and seek to quantify the robustness of that inference to the impact of an unmeasured confound.

Using the school-level outcome and aggregated predictors used in the analysis, I regressed the treatment on the predictors, high school-level teacher retention, to obtain the R^2 from that relationship, and then regressed the outcome on the predictors, without the treatment. These R^2 values were then used to calculate the ITCV for each of the three weighting schemes—the estimated treatment on the margin of indifference (EOTM), the treatment on the treated (TOT) and the treatment on the control (TOC). As noted previously in Chapter 5, when conducting a cluster-level regression for sensitivity

analyses, clusters should be weighted by precision, where precision is calculated as below:

$$\text{Precision} = \text{Weight} = \frac{1}{\tau + V_j} \text{ where } V_j = \frac{\sigma^2}{n_j} \text{ and where } \sigma^2 = np(1 - p)$$

For comparison, I calculated the ITCV using the product of the propensity weight and the precision weight. However, the ITCV estimates were the same and therefore only one set of estimates are produced here.

The ITCV for the EOTM is .05. This implies that the correlations between the unmeasured confound and high teacher retention would have to be .23 or larger, and the correlation between the unmeasured confound and school-level student mobility would have to be .20 or larger to invalidate the inference under the EOTM weighting strategy. In the TOT regression, the ITCV was .04, and the correlation between the unmeasured confound and the treatment would have to be .20, and between the unmeasured confound and the outcome would have to be .21 to invalidate the inference under the TOT strategy. Finally, in the TOC regression, the ITCV again was .04, and the correlations would have to be .22 and .19, respectively. This is a moderate correlation size, especially when comparing it to the size of other measured confounds (see Chapter 5).

Conclusions from Propensity Score Sensitivity Analyses

Sensitivity analyses can be complicated, particularly when applied over a variety of analytic situations, such as here. However, sensitivity analyses should be viewed less as absolute estimates and more as accumulations of evidence. This is the approach taken here—to estimate the treatment effects using a variety of matching algorithms and

methods, and then to assess those methods for sensitivity to unmeasured confounds using all available methods to do so.

The evidence suggests that the observed average treatment effects are somewhat robust to unobserved covariates, but not overly so. Again, this does not mean that there unobserved covariates do or do not exist, but simply that the results here appear to be somewhat robust to deviations from the identifying unconfoundedness assumption.

More importantly, from a substantive perspective, there is reasonable evidence that increasing school-level teacher retention rates can lead to decreases in student mobility. As shown in Chapter 5, increased student mobility has positive implications on achievement. Therefore, it may be a valid policy initiative to look for programs that help to stabilize a teaching force within a given school, or to incentivize teachers to remain within that school.

Limitations of this Analysis

The key limitation in the utilization of a propensity score method is the same problem that analysts face in observational studies: the issue of “hidden bias” from potential unobserved variables. A fundamental assumption of propensity score matching is the assumption of unconfoundedness (Rosenbaum & Rubin, 1983) (otherwise known as the ignorable treatment assumption). This assumption states that, conditional on the set of covariates, the assignment to treatment and control conditions is independent of the outcome of treatment. Using the propensity score allows for the control of all *observed* characteristics, in order to assure that the unconfoundedness assumption is met. However, the question is—could the confound, the variable that affects both assignment to treatment (in this case, high school-level teacher retention) and the outcome (high

school-level student mobility), be an *unobserved* characteristic? If any of the relevant covariates are unobserved, the propensity score matching will produce biased results (Arpino & Mealli, 2008).

I do attempt to address this criticism through the use of sensitivity analyses. These are able to quantify the impact necessary of that unobserved covariate to invalidate the assumptions. The substantive question remains, are these data rich enough to merit this type of analysis? These are relatively limited data—rich in sample size, but limited in the number and type of variables. Important factors, such as school leadership and school culture type of indicators, are not measured. In this way, although propensity scores represent an improvement in the framework to evaluate effects, they are still limited by the covariates available to estimate the propensity score equation, and therefore are still limited by the question of observable factors. If the unobserved characteristic that could invalidate inferences in a multilevel modeling framework is also one that could affect both the assignment to treatment and the outcome, then it still cannot be measured and accounted for in the propensity match itself and therefore still cannot be taken into account. The results will still be biased, because that unobserved characteristic is *unobserved*, and therefore, unmeasurable and uncontrollable.

This can be a rather bleak conclusion. However, even if this is the case, Guo & Fraser (2010) suggest that the matching-based models offer improvements over OLS based methods. Furthermore, the use of multiple methods in tandem with sensitivity analyses produce the best estimates possible from a set of observational data. Finally, it is incumbent upon the education research community to use available skills, methods and

knowledge to investigate a problem, and not to allow the limitations of data to preclude us from providing the evidence we can make available to the research community.

Case Study Schools: An Exploratory Analysis

To further interrogate the differences in outcomes for similar schools using a more descriptive approach, I identified the schools with a very low propensity to have high levels of teacher retention, and then identified those schools that had high retention, despite their low propensity, in order to conduct a descriptive analysis. There were 106 schools with propensity scores lower than .50—in other words, a propensity of less than .50 to have high retention. Of those, 20 had high teacher retention while the other 86 did not. While the overall analysis provides useful information on the entire sample of schools, I wanted to focus in specifically on those schools that were highly unlikely to have high teacher retention who did in fact have it, to see if there were any interesting or informative differences.

In a basic descriptive analysis, I analyzed all of the schools with a less than .50 propensity to have high teacher retention as “case study schools.” For this analysis, I refer to “high retention case study schools” and “low retention case study schools” in reference to only those 106 schools identified as having a propensity of high teacher retention that is less than .50. See Table 6.6.

The high retention case study schools are more likely to be town or rural schools. High retention case study schools had, on average, 55% of their student body as minority students, compared with 69% in the low retention (control) case study schools. However, when comparing case study schools overall with non-case study schools, case study schools (regardless of high or low teacher retention) were much more likely to have

higher minority student populations. Non case-study high schools had an average of 14% of their student body as minority students. Therefore, high retention case study schools had higher rates of minority students than the general population of schools, but lower rates than low retention case study schools.

High retention case study schools had the same average proportion of students with free and reduced lunch as low retention case study schools. Again, case study schools overall had much higher average rates of free and reduced lunch (60% in case study schools, 31% in non-case study schools).

High retention case study schools had higher mean proportion of teachers with professional licenses than low retention case study schools (72% versus 65%), and a lower mean proportion of minority teachers (28% versus 34%). They also had a higher mean of high qualified teachers. None of these differences meet the threshold for statistical significance.¹⁵

Turning to the individual characteristics of students in these schools, there were higher than expected proportions of American Indian and Hispanic students in the high retention case study schools. Between high retention case study schools and low retention case study schools, the proportion of black students in each group was close to the expected value. However, when comparing all case study schools to non-case study schools, the proportion of black students was much higher than expected in case study schools. In other words, case study schools (those with a low propensity to have high

¹⁵ Although they did not meet the statistical significance threshold, the “universe” nature of these data makes simple percentage comparisons informative, as differences are “real” differences and not simply due to sampling bias.

teacher retention, regardless of actual teacher retention level) are disproportionately attended by black students.

The math scale scores were, on average, higher in these high retention case study schools than low retention case study schools, and this effect is statistically significant. High retention case study schools also had higher mean ELA scale scores. Students in high retention case study schools also change schools at a lower rate than those in low retention case study schools.

The multilevel models outlined previously were re-run using only these 106 schools, but there was not a detectable effect of high school-level teacher retention on student mobility under any of the three weighting strategies.

For comparative purposes, I also include in Table 6.6 the characteristics of non-case study schools, for comparison with case study schools. Case study schools were much more likely to be city schools than expected (55% of city schools were case study schools, compared with an expected value of 18%). Case study schools were also more likely to be small schools. The mean percentage of minority students was 66% in case study schools, compared with 14% in non-case study schools, and the mean percentage of students eligible for free/reduced lunch was 60% in case study schools, compared with 31% in non-case study schools. Case study schools has a statistically significantly smaller percentage of teachers with professional licenses (66% in case study schools compared with 87% in non-case study schools), and also had significantly higher mean percentages of minority teachers and new teachers. They were more likely to be charter schools, and had significantly higher mean rates of student mobility.

From the perspective of student characteristics, the students in case study schools were overwhelmingly black students—48% of students in case study schools were black, compared with an expected value of 10%. Case study schools had significantly lower mean mathematics and ELA achievement, and significantly lower mean prior math and ELA achievement.

With the case study schools, then, we see an interesting phenomenon. Within the 106 case study schools, the schools are rather similar in terms of structural characteristics. They differ significantly, on achievement and on student mobility. However, the case study schools, taken as a whole, are much different than the non-case study schools. Treatment case study schools look more like non-case study schools than control case study schools, but are still distinctly different in terms of student composition and achievement.

The fact that all of the case study schools, regardless of treatment or control, are more alike than case study to non-case study schools suggests that there is something unique about the school culture of these schools that influences their ability to retain teachers and to perform more highly on achievement tests than anticipated. While this analysis cannot go deeper in terms of measuring and investigating elements such as school culture, professionalism, and charismatic leadership, this is further evidence to suggest the importance of those factors.

Table 6.1: Testing for Balance in Propensity Score Estimation

	Block 1		Block 2		Block 3	
	<i>Control</i>	<i>Treatment</i>	<i>Control</i>	<i>Treatment</i>	<i>Control</i>	<i>Treatment</i>
N	49	7	27	18	60	115
Avg prop score	0.14	0.13	0.36	0.4	0.64	0.68
Locale						
City	0.73	0.43	0.44	0.44	0.1	0.19
Town	0.02	0	0	0.11	0.05	0.1
Rural	0.06	0.29	0.22	0.17	0.71	0.62
<i>Suburb=reference</i>						
Percent minority	90.8	82.97	61.69	54.41	20.17	21.79
Percent free reduced lunch	66.26	74.33	57.29	52.72	51.93	50.17
School size						
Less than 300 students	0.29	0.43	0.15	0.28	0.24	0.52
Greater than 1000 students	0.26	0	0.3	0.33	0.14	0.19
<i>300-999=reference</i>						
Average math pretest	794.79	793.79	799.76	800.46	804.77	808.02
Average ELA pretest	800.17	797.04	802.82	804.3	804.35	809.29
Charter	0.41	0.29	0.15	0.17	0	0
Magnet	0.06	0.14	0.19	0.11	0.05	0.05
Proportion female students in school	56.1	54.41	56.17	50.58	51.12	53.14
Proportion AmerInd/PI	0.18	0	0.71	0.87	1.87	0.62
Proportion Asian	0.59	3.54	1.25	1.66	0.88	0.99
Proportion black	81.42	74.28	51.52	38.31	10.02	15.86
Proportion Hispanic	7.41	5.04	7.9	10.93	5.12	2.59
Proportion multiethnic	0.28	0.18	0.98	0.37	0.55	0.13
Proportion students in special programs	5.75	15.18	0.37	5.92	1.16	0.21

No statistically significant difference in means detected using two-sample t-tests

Tests for balance performed as part Stata's pscore analytic package

Table 6.1 (cont'd)

N	Block 4		Block 5	
	<i>Control</i>	<i>Treatment</i>	<i>Control</i>	<i>Treatment</i>
Avg prop score				
Locale	21	21	39	94
City	0.56	0.58	0.83	0.86
Town				
Rural	0.03	0.05	0.07	0.08
<i>Suburb=reference</i>	0.05	0.07	0.27	0.2
Percent minority	0.87	0.72	0.32	0.34
Percent free reduced lunch				
School size	8.77	12.34	10.78	11.04
Less than 300 students	38.08	38.17	26.21	23.67
Greater than 1000 students	0.31	0.13	0.1	0.05
300-999= <i>reference</i>	0.03	0.12	0.34	0.49
Average math pretest	808.26	809.91	816.97	816.34
Average ELA pretest	809.45	810.19	816.79	816.5
Charter	0	0	0.02	0.01
Magnet	0.03	0.03	0.12	0.16
Proportion female students in school	47.96	49.86	49.66	49.87
Proportion AmerInd/PI	2.29	1.17	0.48	1.93
Proportion Asian	0.55	0.7	2.35	1.9
Proportion black	2.05	6.82	4.55	3.99
Proportion Hispanic	2.49	2.76	2.72	2.45
Proportion multiethnic	0.21	0.3	0.6	0.78
Proportion students in special programs	3.26	1.28	3.96	1.24

No statistically significant difference in means detected using two-sample t-tests
Tests for balance performed as part Stata's pscore analytic package

**Table 6.2: Logistic Regression for Estimating the Treatment Propensity Score
(Treatment = School-Level Teacher Retention > 85%)**

Independent Variable	Estimate	Std. Error	P value
Percent minority students	-0.06	0.03	0.03
Percent free/reduced lunch students	-0.01	0.01	0.179
Locale			
City	-0.33	0.41	0.417
Town	-0.12	0.42	0.769
Rural	-0.43	0.34	0.212
<i>Suburb=reference</i>			
School Size			
Less than 300 students	-0.28	0.32	0.389
Greater than 1000 students	0.72	0.31	0.021
<i>300-999 students=reference</i>			
Charter	-0.69	0.61	0.258
Magnet	0.56	0.37	0.128
Proportion female students	0.01	0.01	0.994
Proportion asian students	0.09	0.03	0.093
Proportion black students	0.03	0.03	0.186
Proportion Hispanic students	0.04	0.03	0.094
Proportion multiethnic students	0.06	0.04	0.12
Proportion special program students	0.001	0.01	0.927
Average math pretest score	0.001	0.03	0.986
Average ELA pretest score	0.05	0.03	0.093

Table 6.3: Estimated Effect of High School-Level Teacher Retention on Average School-Level Student Mobility

Model	Coefficient	Odds Ratio	Std error	t-ratio	pvalue
Weighted by propensity (EOTM)	-.78	0.46	.14	-2.48	.013
Weighted by propensity (treatment effect for the treated)	-.72	0.49	.16	-2.18	.029
Weighted by propensity (treatment effect for the control)	-.86	.42	.14	-2.60	.009
Unweighted, with covariates	-.57	.57	.16	-2.02	.044
Unweighted, bivariate	-1.62	.19	.04	-8.10	.000
Stratification method (ATT for each strata)**	-0.083		.048	-1.732	.042
Nearest neighbor*	-.091		.046	-1.98	.048
Kernel Matching*	-0.079		.048	-1.63	.103
Radius Matching*	-.09		.042	-2.22	.026

*Estimated with psmatch2

**Estimated with the atts program for stratification matching

Table 6.4: Sensitivity Analyses using Mantel-Haenszel Bounds for Effect on Student Mobility: Nearest Neighbor Matching

Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
1	2.63323	2.63323	0.004229	0.004229
1.05	2.83957	2.43571	0.002259	0.007431
1.1	3.03302	2.244	0.001211	0.012416
1.15	3.21848	2.06125	0.000644	0.019639
1.2	3.39666	1.88665	0.000341	0.029604
1.25	3.56817	1.71947	0.00018	0.042764
1.3	3.73352	1.55911	0.000094	0.059485
1.35	3.89321	1.40501	0.000049	0.080009
1.4	4.04765	1.25669	0.000026	0.104433
1.45	4.19722	1.11372	0.000014	0.1327
1.5	4.34225	0.975715	7.10E-06	0.164603
1.55	4.48304	0.842333	3.70E-06	0.199801
1.6	4.61988	0.713265	1.90E-06	0.237841
1.65	4.75301	0.58823	1.00E-06	0.278189
1.7	4.88264	0.466975	5.20E-07	0.320259
1.75	5.009	0.349271	2.70E-07	0.363443
1.8	5.13226	0.234909	1.40E-07	0.40714
1.85	5.25261	0.123695	7.50E-08	0.450778
1.9	5.37018	0.015456	3.90E-08	0.493834
1.95	5.48515	-0.089971	2.10E-08	0.535845
2	5.59762	-0.053909	1.10E-08	0.521496

Gamma : odds of differential assignment due to unobserved factors

Q_mh+ : Mantel-Haenszel statistic (assumption: overestimation of treatment effect)

Q_mh- : Mantel-Haenszel statistic (assumption: underestimation of treatment effect)

p_mh+ : significance level (assumption: overestimation of treatment effect)

p_mh- : significance level (assumption: underestimation of treatment effect)

Table 6.5: Sensitivity Analyses using Mantel-Haenszel Bounds for Effect on Student Mobility: Stratification Method

Gamma	Q_mh+	Q_mh-	p_mh+	p_mh-
1	1.87972	1.87972	0.030073	0.030073
1.05	2.07905	1.70328	0.018807	0.044258
1.1	2.25903	1.52486	0.011941	0.063647
1.15	2.43158	1.35476	0.007517	0.087748
1.2	2.59735	1.1922	0.004697	0.116592
1.25	2.75692	1.03651	0.002917	0.149981
1.3	2.91079	0.887135	0.001803	0.187503
1.35	3.05939	0.743547	0.001109	0.228575
1.4	3.20313	0.605299	0.00068	0.27249
1.45	3.34234	0.471993	0.000415	0.318466
1.5	3.47736	0.343272	0.000253	0.365697
1.55	3.60845	0.218815	0.000154	0.413397
1.6	3.73587	0.098336	0.000094	0.460833
1.65	3.85986	-0.018424	0.000057	0.50735
1.7	3.98062	-0.131037	0.000034	0.552127
1.75	4.09833	-0.020381	0.000021	0.50813
1.8	4.21319	0.087155	0.000013	0.465274
1.85	4.32534	0.191752	7.60E-06	0.423968
1.9	4.43494	0.293575	4.60E-06	0.384541
1.95	4.5421	0.392775	2.80E-06	0.347243
2	4.64698	0.489493	1.70E-06	0.312246

Gamma : odds of differential assignment due to unobserved factors

Q_mh+ : Mantel-Haenszel statistic (assumption: overestimation of treatment effect)

Q_mh- : Mantel-Haenszel statistic (assumption: underestimation of treatment effect)

P_mh+ : significance level (assumption: overestimation of treatment effect)

P_mh- : significance level (assumption: underestimation of treatment effect)

Table 6.6: Comparison of School and Student Characteristics Between High- and Low-Retention Case Study Schools, and Between Case Study and Non-Case Study Schools

	Case Study Schools: Characteristics of Treatment and Control Schools			Case Study Schools Compared to Non-Case Study Schools		
	Treatment N=20 (19%)	Control N=86 (81%)	N	Case Study N=106 (18%)	Non Case N=474 (82%)	N
Locale						
City	15%	85%	53	55%	45%	96
Suburb	10%	90%	20	12%	88%	161
Town	33%	67%	3	4%	96%	78
Rural	30%	70%	30	12%	88%	245
School Size						
Less than 300 students	24%	76%	34	40%	60%	85
300-999 students	16%	84%	50	16%	84%	307
Greater than 1000 students	18%	82%	22	12%	88%	188
AYP Status						
Fail	18%	82%	61	37%	63%	166
Pass	20%	80%	45	11%	89%	414
Percent minority	55%	69%		66%	14%	
Percent free/reduced lunch	60%	60%		60%	31%	
Percent teachers with professional licences	72%	65%		66%	87%	
Percent new teachers	19%	23%		22%	12%	
Percent minority teachers	28%	34%		33%	2%	
Perct highly qualified teachers	81%	75%		76%	82%	
Charter	18%	82%	27	82%	18%	33
Magnet	30%	70%	10	15%	85%	65
Student mobility	36%	35%		35%	10%	
Prior student mobility	30%	29%		29%	9%	

Table 6.6 (cont'd):	Case Study Schools: Characteristics of Treatment and Control Schools			Case Study Schools Compared to Non-Case Study Schools		
	N=1,856 (19%)	N=7653 (81%)		N=9,509 (10%)	N=87,049 (90%)	
<i>Student Level Characteristics</i>						
Gender						
Male	20%	80%		9%	91%	47,327
Female	19%	81%		11%	89%	49,231
Race*						
American Indian	42%	58%	57	7%	93%	828
Asian	19%	81%	79	4%	96%	2,116
Black	18%	82%	6,755	48%	52%	14,019
Pacific Islander	18%	82%	11	9%	91%	116
White	19%	81%	2,050	3%	97%	75,979
Hispanic	35%	65%	513	18%	82%	2,908
Multi-ethnic	2%	98%	44	7%	93%	592
Pre-test math scale score	810.24	798.07		800.45	815.03	
Pre-test ELA scale score	814.01	803.01		805.16	815.30	
11th grade math scale score	1087.89	1071.58		1074.76	1099.10	
11th grade ELA scale score	1095.18	1080.72		1083.54	1102.91	
Free/Reduced Lunch Eligible	16%	84%	5,230	20%	80%	25,879
Student mobility	15%	85%	2,924	21%	79%	13,823

CHAPTER 7: THE CORRELATES AND IMPACTS OF SCHOOL-LEVEL TEACHER UNDERSUPPLY

The third paper utilizes the demand formula specification and resulting calculations obtained from Chapter 4, as well as the school-specific teacher retention rates obtained from Chapter 5 to address the third core question of the dissertation: what are the predictors of within-school undersupply and how does undersupply affect achievement? This chapter estimates the predictors of undersupply using school characteristics, and also estimates the impact of subject-specific undersupply on student achievement test scores in mathematics, English language arts, and science. Importantly, this analysis will also introduce the school-specific teacher retention rate as a key predictor. The modeling strategy employed in this chapter mimics the one used to estimate the effect of school-level teacher retention on student achievement and student mobility. For the sake of parsimony in the proposal, I will not fully restate the relevant literature that is outlined in the other two papers.

Background to the Problem: Undersupply as an Organizational Characteristic of Schools

The advent of more rigorous graduation requirements, increased instructional demands of NCLB and the upcoming demands related to Race to the Top funding, and labor market pressures away from teaching have led to a situation of potential undersupply of qualified teachers. Many researchers have advanced a teacher shortage thesis that identifies student enrollment increases and teacher retirement increases as a driving factor in teacher shortages. However, following Ingersoll (2001, 2003) and Ingersoll & Perda (2009), I examine teacher shortages from an organizational

perspective. Undersupply can be considered an organizational feature of schools, one that can interact with school structural and compositional factors. While teacher shortages may or may not happen at an aggregate state or national level, individual schools may struggle to adequately staff their schools, particularly in the face of rigorous instructional demands. This undersupply can be exacerbated by teacher turnover. Both of these school-specific aspects of supply, undersupply and teacher turnover, and be considered alongside other factors that shape the culture of the school and, more importantly, how students in that school may perform. Schools that are high in undersupply may have low levels of relational trust (Bryk & Schneider, 2002), or may have high numbers of low-income or minority students.

In a school's supply of teachers, turnover and undersupply are two related but very distinct features. Turnover relates to the frequency with which teachers enter and exit the school. Organizational sociology and the sociology of work have extensively studied the impact turnover has on organizational culture, and in particular, the effectiveness of the organization in achieving its goals (Hom & Griffeth, 1995; Kalleberg & Mastekaasa, 1998; Mueller & Price, 1990) Undersupply is related to turnover—a high rate of turnover can relate to being undersupplied—but distinct in that a school could have a high rate of turnover, but not be undersupplied. An undersupply of teachers relates to the school's ability to staff all its positions adequately, and therefore can be considered a resource question. How does the lack of resources as quantified by qualified personnel relate to a school's culture, its hospitableness as a work environment, and its ability to teach and train children? While teacher turnover and other features of teacher supply and demand have been heavily studied (see Darling-Hammond, 2000; Ingersoll, 1995, 2000,

2001), the relationship of a school-specific undersupply of teachers, coupled with teacher turnover in that school, to student achievement and mobility has not been studied. This analysis also moves teacher undersupply away from a general overall estimate of teacher supply into a school-specific characteristic and quality of individual schools, focusing on undersupply as an organizational characteristic rather than a more meta-level economic condition.

There are two relevant questions that arise when considering school-level teacher undersupply. The first is, are undersupply rates distributed unevenly over school characteristics? Are there certain types of schools that are more likely to be undersupplied than other types of schools? Particularly in a situation where a state may appear to have an adequate number of teachers to meet instructional demands, are there schools that are still struggling to recruit and retain staff so that they can provide appropriate instructional to all students? This has implications for the equitable distribution of teachers, as well as for closing achievement gaps, both key aspects of both NCLB and the new Race to the Top funding.

The second important question is, once each school's level of undersupply is ascertained, what is the relationship between that undersupply as a school characteristic and student achievement outcomes? Do students in undersupplied schools perform more poorly than students in adequately supplied schools? One would assume this to be the case; it is difficult to imagine that the learning experience and achievement of students would be better in a school where there are not enough mathematics teachers and class sizes are extremely high or students are forced to take study halls or independent studies in order to obtain instruction. However, this needs to be investigated empirically. If there

is evidence to suggest that undersupply negatively relates to student achievement, then it becomes incumbent upon districts and the state itself to ensure that each school has, at minimum, an adequate *number* of teachers to meet instructional demands.

Analytic Methods

Part I: Model predicting undersupply

Utilizing both the tested demand formula, and the refined supply calculations from the second paper, potential undersupply will be estimated for each school in the state of Michigan. The first stage in the analysis examines the predictors of undersupply using a school-level ordinary least squares regression analysis and a logistic regression analysis. Tables 7.1, 7.3 and 7.5 present the results of OLS regressions predicting undersupply in mathematics, English language arts, and science, respectively, with undersupply as a continuous outcome; Tables 7.2, 7.4, and 7.6 present the same analysis, but with the outcome as dichotomous, where only those schools with greater than one FTE of undersupply are identified as “undersupplied.”¹ Since undersupply is a continuous measure centered around zero, with positive numbers indicating increasing amounts of undersupply, using it as a continuous outcome represents the relationship between a decreasing supply of teachers (relative to demand) and the various predictors. A school may have an undersupply of -2, which means they are actually oversupplied, given the assumptions of the formula. This analysis estimates the impact of other school-level characteristics on that undersupply number, so even if a school is not “undersupplied” in terms of the cut point of one FTE or greater, they still may have lower supply.

¹ Social studies is not represented here, as Michigan’s social studies achievement test is one of it’s least robust, and as social studies is not a matter of key focus in many federal and state initiatives.

The adequate supply of teachers across schools is strongly related to the equitable distribution of teachers. Are schools that serve low-income or minority students more likely to be undersupplied? Are schools in certain locales more likely to struggle to adequately staff their courses to meet the Michigan Merit Curriculum? While this analysis cannot estimate teacher quality, an important predecessor to quality is ensuring that schools have an adequate *quantity* of certified teachers to provide instruction in the MMC areas, without excessively large class sizes.

Findings: Predictors of Subject-Specific Undersupply

The goal is to investigate the type of school-level characteristics that are associated with a subject-specific undersupply—in this case, mathematics. Table 7.1 presents the results of the OLS regression predicting mathematics teacher undersupply. Small schools have significantly lower levels of mathematics teacher undersupply and large schools have significantly higher levels of mathematics teacher undersupply than schools with 300-999 students. Rural schools have lower rates of undersupply than their suburban counterparts. Schools with a higher average proportion of minority teachers have higher rates of mathematics teacher undersupply, with each one unit increase in the proportion of minority teachers associated with a corresponding .02 unit increase in mathematics teacher undersupply. Interestingly, schools with higher proportions of female students appear to have higher rates of mathematics undersupply, with each one unit increase in the proportion of students in a school who are female associated with a .02 increase in math teacher undersupply (significant at the .001 level). Also interestingly, schools with higher proportions of Asian, black, and Hispanic students

appear to have lower rates of mathematics teacher undersupply, relative to the proportion of white students in a school.

Using the dichotomous outcome variable allows for the estimation of the likelihood of being “significantly” undersupplied in mathematics, where significant is defined by an undersupply of greater than one FTE (see Table 7.2). Small schools are much less likely to be significantly undersupplied in mathematics, with an odds ratio of .10, which means that small schools are .10 times as likely as schools with 300-999 students to be significantly undersupplied in mathematics. Large schools are four times more likely to be significantly undersupplied in mathematics than schools with 300-999 students. This suggests that adequately staffing a school with mathematics teachers is highly dependent on school size. Schools with higher proportions of student mobility are also more likely to be significantly undersupplied, and again, schools with higher proportions of female students are more likely to be significantly undersupplied in mathematics.

In English language arts, the school size relationship observed in the mathematics undersupply predictions appears to be less salient, and the key predictors are school instructional staff compositional characteristics (see Table 7.3 for the results of the OLS regression). Schools with a higher proportion of minority teachers experience higher rates of undersupply in English language arts, while schools with higher proportions of highly qualified teachers have lower rates of undersupply in ELA. Higher school-level prior ELA achievement is related to lower levels of ELA undersupply, with each one unit increase in school-level prior ELA achievement related to a .03 unit decrease in ELA teacher undersupply. The relationships between the racial composition of the student

body and undersupply observed in the mathematics analyses appears to have disappeared, with the exception of schools with higher proportions of black students, which have lower levels of ELA teacher undersupply.

Table 7.4 presents the results of the logistic regression predicting significant ELA teacher undersupply. Again, this model is used to look only at the likelihood of being significantly undersupplied (greater than one FTE of undersupply). Schools with a higher proportion of minority teachers are 1.04 times more likely to be undersupplied than those with lower proportions of minority teachers. Schools with high levels of student prior achievement in ELA are .91 times as likely to be significantly undersupplied, while schools with higher levels of female students, Asian students, and American Indian students are all more likely to be significantly undersupplied.

Turning to science teacher undersupply, schools with higher proportions of students who are free and reduced lunch eligible have higher levels of science teacher undersupply (see Table 7.5 for the OLS regression). Large schools have lower rates of science teacher undersupply. Schools with higher student mobility have higher science teacher undersupply, and schools with higher Asian student populations and higher black student populations have lower levels of science teacher undersupply.

When looking exclusively at significant science teacher undersupply (see Table 7.6), schools with higher proportions of free and reduced lunch eligible students are 1.05 times more likely to be significantly undersupplied in science, and schools with higher proportions of student mobility are 1.02 times more likely to be undersupplied in science.

This analysis identifies several common themes regarding the predictors of teacher undersupply. The first is that schools with higher proportions of minority

students, particularly those with higher proportions of black students, are less likely to be undersupplied in all three subject areas. This is a counterintuitive finding, but suggests that Michigan does not have an inequitable distribution of the supply of teachers to schools that serve minority students. The second is the strong relationship between school-level student mobility and teacher undersupply. In all three subjects, higher student mobility was linked to greater levels of teacher undersupply. Third, it is important to note that school-level teacher retention was not a significant predictor of subject-specific undersupply.

The Relationship Between Teacher Undersupply and Student Achievement

This study utilizes a multilevel modeling approach to estimate the contextual effects of the key predictor of interest, school-specific teacher undersupply in mathematics, English language arts and science. Although fixed effects models can be used to control for school factors without making parametric assumptions, I choose to express these models in a multilevel modeling framework in order to estimate effects simultaneously at both levels. To estimate the relationship between school-level teacher undersupply, school-level teacher retention rate (from Chapter 5), and other school contextual factors and student achievement outcomes, I controlled for student level factors such as prior ability, gender, race, and free or reduced lunch eligibility. The models are random intercept models, which model each school mean as a function of the key predictor of interest, school-level teacher undersupply, and other school covariates. There are no random slopes estimated. It is important to remember that these models consist of students nested within schools; although teacher data is available and used to calculate school-level teacher undersupply and retention, as well as other instructional

workforce characteristic variables, teachers are not currently linked to students in the Michigan data, and thus, estimating the impact of a given teacher's retention or mobility on a given student's achievement outcomes is not possible. Moreover, this study takes an organizational approach, in which it is hypothesized that it is the composition of the instructional workforce and the organizational culture of a school that impact student achievement outcomes, and therefore modeling individual outcomes as a function of school-level predictors is appropriate.

Data Source

The data source utilized for this paper is the same as that used in throughout the dissertation (see Chapter 2: Data and Methods for more information if necessary). This study utilizes a set of rich longitudinal administrative data from the state of Michigan, including data on *all* teachers, students and schools in the entire state, collected longitudinally over a period of four years. The sample of high schools is the same as defined in Chapter 5.

Measures

Main independent variable: The key predictor of interest is school-level teacher undersupply rate in mathematics, English language arts, and science. Teacher undersupply in each subject is calculated by estimating demand using the refined demand formula outlined in Chapter 4, calculating supply from the 2008 Registry of Educational Personnel, and subtracting supply from demand (Chapter 1 and Chapter 4 provide greater

detail on this method). This yields a continuous measure of supply, with positive numbers indicating potential undersupply.²

This analysis utilizes as a key predictor the main independent variable from Chapter 5, school-level teacher retention rate. Teacher churn rate is calculated by the teacher retention rate, the number of teachers in a given school who remain the same from one year to the next.³ Teacher retention rate calculated over four years, three retention time points: school years 2004-2005, 2005-2006, 2006-2007 and 2007-2008. These rates are averaged to generate an average retention rate for each high school in the sample. Retention is based on a teacher remaining in the same school from one year to the next. If a teacher is in multiple schools, they are counted as either a stayer or mover from all schools in which they teach. This calculation is at the person-level.⁴

Dependent variables: The outcome variable for the student achievement models is the scale score on mathematics, English language arts, and science achievement tests from the 2009 Michigan Merit Examination.

School- and student-level covariates: The multilevel models are estimated with the same set of covariates described in Chapter 2 and in Chapter 5. They are briefly listed here, and discussed in greater detail above. School-level covariates include: the proportion of teachers with professional licenses, the proportion of minority teachers, the

² For the analyses predicting undersupply, this variable was used both as continuous and dichotomous. In the multilevel models, it is used in its continuous format only.

³ This method is used rather than calculating the number of “new” teachers in a school in a given year because of the difficulty of defining the denominator for that calculation.

⁴ Like Ingersoll (2001), I analyze all turnovers or departures, and do not distinguish between teacher attrition (from the profession) and teacher mobility (between schools). If a teacher is in the same school from one year to the next, they are retained; if they are not, they have “turned over.” This focuses on the organizational aspect of teacher retention, as the consequences to the organization are the same regardless of whether a teacher leaves the field entirely, or simply moves to another school (Ingersoll, 2001).

proportion of highly qualified teachers, the percent minority students, the percent of students eligible for free and reduced lunch, locale, school size, the mean achievement pretest in each subject, and the student mobility rate. Student-level covariates include student mobility, achievement pre-test in a given subject, free/reduced lunch eligibility, racial/ethnic identification, gender, and special program eligibility.

ANALYTIC APPROACH

This study posits school-level teacher undersupply as a critical factor in explaining variations in student achievement in mathematics, English language arts, and science. Therefore, the first step in the analysis is to calculate subject-specific teacher undersupply for each school. Simple descriptive statistics were then calculated for schools by subject-specific undersupply, and for all student- and school-level variables that are used in the multilevel models. Tables 3.10, 3.11, and 3.12 (in the descriptive analysis chapter) present the mean undersupply rates by school characteristics in each subject (details not repeated here).

The Multilevel Models to Be Estimated

To test the effects of teacher undersupply on student achievement outcomes, multilevel models (i.e., models with random effects) of student achievement in mathematics, English language arts, and science with high school students nested within schools are estimated. These models are the same as those used in Chapter 5, and will not be repeated here in any length.

The general model specification for the HLM models is as follows:

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{student mobility}) + \beta_j Z' + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{undersupply}) + \gamma_{02}(\text{teacher retention}) + \gamma_j \mathbf{Q}' + u_{0j}$$

where

Y_{ij} = outcome (mathematics, ELA or science scale score) for each student i in school j

β_{0j} = each school mean, represented as a function of the grand mean, student mobility, the matrix of student-level predictors, the school teacher undersupply rate, and the matrix of school level predictors

β_{1j} = coefficient for student mobility

β_j = vector of coeffs for school j

\mathbf{Z}' = vector of student covariates for school j

γ_{00} = grand mean (intercept)

γ_{01} = effect of subject-specific undersupply on B_{0j} (each school mean)

γ_{02} = effect of school-level teacher retention rate (from Chapter 5) on B_{0j} (each school mean)

γ_j = vector of school-level predictors)

\mathbf{Q}' = vector of school covariates

u_{0j} = the residual error of β_{0j} , distributed iid $N(0, \tau_{00})$

r_{ij} = level 1 variance (student error term), r_{ij} distributed iid $N(0, \sigma^2)$

One key challenge to this analysis might be that the causal path could be in the opposite direction—that student achievement predicts teacher undersupply. In order to attempt to address that criticism, the longitudinal nature of the data is utilized. Prior individual student academic achievement from their 8th grade pretest, as well as a prior cohort-level mean mathematics achievement from 8th grade are included, to control for both prior individual and group achievement. Teacher undersupply is calculated based on 2008 data, while the student outcome is from the 2009 testing year. Therefore, the school-level undersupply predictor is causally prior to the student outcome.

A baseline for the HLM models is established by estimating an unconditional random effects ANOVA (output not reported). This allows for the calculation of the intraclass correlation, or the proportion of variance that is between schools. Following the

estimation of this baseline model, a series of multilevel models were estimated, for mathematics, English language arts, and science. Four different multilevel models are presented in Table 7.7 (mathematics), Table 7.8 (English language arts), and Table 7.9 (science). The first, a bivariate model, estimates the bivariate relationship between subject-specific school-level teacher undersupply and student achievement in that subject. The second (Model 2) includes a second predictor shown in the preliminary analyses to be highly correlated with the student achievement, free lunch eligibility, as well as a pretest measure of achievement in the appropriate subject and the student mobility indicator at the student level, in order to account for student prior ability. The third model adds in the school-level teacher retention rate. The final model includes all predictors at level 1 and level 2.

Sensitivity Analyses

When conducting research using observational data, there is cause for concern regarding the impact of an unobservable characteristic on the outcome, one that might invalidate the inferences drawn from the study. State administrative data, such as those used in this study, are rich in observations but often does not include a large number of variables. Sensitivity analyses are conducted to test the robustness of the inferences to the influence of other unobserved characteristics, characterizing the robustness of these inferences to the potential impact of confounding variables (Frank, 2000). I utilized this sensitivity analysis strategy in Chapter 5 as well.

RESULTS⁵

Mathematics Achievement Outcomes⁶

⁵ The descriptive statistics are described in Chapter 3 and are not repeated here. Note, however, that when this is submitted as an individual journal article, the descriptive information would be included here.

To assess the relationship between student mathematics achievement and school-level mathematics teacher undersupply, a series of multilevel models are estimated (see Table 7.7). Using the series of multilevel models outlined previously, I find that school-level mathematics teacher undersupply is negatively and significantly associated with school mean mathematics achievement scores. In the bivariate regression (Model 1), schools with higher rates of mathematics teacher undersupply have mean mathematics achievement scores that are 1.09 scale score points lower than students in schools with lower rates of mathematics teacher undersupply ($p \leq .05$). The proportion of variance explained between the unconditional model and the bivariate model is 18%. This effect for teacher undersupply remains in Model 2, with the addition of the measure of student prior ability, the math pretest score, as well as student free and reduced lunch eligibility and student mobility. For each one unit increase in mathematics teacher undersupply, there is a corresponding decrease of 0.46 scale score points in student mathematics achievement ($p \leq .022$). Prior student mathematics achievement, aggregated to the school level, is also an important predictor, with each one unit increase in average school-level student mathematics achievement associated with a 1.23 point increase in mean scale scores. At the student level, mathematics pretest and free/reduced lunch eligibility are both important predictors, with mathematics prior ability associated with higher mathematics achievement outcomes, and free/reduced lunch eligibility associated with significantly lower mathematics achievement outcomes. This model now explains 87% of the variance at level 2, and 53% of the variance at Level 1, which suggests that teacher undersupply and prior school-level mathematics achievement account for much of the

⁶ Results from the baseline model, a oneway ANOVA with random effects, yield an intraclass correlation of 18% for the mathematics achievement outcome.

variation in school mean mathematics achievement. In Model 3, the other key predictor of interest, school-level teacher retention rate, is introduced. This is positively and significantly related to mathematics achievement, with each one unit increase in school-level teacher retention rate associated with a .21 unit increase in mean mathematics scale scores ($p \leq .000$).

In the full model, mathematics teacher undersupply continues to be associated with decreased student mathematics achievement outcome, with a one unit increase in mathematics teacher undersupply leading to a corresponding .35 point decrease in mathematics scale scores. The relationship between school-level teacher retention rate and mathematics achievement has been reduced to statistical nonsignificance, however, although it remains in the same direction. This suggests that, controlling for all other school structural and compositional characteristics, mathematics teacher undersupply is an important predictor of student mathematics achievement, even when controlling for school-level teacher retention rate.

The percent of students in a school who are free and reduced lunch eligible is also linked to decreased mathematics achievement scores. Prior school-level mathematics achievement continues to be an important predictor of current student mathematics achievement. Turning to school structural characteristics, rural schools have higher mean mathematics achievement, with mean scale scores that are 1.43 scale score points higher ($p \leq .009$). In terms of school compositional characteristics, schools with higher rates of teachers with professional licenses have higher mean mathematics achievement, with each one unit increase in the proportion of teachers with professional licenses associated with a corresponding .09 increase in mean mathematics scale scores ($p \leq .002$). Schools

with higher rates of minority teachers have lower mean mathematics achievement, where each one unit increase in the proportion of minority teachers is associated with a .07 decrease in mean mathematics scale scores ($p \leq .001$). At the student level, higher levels of prior mathematics achievement are associated with higher levels of current mathematics achievement ($\gamma_{20} = .90$, $p \leq .000$). American Indian ($\gamma_{40} = -2.17$), black ($\gamma_{60} = -7.11$), Hispanic ($\gamma_{70} = -2.11$) and multiracial ($\gamma_{80} = -2.85$) students all had significantly lower mathematics achievement scores than their white counterparts, while Asian students had significantly higher scores ($\gamma_{50} = 1.52$). Finally, students who changed schools between their freshmen year and the time of the test had significantly lower mathematics achievement ($\gamma_{10} = -3.08$, $p \leq .000$).

English Language Arts Achievement Outcomes

The modeling strategy for the relationship between English language arts teacher undersupply and English language arts achievement is the same as that used to assess the mathematics relationship (see Table 7.8).⁷ In the bivariate model, higher rates of English language arts teacher undersupply is associated with lower ELA scores, with each one unit increase in undersupply associated with a corresponding 1.31 unit increase in mean ELA score ($p \leq .003$). This relationship remains in Model 2, which includes English language arts pretest and free/reduced lunch as key predictors. In Model 2, increased undersupply is associated with decreased ELA score ($\gamma_{01} = -0.34$) while higher levels of prior student level ELA achievement is associated with higher ELA scores ($\gamma_{01} = 1.02$, $p \leq .000$). Free and reduced lunch eligibility is associated with significantly lower rates of mean ELA achievement, with students who are free and reduced lunch eligible

⁷ From the unconditional ANOVA model (output not reported), the intraclass correlation is 20%.

demonstrating ELA scale scores that are, on average, 4.41 scale score points lower than their non-free/reduced lunch counterparts ($p \leq .000$). When adding school-level teacher retention as a predictor in Model 3, we see that each one unit increase in school-level teacher retention rate is associated with a 0.14 unit increase in mean ELA scale scores ($p \leq .000$). Finally, in the full model (Model 4), ELA teacher undersupply is still negatively and significantly associated with ELA achievement scores, with each one unit increase in ELA teacher undersupply associated with a 0.37 unit decrease in mean ELA achievement score ($p \leq .000$). However, the relationship between school-level teacher retention and ELA achievement has been reduced to non-significance, likely due to the inclusion of other workforce composition variables.

Looking at school workforce characteristics, schools with higher proportions of teachers with professional licenses have higher mean ELA achievement scores, while those with higher proportions of minority teachers have lower mean ELA achievement scores. For school structural characteristics, an increased proportion of minority students is associated with higher ELA scores ($\gamma_{06} = .06$, $p = .005$) while the proportion of free and reduced lunch students is associated with lower ELA scores ($\gamma_{07} = -0.14$, $p = .000$). The average school-level ELA pretest score is also associated with higher mean ELA achievement ($\gamma_{015} = 0.94$, $p = .000$). Finally, student characteristics are important in understanding ELA achievement. Student mobility is negatively associated with mean ELA achievement, with students who change schools exhibiting ELA scores that are 2.74 scale score points lower than those that stay in the same school. Free and reduced lunch eligibility is also associated with lower mean ELA achievement scores ($\gamma_{30} = -3.94$, $p = .000$). American Indian, black, Hispanic and multiethnic students all have significantly

lower ELA achievement scores than their white counterparts, while Asian students had significantly higher ELA achievement scores. Female students also had significantly higher achievement scores than male students ($\gamma_{90}=2.02$, $p=.000$).

Science Achievement Outcomes

All of the science teacher undersupply relationships with student science achievement are in the same direction and the same magnitude as the ELA relationships described above, and for the sake of parsimony, are not repeated here. See Table 7.9.

Comparison of Main Effects under Different Demand Estimations

In Chapter 4, the demand formula specification was justified and a formula was selected, which has been utilized here. For comparative purposes, and to see if inferences are altered regarding the impact of school-level teacher undersupply on student achievement, these models were ran under three different demand assumptions: 1) using a 0.7 smoothing constant for enrollment (the optimal formula identified by Chapter 4, and the one used as the basis for all previous analyses in this chapter), 2) the original formula, and 3) a 0.7 smoothing constant on enrollment and distributional assumptions on class size ($\sim N(25, .5)$) and courses taught per FTE ($\sim N(5, 1)$). Table 7.10 compares the main effects from each of these methods for mathematics, Table 7.11 for English language arts, and Table 7.12 for science.

In Table 7.10, we see that the coefficients for the relationship between mathematics undersupply and student mathematics achievement is in the same direction, with similar standard errors, and similar significance levels. The formula with distributional assumptions actually detects a stronger relationship between mathematics undersupply and student mathematics outcomes, which indicates that the primary formula

utilized in this chapter may be a more conservative formula, which helps avoid Type I errors. For English language arts (Table 7.11) and science (Table 7.12), the estimates and standard errors are nearly identical.

Sensitivity Analyses

As discussed in Chapter 5, the question of sensitivity analyses in this context is: to what extent could an unobserved confounding variable alter the inferences regarding the relationship between school-level teacher undersupply in mathematics, English language arts, and science and related student achievement outcomes. To do this, I make use of the Impact Threshold for Confounding Variables (ITCV), to quantify how powerful the impact of an unmeasured confound would have to be in order to negate these inferences (Crosnoe, 2009; Frank, 2000). This method is more fully developed in Chapter 5, and in Technical Appendix D. Sensitivity analyses are presented below for each of the three outcomes (mathematics, English language arts, and science).⁸

The impact of an unmeasured confound (recall $\text{impact} = r_{v \cdot y} \times r_{v \cdot x}$, see Frank, 2000) would have to have magnitude greater than .02 to invalidate the inference regarding mathematics teacher undersupply and its relationship to student mathematics achievement outcomes. Thus to invalidate the inference that high levels of teacher undersupply have a negative impact on student achievement outcomes, a confounding

⁸ These calculations are obtained by conducting school-level regressions with aggregate student variables, and regressing all of the predictors on the treatment (school-level teacher undersupply in each subject) and obtaining the R^2 value, and then regressing all predictors on the outcome without the treatment and obtaining the R^2 value. As the outcome was a student-level outcome, the variables were aggregated to average achievement scores and the continuous teacher undersupply predictor was used. Seltzer, Kim & Frank (2006) suggest that when performing a cluster-level regression with unbalanced data, precision weights should be used. These weights are calculated by $1/\tau + V_j$, where $V_j = \sigma^2/n_j$. See Technical Appendix D for more detail.

variable would have to be correlated with teacher undersupply at 0.23 *and* with school-level mathematics achievement at .09. These are small-to-moderate correlations. It is important to note that this assumes that the unmeasured confound is uncorrelated with the measured covariates (Frank, 2000).

Turning to English language arts, the findings are very similar. For English language arts, the impact would have to be greater than .03 to invalidate the inference between school-level ELA teacher undersupply and student ELA achievement. The confounding variable would have to be correlated with ELA teacher undersupply at .25 and with student ELA achievement at .25.

In science, the impact would have to be greater than is .02. An unobserved confound would have to be correlated with science teacher undersupply at .22 and with student science achievement at .09.

To put these numbers in context, it is helpful to compare the threshold for an unmeasured confounding variable with the impacts for *measured* covariates. One of the strongest predictors is the school-level aggregate of student mobility. The impact of school-level student mobility on the coefficient for school-level teacher mathematics undersupply on student achievement is .07, which is the product of the correlation with mathematics teacher undersupply (.09) and the correlation with school-level mathematics achievement (-.17). Therefore, we see that the impact of an unmeasured covariate that would invalidate the inference between school-level undersupply and student achievement is smaller than the impact of the key predictors, but not by a large degree. This suggests that, although an unmeasured confound would have to be correlated with

the outcome and the treatment at moderate levels, these levels are in the same range as the strongest measured covariate.

Conclusions and Limitations

The evidence presented here suggests that school-level teacher undersupply is an important factor in student achievement outcomes. This is not necessarily a surprising relationship. It does, however, provide empirical evidence to support initiatives that focus on adequately staffing individual schools, rather than overall teacher supply at the state or national level. Michigan has proposed several of these policies in their recent Race to the Top application, including the Woodrow Wilson Teaching Fellows program (aimed at matching high-quality mathematics and science teachers with areas of high need) and a specific focus on recruiting and retaining high-quality teachers *for schools in which there is an identified need*. It is this matching of qualified teachers to undersupplied schools that is critical, not merely widescale teacher recruitment efforts. This may require incentives or specific supports, such as assistance with getting through human resource “hoops.”

This analysis also suggests that having an adequate supply of teachers is important, regardless of teacher quality. This is not to say that teacher quality is unimportant, but simply that, regardless of quality, providing an adequate quantity of instructional staff to meet instructional needs has a positive impact on student achievement. For many struggling schools, the addition of more people could have beneficial consequences, while the longer-term work of improving quality can continue. It is important to note, however, that this analysis only considers licensed instructional staff; this is not to imply that simply putting a body in the classroom will yield results,

but rather that providing a licensed and endorsed teacher of any career stage or ability level is an important first step in raising achievement. This is a problem that has a solution—find areas of shortage and staff them adequately, and in today’s educational climate, may be an important step to take.

Table 7.1: OLS Regression Predicting Mathematics Teacher Undersupply

	<i>Coeff</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	0.004	0.004	0.363
Racial composition of the school			
Percent American Indian/PI	0.005	0.009	0.565
Percent Asian	-0.015	0.017	0.375
Percent black	-0.01	0.004	0.003
Percent Hispanic	-0.01	0.007	0.051
Percent multiethnic	0.001	0.01	0.933
<i>reference=percent white</i>			
Locale			
City	-0.33	0.18	0.064
Town	-0.2	0.17	0.231
Rural	-0.3	0.14	0.037
<i>reference=suburban</i>			
School Size			
Less than 300 students	-0.52	0.15	0.001
Greater than 1000 students	0.76	0.12	.000
<i>reference=300-999 students</i>			
Average teacher retention rate	-0.01	0.008	0.190
Proportion teachers with prof licenses	0.01	0.006	0.370
Proportion minority teachers	0.02	0.006	0.029
Proportions highly qualified teachers	-0.01	0.004	0.083
Charter	-0.059	0.319	0.853
Magnet	-0.29	0.146	0.045
Average mathematics pretest	-0.02	0.008	0.056
Percent female students	0.02	0.006	0.002
Percent of student mobility	0.004	0.002	.093
<i>Adj R-squared</i>	<i>0.1775</i>		
<i>N</i>	<i>578</i>		

Table 7.2: Logistic Regression Predicting Significant Mathematics Teacher Undersupply

	<i>Odds Ratio</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	1.01	0.01	0.921
Racial composition of the school			
Percent American Indian/PI	1.02	0.03	0.55
Percent Asian	0.99	0.04	0.813
Percent black	0.98	0.01	0.075
Percent Hispanic	0.99	0.02	0.509
Percent multiethnic	0.98	0.05	0.734
<i>reference=percent white</i>			
Locale			
City	1.18	0.45	0.66
Town	0.66	0.24	0.265
Rural	0.61	0.2	0.132
<i>reference=suburban</i>			
School Size			
Less than 300 students	0.1	0.08	.002
Greater than 1000 students	4.25	1.15	.000
<i>reference=300-999 students</i>			
Average teacher retention rate	0.99	0.02	0.915
Proportion teachers with prof licenses	0.98	0.02	0.386
Proportion minority teachers	1.02	0.01	0.112
Proportions highly qualified teachers	0.98	0.01	0.125
Charter	0.49	0.48	0.464
Magnet	0.91	0.31	0.774
Average mathematics pretest	1	0.02	0.954
Percent female students	1.04	0.02	0.022
Percent of student mobility	1.01	0.01	0.012
<i>Pseudo R2</i>	<i>0.2076</i>		
<i>N</i>	<i>578</i>		

Table 7.3: OLS Regression Predicting ELA Teacher Undersupply

	<i>Coeff</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	-0.003	0.01	0.56
Racial composition of the school			
Percent American Indian/PI	-0.002	0.01	0.888
Percent Asian	-0.04	0.02	0.133
Percent black	-0.01	0.01	0.017
Percent Hispanic	-0.002	0.01	0.859
Percent multiethnic	0.0001	0.02	0.998
<i>reference=percent white</i>			
Locale			
City	0.04	0.25	0.873
Town	-0.05	0.23	0.814
Rural	0.25	0.2	0.207
<i>reference=suburban</i>			
School Size			
Less than 300 students	0.22	0.21	.295
Greater than 1000 students	-0.21	0.17	.236
<i>reference=300-999 students</i>			
Average teacher retention rate	-0.01	0.01	0.277
Proportion teachers with prof licenses	0.02	0.01	0.066
Proportion minority teachers	0.02	0.01	0.018
Proportions highly qualified teachers	-0.02	0.01	0.002
Charter	-0.05	0.45	0.902
Magnet	-0.13	0.2	0.528
Average ELA pretest	-0.03	0.01	0.007
Percent female students	0.02	0.01	0.1
Percent of student mobility	0.001	0.003	0.706
<i>Adj R-squared</i>	<i>0.0706</i>		
<i>N</i>	<i>578</i>		

Table 7.4: Logistic Regression Predicting Significant ELA Teacher Undersupply

	<i>Odds Ratio</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	0.99	0.02	0.414
Racial composition of the school			
Percent American Indian/PI	1.06	0.03	0.035
Percent Asian	1.12	0.05	0.012
Percent black	0.98	0.02	0.173
Percent Hispanic	1.02	0.02	0.476
Percent multiethnic	1.02	0.03	0.626
<i>reference=percent white</i>			
Locale			
City	0.73	0.43	0.593
Town	0.14	0.15	0.075
Rural	0.71	0.41	0.551
<i>reference=suburban</i>			
School Size			
Less than 300 students	0.3	0.28	.191
Greater than 1000 students	2.91	1.44	.031
<i>reference=300-999 students</i>			
Average teacher retention rate	1.01	0.04	0.704
Proportion teachers with prof licenses	1.03	0.03	0.336
Proportion minority teachers	1.04	0.02	0.057
Proportions highly qualified teachers	0.98	0.02	0.336
Charter	3.47	5.18	0.404
Magnet	2.37	1.15	0.076
Average ELA pretest	0.91	0.04	0.014
Percent female students	1.05	0.02	0.037
Percent of student mobility	1	0.01	0.918
<i>Pseudo R2</i>	<i>0.2173</i>		
<i>N</i>	<i>578</i>		

Table 7.5: OLS Regression Predicting Science Teacher Undersupply

	<i>Coeff</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	0.01	0.004	0.001
Racial composition of the school			
Percent American Indian/PI	-0.004	0.009	0.676
Percent Asian	-0.102	0.02	0.000
Percent black	-0.01	0.004	0.033
Percent Hispanic	-0.01	0.01	0.106
Percent multiethnic	0.001	0.01	0.962
<i>reference=percent white</i>			
Locale			
City	-0.11	0.17	0.516
Town	0.03	0.16	0.835
Rural	0.09	0.14	0.501
<i>reference=suburban</i>			
School Size			
Less than 300 students	-0.21	0.15	.160
Greater than 1000 students	-0.44	0.12	.000
<i>reference=300-999 students</i>			
Average teacher retention rate	-0.01	0.01	0.505
Proportion teachers with prof licenses	-0.005	0.006	0.424
Proportion minority teachers	0.003	0.006	0.615
Proportions highly qualified teachers	-0.002	0.004	0.579
Charter	-0.131	0.306	0.667
Magnet	-0.056	0.139	0.689
Average science pretest	-0.012	0.008	0.111
Percent female students	-0.002	0.005	0.693
Percent of student mobility	0.01	0.002	0.01
<i>Adj R-squared</i>	<i>0.2387</i>		
<i>N</i>	<i>578</i>		

Table 7.6: Logistic Regression Predicting Significant Science Teacher

	<i>Odds Ratio</i>	<i>Std. Er</i>	<i>P-value</i>
Percent free and reduced lunch	1.05	0.02	0.026
Racial composition of the school			
Percent American Indian/PI	0.82	0.22	0.456
Percent Asian	0.89	0.13	0.429
Percent black	0.98	0.02	0.229
Percent Hispanic	0.99	0.03	0.577
Percent multiethnic	0.41	0.24	0.122
<i>reference=percent white</i>			
Locale			
City	0.89	0.69	0.879
Town	0.56	0.45	0.466
Rural	0.2	0.17	0.064
<i>reference=suburban</i>			
School Size			
Less than 300 students	0.06	0.08	.040
Greater than 1000 students	1.7	1.06	.394
<i>reference=300-999 students</i>			
Average teacher retention rate	0.98	0.03	0.539
Proportion teachers with prof licenses	0.98	0.03	0.57
Proportion minority teachers	1.01	0.02	0.765
Proportions highly qualified teachers	0.99	0.02	0.815
Charter	1.2	1.87	0.909
Magnet	0.73	0.59	0.696
Average science pretest	1.04	0.04	0.347
Percent female students	1.04	0.03	0.125
Percent of student mobility	1.02	0.01	0.002
<i>Pseudo R2</i>	<i>0.2389</i>		
<i>N</i>	<i>578</i>		

Table 7.7: Hierarchical Linear Model Predicting Student Mathematics Achievement as a Function of Mathematics Teacher Undersupply

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>
	<i>(se)</i>		<i>(se)</i>		<i>(se)</i>		<i>(se)</i>	
Intercept γ_{00}	1093.63	.000	1093.52	.000	1093.41	.000	1093.55	.000
	(.56)		(.22)		(.21)		(.18)	
<i>Level 2 Variables</i>								
Mathematics teacher undersupply γ_{01}	-1.09	.058	-0.46	.022	-0.44	.021	-0.35	.028
	(.57)		(.20)		(.19)		(.16)	
<i>Instructional Workforce Characteristics</i>								
School-level teacher retention rate γ_{02}					0.21	.000	0.04	.239
					.04)		(.03)	
Proportion teachers w/prof licenses γ_{03}							0.09	.002
							(.03)	
Proportion minority teachers γ_{04}							-0.07	.001
							(0.02)	
Proportion HQ teachers γ_{05}							0.04	.020
							(.02)	
<i>School Structural Characteristics</i>								
Percent minority students γ_{06}							0.03	.069
							(.01)	
Percent free/reduced lunch students γ_{07}							-0.05	.006
							(.02)	
City γ_{08}							-0.06	.931
							(.65)	
Town γ_{09}							0.71	.239
							(.60)	
Rural γ_{010} (suburb=reference)							1.43	.009
							(.54)	
Small school γ_{011}							-1.14	.124
							(.74)	
Large school γ_{012} Medium school=reference							0.63	.167
							(.46)	
Charter γ_{013}							2.03	.232
							(.169)	
Magnet γ_{014}							0.04	.953
							(.62)	
<i>Group Means for Level 1 Variables</i>								
Mean mathematics pretest γ_{015}			1.23	.000	1.14	.000	0.96	.000
			(.03)		(.03)		(.03)	
Mean different school γ_{016}					-0.03	.004	-0.01	.476
					(.01)		(.01)	

Table 7.7 (cont'd)

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>
<i>Student Characteristics**</i>								
Student mobility γ_{10}					-3.81 (.42)	.000	-3.08 (.42)	.000
Math pretest γ_{20}			0.91 (.01)	.000	0.91 (.01)	.000	0.90 (.01)	.000
Free/reduced lunch eligible γ_{30}			-3.82 (.20)	.000	-3.66 (.20)	.000	-3.16 (.19)	.000
American Indian, γ_{40}							-2.17 (.74)	.004
Asian, γ_{50}							1.52 (.51)	.004
Black, γ_{60}							-7.11 (.39)	.000
Hispanic, γ_{70}							-2.11 (.51)	.000
Multiethnic, γ_{80}							-2.85 (.96)	.003
Female γ_{90}							-0.21 (.14)	.145
Program eligible γ_{10}							-5.42 (.73)	.000
<i>Random Effects</i>								
Level 1 variance component σ^2	805.75	576	381.99	575	381.16	573	378.09	560
Level 2 variance component, τ_{00}	172.69		22.18		19.82		13.52	
Proportion variance explained, Level 1	n/a		53%		53%		53%	
Proportion variance explained, Level 2	18%		87%		89%		92%	

* All level 2 variables are grand mean centered.

**Pretest and same school are group mean centered, with group means included at level 2. All other level 1 predictors are grand mean centered.

Table 7.8: Hierarchical Linear Model Predicting Student English Language Arts Achievement as a Function of English Language Arts Teacher Undersupply

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>
	(<i>se</i>)		(<i>se</i>)		(<i>se</i>)		(<i>se</i>)	
Intercept γ_{00}	1097.75	.000	1097.62	.000	1097.5	.000	1097.66	.000
	(.49)		(.22)		4		(.20)	
<i>Level 2 Variables</i>								
ELA teacher undersupply γ_{01}	-1.31	.003	-0.34	.034	-0.35	.017	-0.37	.000
	(.44)		(.16)		(.14)		(.10)	
<i>Instructional Workforce Characteristics</i>								
School-level teacher retention rate γ_{02}					0.14	.000	-.01	.821
					(.04)		(.03)	
Proportion teachers w/prof licenses γ_{03}							0.09	.005
							(.03)	
Proportion minority teachers γ_{04}							-0.07	.005
							(0.02)	
Proportion HQ teachers γ_{05}							0.03	.077
							(.02)	
<i>School Structural Characteristics</i>								
Percent minority students γ_{06}							0.06	.005
							(.02)	
Percent free/reduced lunch students γ_{07}							-0.14	.000
							(.02)	
City γ_{08}							1.29	.054
							(.67)	
Town γ_{09}							0.30	.636
							(.63)	
Rural γ_{010} (suburb=reference)							0.61	.310
							(.60)	
Small school γ_{011}							0.15	.863
							(.88)	
Large school γ_{012} (medium school=reference)							-.24	.627
							(.49)	
Charter γ_{013}							0.87	.616
							(1.73)	
Magnet γ_{014}							-0.07	.920
							(.65)	
<i>Group Means for Level 1 Variables</i>								
Mean ELA pretest γ_{015}			1.28	.000	1.21	.000	0.94	.000
			(.03)		(.04)		(.04)	
Mean different school γ_{016}					-0.03	.019	-0.02	.093
					(.01)		(.01)	

Table 7.8 (cont'd)

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>
<i>Student Characteristics**</i>								
Student mobility γ_{10}					-3.29 (.37)	.000	-2.74 (.35)	.000
ELA pretest γ_{20}			1.02 (.01)	.000	1.02 (.01)	.000	1.00 (.01)	.000
Free/reduced lunch eligible γ_{30}			-4.55 (.20)	.000	-4.41 (.19)	.000	-3.94 (.18)	.000
American Indian, γ_{40}							-2.71 (.83)	.001
Asian, γ_{50}							2.06 (.71)	.004
Black, γ_{60}							-6.23 (.39)	.000
Hispanic, γ_{70}							-3.45 (.45)	.000
Multiethnic, γ_{80}							-2.77 (.96)	.004
Female γ_{90}							2.02 (.15)	.000
Program eligible γ_{100}							-5.59 (.98)	.000
<i>Random Effects</i>								
Level 1 variance component σ^2	817.17	575	385.47	575	384.84	573	381.36	560
Level 2 variance component, τ_{00}	130.79		22.92		21.82		15.51	
Proportion variance explained, Level 1	n/a		53%		53%		53%	
Proportion variance explained, Level 2	3%		83%		84%		89%	

* All level 2 variables are grand mean centered.

**Pretest and same school are group mean centered, with group means included at level 2. All other level 1 predictors are grand mean centered.

Table 7.9: Hierarchical Linear Model Predicting Student Science Achievement as a Function of Science Teacher Undersupply

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>	<i>Coeff</i>	<i>p-val</i>
	<i>(se)</i>		<i>(se)</i>		<i>(se)</i>		<i>(se)</i>	
Intercept γ_{00}	1098.34	.000	1098.23	.000	1098.17	.000	1098.31	.000
	(.59)		(.21)		(.22)		(.20)	
<i>Level 2 Variables</i>								
Science teacher undersupply γ_{01}	-3.43	.000	-0.87	.000	-0.86	.000	-0.77	.000
	(.58)		(.18)		(.18)		(.17)	
<i>Instructional Workforce Characteristics</i>								
School-level teacher retention rate γ_{02}					0.09	.026	-.02	.661
					(.04)		(.04)	
Proportion teachers w/prof licenses γ_{03}							0.08	.010
							(.03)	
Proportion minority teachers γ_{04}							-0.07	.003
							(0.02)	
Proportion HQ teachers γ_{05}							0.05	.001
							(.02)	
<i>School Structural Characteristics</i>								
Percent minority students γ_{06}							0.04	.040
							(.02)	
Percent free/reduced lunch students γ_{07}							-0.10	.000
							(.02)	
City γ_{08}							0.10	.890
							(.71)	
Town γ_{09}							0.02	.971
							(.62)	
Rural γ_{010} (suburb=reference)							0.13	.822
							(.60)	
Small school γ_{011}							-0.90	.318
							(.90)	
Large school γ_{012} (medium school=reference)							0.31	.500
							(.46)	
Charter γ_{013}							2.71	.105
							(1.67)	
Magnet γ_{014}							0.85	.146
							(.58)	
<i>Group Means for Level 1 Variables</i>								
Mean science pretest γ_{015}			1.34	.000	1.29	.000	1.01	.000
			(.03)		(.03)		(.05)	
Mean different school γ_{016}					-0.02	.067	-0.01	.240
					(.01)		(.01)	

Table 7.9 (cont'd)

	Bivariate		Model 2		Model 3		Model 4	
	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>	<i>Coeff</i> (<i>se</i>)	<i>p-val</i>
<i>Student Characteristics**</i>								
Student mobility (1=different school) γ_{10}					-4.48 (.50)	.000	-3.77 (.51)	.000
Science pretest γ_{20}			0.98 (.01)	.000	0.98 (.01)	.000	0.97 (.01)	.000
Free/reduced lunch eligible γ_{30}			-4.03 (.26)	.000	-3.85 (.25)	.000	-3.43 (.25)	.000
American Indian, γ_{40}							-1.21 (.88)	.168
Asian, γ_{50}							8.58 (.74)	.000
Black, γ_{60}							-6.41 (.49)	.000
Hispanic, γ_{70}							-0.51 (.63)	.414
Multiethnic, γ_{80}							-3.45 (1.71)	.044
Female γ_{90}							0.32 (.18)	.081
Program eligible γ_{100}							-5.62 (1.04)	.000
<i>Random Effects</i>								
Level 1 variance component σ^2	1099.2	576	637.02	575	635.89	573	631.68	560
Level 2 variance component, τ_{00}	184.86		19.08		18.77		13.91	
Proportion variance explained, Level 1	n/a		42%		42%		43%	
Proportion variance explained, Level 2	8%		90%		91%		93%	

* All level 2 variables are grand mean centered.

**Pretest and same school are group mean centered, with group means included at level 2. All other level 1 predictors are grand mean centered.

Table 7.10: Comparison of Main Effects for Math Undersupply on Student Mathematics Achievement Using Different Demand Formula Assumptions

	<i>Coeff</i>	<i>Std Error</i>	<i>P-value</i>	<i>Level 1 variance</i>	<i>Level 2 variance</i>
Enrollment with smoothing constant**	-0.35	0.16	0.028	378.09	13.52
Enrollment with smoothing constant and assumptions on class size and	-0.42	0.15	0.006	378.09	13.69
Original demand formula	-0.35	0.15	0.023	378.09	13.8

**All models are the full models, with all school- and student-level covariates specified, as in Chapter 7*

***Estimate utilized in all undersupply analyses in Chapter 7*

Table 7.11: Comparison of Main Effects for English Language Arts Undersupply on Student ELA Achievement Using Different Demand Formula Assumptions

	<i>Coeff</i>	<i>Std Error</i>	<i>P-value</i>	<i>Level 1 variance</i>	<i>Level 2 variance</i>
Enrollment with smoothing constant**	-0.37	.10	.000	381.36	15.51
Enrollment with smoothing constant and assumptions on class size and					
FTE	-0.38	0.10	.000	381.36	15.67
Original demand formula	-0.39	0.10	.000	381.36	15.64

**All models are the full models, with all school- and student-level covariates specified, as in Chapter 7*

***Estimate utilized in all undersupply analyses in Chapter 7*

Table 7.12: Comparison of Main Effects for Science Undersupply on Student Science Achievement Using Different Demand Formula Assumptions

	<i>Coeff</i>	<i>Std Error</i>	<i>P-value</i>	<i>Level 1 variance</i>	<i>Level 2 variance</i>
Enrollment with smoothing constant**	-0.77	.17	.000	631.68	13.91
Enrollment with smoothing constant and assumptions on class size and					
FTE	-0.74	.16	.000	631.69	13.98
Original demand formula	-0.79	.16	.000	631.69	13.9

**All models are the full models, with all school- and student-level covariates specified, as in Chapter 7*

***Estimate utilized in all undersupply analyses in Chapter 7*

CHAPTER 8: MOVING BEYOND TEACHER DEMAND, SUPPLY, AND UNDERSUPPLY:

CONCLUSIONS AND DISCUSSION

This dissertation investigates each of the three aspects of teacher supply in the state of Michigan: estimating demand, understanding supply and how turnover relates to supply, and teacher undersupply. More importantly, it marshals evidence regarding the relationship between teacher retention and teacher undersupply, and student achievement and mobility. Using multiple rigorous methods, including multilevel modeling with longitudinal data, and propensity score matching, and quantifying the robustness of the inferences to unobserved covariates using sensitivity analyses, this dissertation seeks to provide a wealth of evidence, grounded in rigorous methods, to provide policy-relevant information on teacher labor supply in Michigan. Below, I outline the major lessons learned from each of the sections, and then move to a discussion of the policy implications.

Estimating Teacher Demand

Through an iterative series of analyses that tested various specifications of the demand formula, the importance of using a more sophisticated calculation for student enrollment was identified. For this study, student enrollment is now calculated using three years of enrollment data and a 0.7 smoothing constant in order to more accurately predict enrollment for use in the demand formula. Overestimating enrollment leads to an overestimate of demand, which in turn indicates a greater amount of undersupply than may in fact be occurring.

The second important lesson from testing the demand formula was that overall, there is not differential formula functioning by school types. The distribution of

undersupply over types of schools does not differ regardless of the demand formula specification. Changes in classification of significant undersupply are largely due to the shape of the distributions, and the fact that imposing a 1.0 FTE cut point as the threshold for “significant” undersupply places that cut point at different locations on the varying distributions. In the original formula, more schools fell above that cut point, because the formula was more generous, and also, because enrollment was likely overestimated by using a simple average, particularly in large schools. Therefore schools were classified as undersupplied who may not actually have experienced undersupply. This problem has been corrected with the revised formula by estimating enrollment using a projection with a smoothing constant, rather than a three year average.

The purpose of this formula is to provide a useful tool to practitioners and researchers alike to allow for demand calculations that are sensitive to school-specific variations in demand and supply; that takes into account the context of curricular requirements; and that can be used as a planning tool by practitioners in which they can adjust the assumptions to reflect the current or future conditions in their school in order to inform decision making.

Teacher Retention

School-level teacher retention is clearly an essential characteristic of schools, and is important to consider from an organizational standpoint. This analysis identified a mechanism, student mobility, through which school-level teacher retention appears to affect student achievement. Although teacher retention appears to have a direct effect on student achievement outcomes via an interaction with student mobility, this relationship can be explained with other covariates. The results from the first analysis demonstrate

that remaining in the same school has a positive impact on student achievement outcomes. The second analysis in turn demonstrates that students in schools with higher levels of teacher retention are more likely to stay in the same school, which in turn has a positive impact on achievement. Therefore, teacher retention appears to affect student achievement via the mechanism of student mobility. This suggests that there is an interrelated nature of the two “levels” of churn—teacher and student—and that stabilizing the teacher population can help to stabilize the student population, which in turn can positively impact achievement.

Although not directly measured by this study due to the lack of available variables, the evidence strongly suggests that it is the *culture* of the school, which student and teacher mobility serve as a proxy for in this analysis, which is an important characteristic in student achievement. As much prior research has shown, schools that have higher levels of trust, stronger reciprocal relationships, and positive student-teacher relationships are more successful and are able to support higher levels of student achievement. This analysis suggests that teacher and student mobility are a critical aspect of this school-level culture.

These findings were tested and validated through the use of propensity score analyses, which test whether or not high school-level teacher retention has an effect on student mobility, when controlling for selection bias by matching on observed characteristics. The results from the propensity scores demonstrate that increasing school-level teacher retention rates can lead to decreases in student mobility. As shown in Chapter 5, increased student mobility has positive implications on achievement. Therefore, a valid policy initiative is to identify ways to increase school-level teacher

retention, such as establishing programs that help to stabilize a teaching force within a given school, or to incentivize teachers to remain within that school.

Teacher Undersupply

In mathematics, English language arts, and science, teacher undersupply was related to decreased student achievement. This supports initiatives that focus on adequately staffing individual schools, such as Michigan's new Woodrow Wilson Teaching Fellows program (aimed at matching high-quality mathematics and science teachers with areas of high need) and a specific focus on recruiting and retaining high-quality teachers *for schools in which there is an identified need*. It is this matching of qualified teachers to undersupplied schools that is critical, not merely widescale teacher recruitment efforts.

While the quality of teachers is very important, this analysis shows that, particularly in the context of increased graduation requirements, it is imperative to ensure that *all* schools are adequately staffed as a minimum requirement for supporting schools for success. This is not to say that teacher quality is unimportant, but simply that, regardless of quality, providing an adequate quantity of instructional staff to meet instructional needs has a positive impact on student achievement. For many struggling schools, the addition of more people could have beneficial consequences, while the longer-term work of improving quality can continue. IThis is a problem that has a solution—find areas of shortage and staff them adequately, and in today's educational climate, may be an important step to take.

Policy Implications

As Michigan and other states continue to modify graduation requirements and curricular standards to comply with ever-changing federal mandates, as well as face an economic situation that requires difficult decisions regarding staffing, class size, and school configurations, this analysis provides important information for policymakers. The revised demand formula can be used by the state, districts, or even individual schools to help them assess areas of possible undersupply. It can also be used to test the undersupply implications under various proposed changes. For example, if a school is facing a potential undersupply, they can adjust the class size, or courses taught per FTE, to see the impact on their undersupply. With the modifications outlined here, the formula is a rigorous and effective, yet computationally straightforward, tool for policymakers.

Michigan is also implementing alternative routes to teacher certification, and beginning to focus some efforts on not only recruitment, but on the targeting of teachers to schools. The distribution of both retention rates and undersupply over schools suggests that this is a wise strategy. While Michigan does not face an aggregate undersupply of teachers, or a statewide high teacher mobility rate, there are certain schools that are differentially impacted by these challenges. Particularly as Michigan moves forward with turning around low-performing schools and continuing to focus attention on closing achievement gaps, targeting teaching resources to areas of highest need, and being able to accurately diagnose those areas of need, is vitally important.

Finally, the implications from the teacher retention analyses suggest that Michigan needs to continue and enhance efforts that retain qualified teachers not only in the profession, but specifically in a given school. Finding ways to incentivize teachers to

remain in their schools, especially when those schools are struggling schools, may have strong impacts on improving student achievement.

Looking Ahead

Michigan's changing data landscape will provide a number of exciting analytic opportunities that can be brought to bear on this problem. The first is the newly implemented teacher-student link, and related measures of teacher effectiveness and student growth. As these data become available, teacher effectiveness will be an important predictor and covariate alongside school-level teacher retention and undersupply rates. Additionally, this teacher-student link will allow for the estimation of multilevel models with students nested within teachers, so that an individual teacher's mobility can be evaluated for its impact on their student's achievement. The teacher-student link will also allow for the estimation of school fixed effects models, which are useful in accounting for school-level time-invariant unobserved characteristics. Three-level multilevel models, with students nested within teachers within schools, can also be estimated, which allows for the further partitioning of effects into the appropriate levels.

Another development is Michigan's new electronic transcript information. As of winter 2011, Michigan will begin collecting and making available individual student level data regarding coursetaking and grades. This provides a whole new avenue of research. For example—how do coursetaking patterns differ among students in schools with high rates of teacher mobility?

Finally, through the transcript system and a concerted data effort, Michigan is making major inroads on its P-20 data system, with the inclusion of new postsecondary data. Understanding how teacher retention, student mobility, and teacher undersupply

relate to student preparation for and transition to college is a critical outcome. Are students who have a large number of inexperienced or transitory teachers less likely to transition to college? Do schools with a great deal of instability in teaching force struggle to prepare students for college, and to create a culture that supports college going?

Outside of the bounds of a specific teacher labor supply analysis, the continued use of administrative data to evaluate timely and relevant questions of interest to the state and to the education policy discourse at large is the fourth critical outgrowth of this study, and other work that preceded it. A popular idea is creating a “culture of quality data use,” which will remain just that—an idea—without researchers with appropriate training and a strong interest in education policy taking an active role in pioneering data use, sharing information with the state, and producing rigorous results for dissemination in the academic community.

APPENDIX A

Complete List of Variable Descriptions and Summary Statistics

Table A1: Complete List of Variable Descriptions and Summary Statistics						
Variable	Description	Mean	SD	Min	Max	Chap
Teacher Retention						
Average teacher retention rate	Percentage of teachers who were retained in a school from one year to the next. Averaged over four years (2005-2008), three retention time points	86.20	8.57	0	100	5,6,7
Mathematics undersupply (new demand formula)	Undersupply in mathematics as calculated using the new demand formula. Positive numbers indicate undersupply	0.48	1.2	-5.7	4.5	4,7
English language arts undersupply (new formula)	Undersupply in English language arts as calculated using the new demand formula. Positive numbers indicate undersupply	-0.78	1.58	-15	5.17	4,7
Science undersupply (new demand formula)	Undersupply in science as calculated using the new demand formula. Positive numbers indicate undersupply	-0.77	1.2	-9.1	3.56	4,7
Student Achievement						
Math achievement score	Math scale score on 2009 Michigan Merit Examination, Michigan's high school assessment. Dependent variable in multilevel models	1096.7	31.1	950	1250	5,7
English language arts achievement score	English language arts scale score on 2009 Michigan Merit Examination. Dependent variable in multilevel models	1101	30.8	950	1250	5,7
Science achievement score	Science scale score on 2009 Michigan Merit Examination. Dependent variable in multilevel models	1101.6	35.8	950	1250	5,7
Student mobility (Different school)	Dummy, 1 = different school in 9th grade than 11th grade. Dependent variable in teacher retention multilevel models	0.14	0.35	0	1	5
Math pretest achievement score	Math scale score on 2006 MEAP test, last test administered prior to high school	813.59	24.1	470	952	5,6,7

Table A1 (cont'd)						
English language arts pretest achievement score	ELA scale score on 2006 MEAP test	814.3	21.3	458	908	5,6,7
Science pretest achievement score	Science scale score on 2006 MEAP test	823.76	23.3	485	959	5,6,7
Gender	Dummy, 1 = female	0.51	0.5	0	1	5,7
Racial/ethnic code						5,7
American Indian/Pacific Islander	Dummy, 1 = yes. Combined American Indian and Pacific Islander due to small sample sizes	0.01	0.1	0	1	5,7
Asian	Dummy, 1 = yes	0.02	0.15	0	1	5,7
Black	Dummy, 1 = yes	0.14	0.35	0	1	5,7
White	Reference	0.79	0.41	0	1	5,7
Hispanic	Dummy, 1 = yes	0.03	0.17	0	1	5,7
Multiple race/ethnicity	Dummy, 1 = yes	0.01	0.08	0	1	5,7
Free/reduced lunch eligible	Dummy, 1=yes. Indicates whether a student is eligible for free or reduced lunch.	0.27	0.44	0	1	5,7
Special programs	Dummy, 1 = eligible for Title I, special education, Section 504, Limited English Proficient, or migrant programs	0.01	0.11	0	1	5,7
School Level Variables (n=580 high schools)						
<i>School-Level Student Aggregate Variables</i>						
Aggregate student mobility	School-aggregate of student mobility; the proportion of students in a given school who changed schools between the fall of their freshmen year (2006) and their junior year at the MME test (2009)	14.70	24.5	0	100	5,6
Aggregate mathematics score (2009)	School-aggregate mathematics score. Aggregated for each student in a given school who took the test. Used in sensitivity analyses.	1092.9	14.5	1034	1141	7
Aggregate English language arts score (2009)	School-aggregate ELA score. Aggregated for each student in a given school who took the test. Used in sensitivity analyses.	1096.9	14	997	1147	7

Table A1 (cont'd)						
Aggregate science score (2009)	School-aggregate science score. Aggregated for each student in a given school who took the test. Used in sensitivity analyses.	1097.4	16.9	950	1142	7
Aggregate mathematics pretest score (2006)	School-aggregate mathematics pretest score. Aggregated the 2005 MEAP pretest scores for each student in the school in this analysis (current school, not the school where the test was taken)	810.7	9.82	784	853	5,6,7
Aggregate English language arts pretest score (2006)	School-aggregate ELA pretest score. Aggregated the 2005 MEAP pretest scores for each student in the school in this analysis (current school, not the school where the test was taken)	811.66	8.13	784	838	5,6,7
Aggregate science pretest score (2006)	School-aggregate science pretest score. Aggregated the 2005 MEAP pretest scores for each student in the school in this analysis (current school, not the school where the test was taken)	821.45	10.4	771	863	5,6,7
School-level student mobility (2007 cohort)	School-aggregate of the proportion of the 2007 cohort of MME testers who changed schools since the fall of 2005.	13.02	15.9	0	100	5,6
Proportion female	School-level aggregate of the proportion of students who are female	50.90	8.78	0	100	5, 6, 7
Proportion American Indian/Pacific Islander	School-level aggregate of the proportion of students who are American Indian/Pacific Islander	1.42	5.03	0	64.7	5, 6, 7
Proportion black/African American	School-level aggregate of the proportion of students who are black or African American	15.55	28.8	0	100	5, 6, 7
Proportion Hispanic	School-level aggregate of the proportion of students who are Hispanic	3.60	8.34	0	100	5, 6, 7
Proportion multiracial	School-level aggregate of the proportion of students who are multiracial	0.56	3.93	0	90.3	5, 6, 7

Table A1 (cont'd)						
Proportion Asian	School-level aggregate of the proportion of students who are Asian	1.45	3.13	0	38.5	5, 6, 7
Proportion white (reference)	School-level aggregate of the proportion of students who are white	77.42	30.3	0	100	5, 6, 7
<i>School Workforce Composition Variables</i>						
Percent of teachers with professional licenses	Percentage of teachers who have professional licenses in a school. Averaged over four years (2005-2008)	83.29	12.1	25.2	100	3,5,6,7
Percent minority teachers	Percentage of teachers in a school who are minority teachers. Averaged over four years	7.97	17.5	0	100	3,5,6,7
Percent highly qualified teachers	Percentage of teachers in a school who are highly qualified. Averaged over four years (2005-2008)	80.61	12.5	0	100	3,5,6,7
Percent free/reduced lunch	Percentage of the student body eligible for free/reduced lunch	35.88	19.9	0	99.6	3,5,6,7
Percent minority	Percentage of the student body that is minority students	23.27	30	0	100	3,5,6,7
Locale						3,4,5,6,7
City	Dummy, 1=yes	0.17	0.37	0	1	
Suburb	Reference	0.28	0.45	0	1	
Town	Dummy, 1=yes	0.13	0.34	0	1	
Rural	Dummy, 1=yes	0.42	0.49	0	1	
Charter	Dummy, 1 = school is a charter school	0.06	0.23	0	1	4,5,6,7
Magnet	Dummy, 1 = school is a magnet school	0.11	0.32	0	1	4,5,6,7
School size						3,4,5,6,7
Small school	School has less than 300 students	0.15	0.35	0	1	
Medium school	School has 300-999 students	0.53	0.5	0	1	
Large school	School has over 1000 students	0.32	0.47	0	1	
Retention difference	Difference in retention rates from 2008 to 2006. Used as an indicator is included in analyses to control for the direction of change in average retention rates.	0	0.11	-0.4	0.69	2,5

Table A1 (cont'd)						
<i>Supply, Demand and Undersupply Variables</i>						
Mathematics demand (original formula)	Original formula; used an average three year enrollment and assumed class size and courses per FTE	6.7	4.54	0.17	22.7	4,7
Mathematics demand (new formula)	Modified formula identified for use, based on analyses in Chapter 4. Includes smoothing constant on three year enrollment estimate; assumes class size and courses per FTE to be 25 and 5.	6.49	4.43	0.15	22.1	4,7
Mathematics demand (comparative formula)	Formula tested for possible use; includes assumed distribution on class size and courses per FTE. Not selected as optimal; used for comparative purposes, to test possible changes in estimates	6.52	4.49	0.15	23.2	4,7
English language arts demand (original formula)	Original formula; used an average three year enrollment and assumed class size and courses per FTE	6.7	4.54	0.17	22.7	4,7
English language arts demand (new formula)	Modified formula identified for use, based on analyses in Chapter 4. Includes smoothing constant on three year enrollment estimate; assumes class size and courses per FTE to be 25 and 5.	6.49	4.43	0.15	22.1	4,7
English language arts demand (comparative formula)	Formula tested for possible use; includes assumed distribution on class size and courses per FTE. Not selected as optimal; used for comparative purposes, to test possible changes in estimates	6.52	4.49	0.15	23.2	4,7
Science demand (original formula)	Original formula; used an average three year enrollment and assumed class size and courses per FTE	5.03	3.41	0.13	17.1	4,7
Science demand (new formula)	Modified formula identified for use, based on analyses in Chapter 4. Includes smoothing constant on three year enrollment estimate; assumes class size and courses per FTE to be 25 and 5.	4.87	3.32	0.11	16.5	4,7

Table A1 (cont'd)						
Science demand (comparative formula)	Formula tested for possible use; includes assumed distribution on class size and courses per FTE. Not selected as optimal; used for comparative purposes, to test possible changes in estimates	4.89	3.37	0.11	17.4	4,7
Mathematics supply	Sum of FTEs in mathematics for each school	6.03	4.13	0	24.7	4, 7
English language arts supply	Sum of FTEs in English language arts for each school	7.29	4.94	0	23.6	4, 7
Science supply	Sum of FTEs in science for each school	5.65	3.93	0	21.2	4, 7
Mathematics undersupply (original demand formula)	Undersupply in mathematics as calculated using the original demand formula. Positive numbers indicate undersupply	0.69	1.28	-5.3	5.2	4, 7
Mathematics undersupply (comparative demand formula)	Undersupply in mathematics as calculated using the comparative demand formula. Positive numbers indicate undersupply	0.51	1.27	-4.7	5.2	4,7
English language arts undersupply (original demand formula)	Undersupply in English language arts as calculated using the original demand formula. Positive numbers indicate undersupply	-0.57	1.62	-15	5.62	4, 7
English language arts undersupply (comparative demand formula)	Undersupply in English language arts as calculated using the comparative demand formula. Positive numbers indicate undersupply	-0.75	1.62	-15	4.99	4, 7
Science undersupply (original demand formula)	Undersupply in science as calculated using the original demand formula. Positive numbers indicate undersupply	-0.61	1.19	-8.8	3.74	4, 7
Science undersupply (comparative demand formula)	Undersupply in science as calculated using the comparative demand formula. Positive numbers indicate undersupply	-0.74	1.23	-9.5	3.66	4, 7

Table A1 (cont'd)						
Misclassification Situation 1 & 2	Identifies those schools that are classified as undersupplied by one demand formula and not by another. Dummy variable, 1=misclassified. Situation 1 (original formula) and Situation 2 (new formula; smoothing constant on enrollment estimate)					4
Misclassified		0.06				
Not misclassified		0.94				
Misclassification Situation 2 & 4	Identifies those schools that are classified as undersupplied by one demand formula and not by another. Dummy variable, 1=misclassified. Situation 2 (new formula; smoothing constant on enrollment estimate) and Situation 4 (comparative formula; smoothing constant on enrollment and distributional assumptions on class size and courses per FTE)					
Misclassified		0.08				
Not misclassified		0.92				
<i>Weights</i>						
Propensity score	School-level propensity of having high teacher retention	0.69	0.46	0	1	6
Propensity weight: EOTM (estimate on the margin of indifference)	Propensity weight for the treatment effect for those schools on the margin of indifference. Calculated by $\omega(t,x) = t/e(x) + (1-t)/1-e(x)$	2.04	2.26	1.02	18	6
Propensity weight: TOTW (treatment on the treated)	Propensity weight for the treatment effect on the treated. Calculated by $\omega(t,x) = t + 1-t/1-e(x)$	1.71	2.06	1	18	6
Propensity weight: TOCW (treatment on the control)	Propensity weight for the treatment effect on the control. Calculated by $\omega(t,x) = t/e(x) + 1-t$	1.33	1.15	1	18	6

Table A1 (cont'd)						
Precision weight: mathematics undersupply sensitivity analysis	Weight used in the sensitivity analysis for mathematics undersupply (Seltzer, Kim, & Frank, 2007). Equals $1/\tau + V_j$, where $V_j = \sigma^2/n_j$.	0.005	0	0	0.01	7
Precision weight: English language arts undersupply sensitivity analysis	Weight used in the sensitivity analysis for ELA undersupply (Seltzer, Kim, & Frank, 2007). Equals $1/\tau + V_j$, where $V_j = \sigma^2/n_j$.	0.007	0	0	0.01	7
Precision weight: Science undersupply sensitivity analysis	Weight used in the sensitivity analysis for science undersupply (Seltzer, Kim, & Frank, 2007). Equals $1/\tau + V_j$, where $V_j = \sigma^2/n_j$.	0.004	0	0	0.01	7
Precision weight: teacher retention sensitivity analysis	Weight used in the sensitivity analysis for teacher retention analysis (Seltzer, Kim, & Frank, 2007). Equals $1/\tau + V_j$, where $V_j = \sigma^2/n_j$. Since the outcome is dichotomous, there is no σ^2 produced; this is approximated by $np(1-p)$, where p =proportion of students in a given school who changed schools.	.30	0.01	0.28	0.3	5

**APPENDIX B: ADDITIONAL HISTOGRAMS FROM CHAPTER 4: ESTIMATING TEACHER
DEMAND USING STATE ADMINISTRATIVE DATA: CHALLENGES AND RESOLUTIONS**

TECHNICAL APPENDIX B:
ADDITIONAL HISTOGRAMS FROM CHAPTER 4: ESTIMATING TEACHER DEMAND USING
STATE ADMINISTRATIVE DATA: CHALLENGES AND RESOLUTIONS
Demand Histograms under Varying Enrollment Calculations

English Language Arts

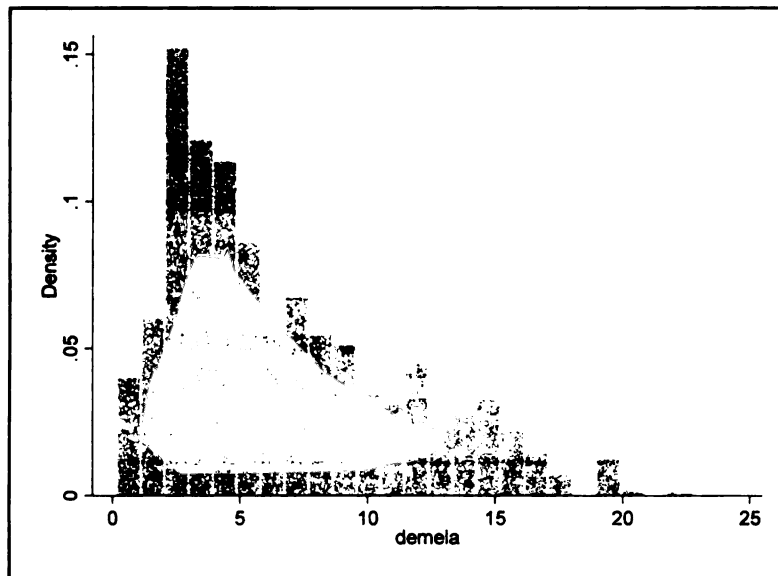


Figure C1: ELA demand distribution: Original formula (average enrollment only)

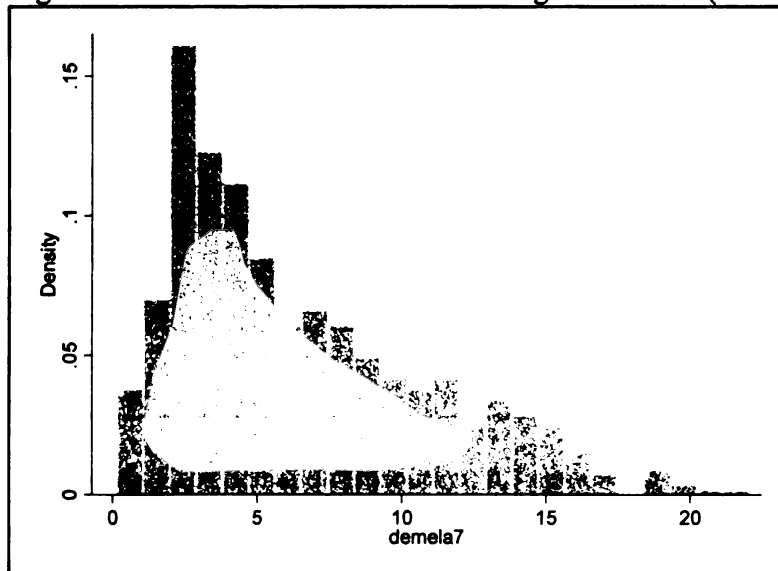


Figure C2: ELA demand distribution: New formula (0.7 smoothing constant on enrollment)

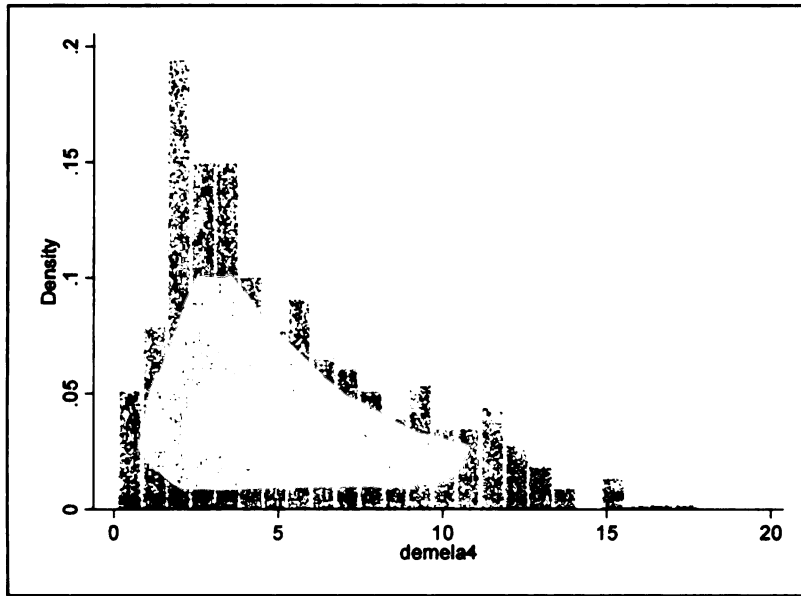


Figure C3: ELA demand distribution: 0.4 smoothing constant on enrollment

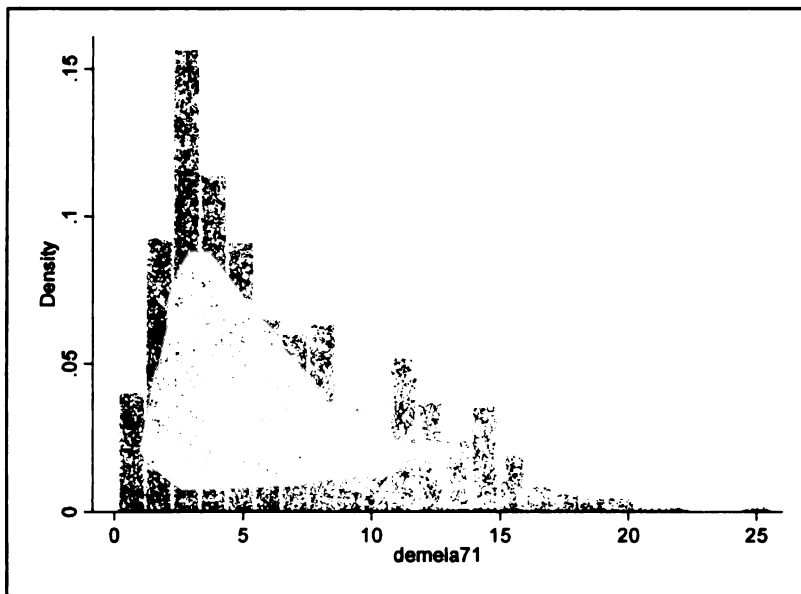


Figure C4: ELA demand distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, 1)$ and courses per FTE $\sim N(5, .5)$.

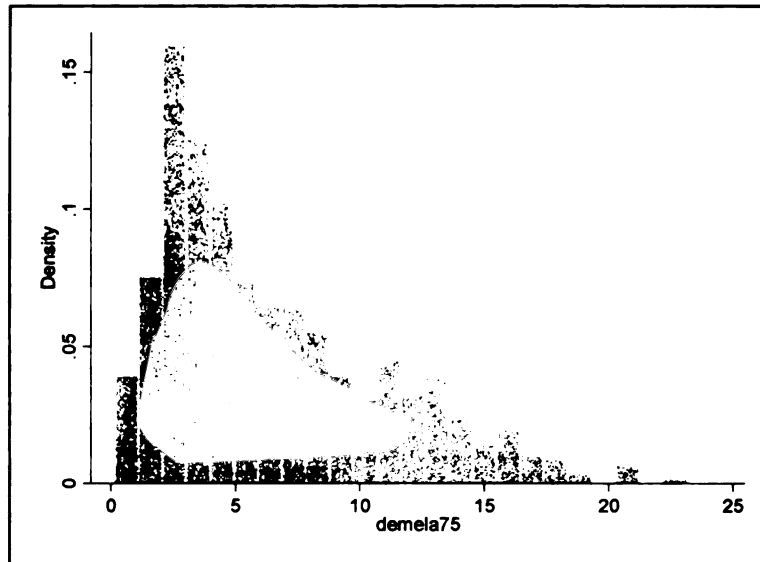


Figure C5: ELA demand distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, .5)$ and courses per FTE $\sim N(5, .5)$.

Science

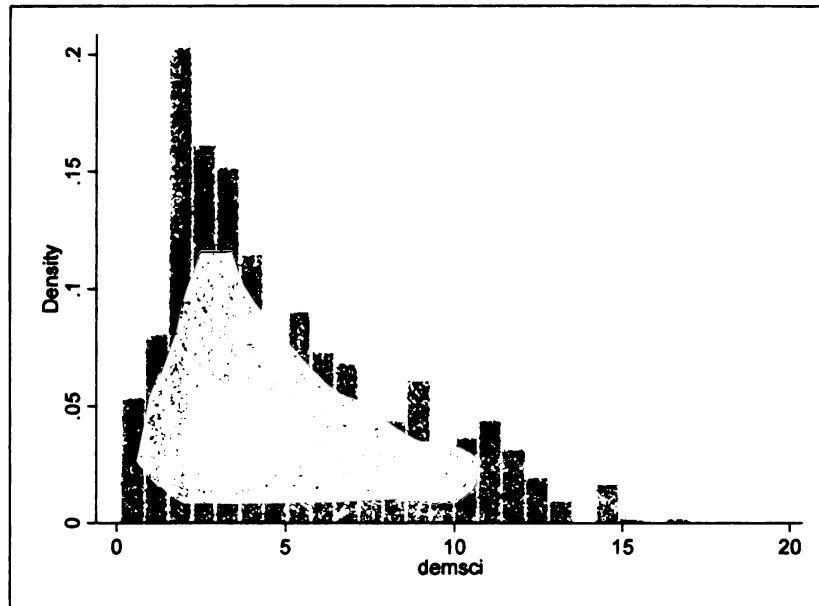


Figure C6: Science demand distribution: Original formula (average enrollment only)

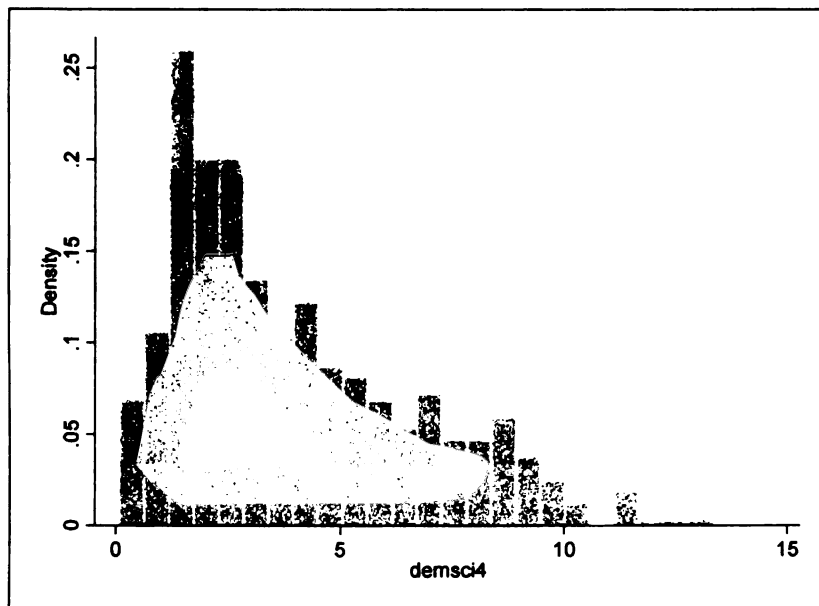


Figure C7: Science demand distribution: 0.4 smoothing constant on enrollment

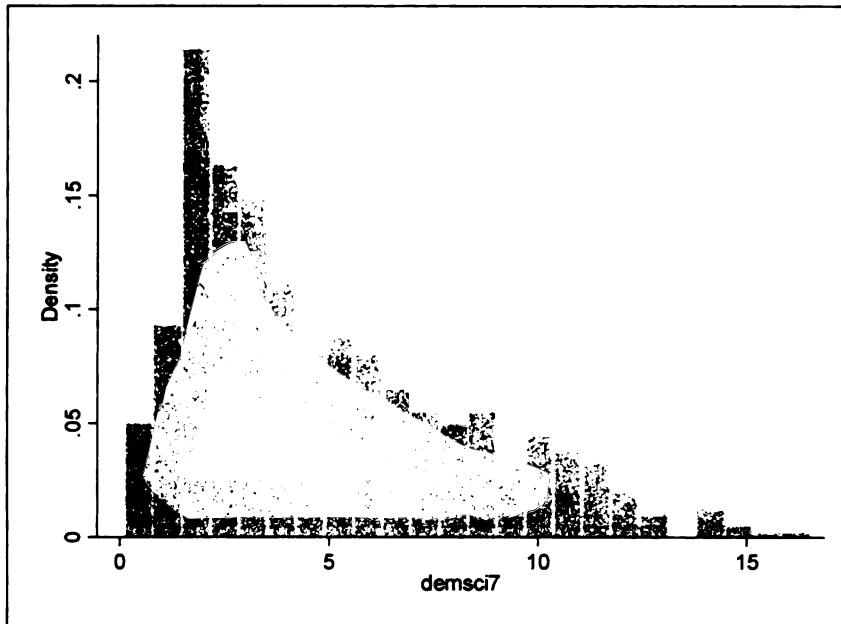


Figure C8: Science demand distribution: 0.7 smoothing constant on enrollment

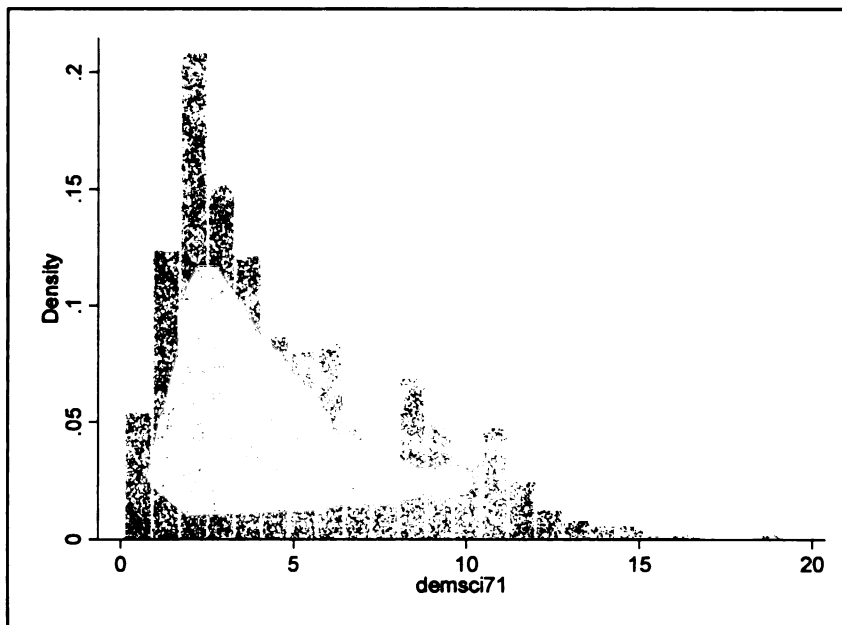


Figure C9: Science demand distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, 1)$ and courses per FTE $\sim N(5, .5)$.

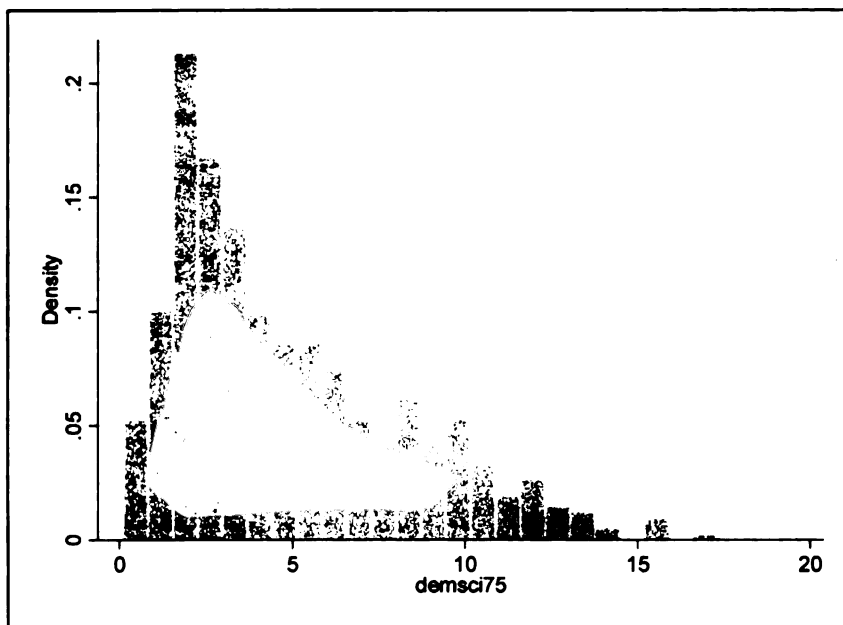


Figure C10: Science demand distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, .5)$ and courses per FTE $\sim N(5, .5)$.

Undersupply Histograms under Varying Enrollment Estimates

English Language Arts

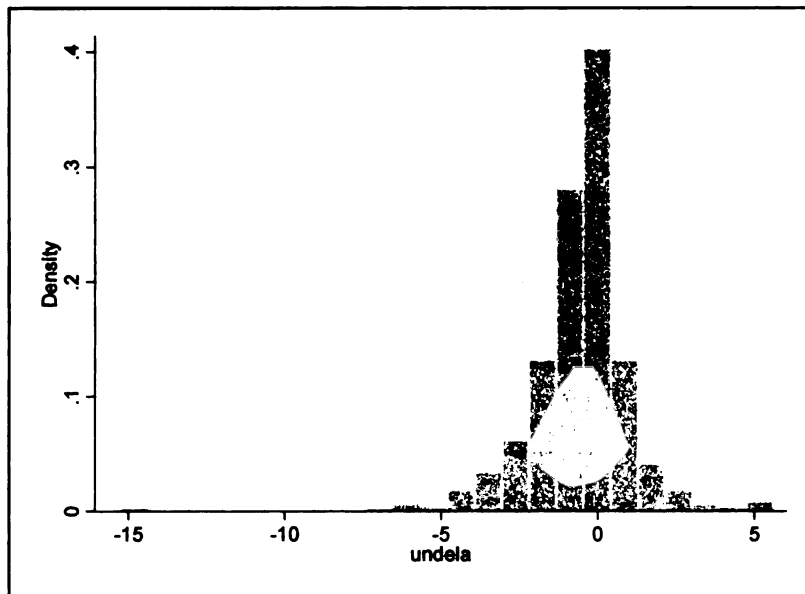


Figure C11: ELA Undersupply distribution: original formula (average enrollment)

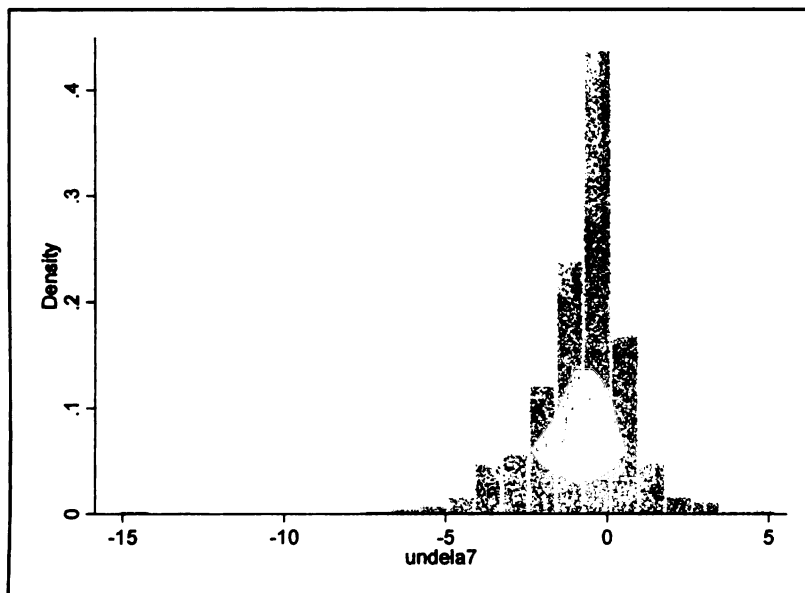


Figure C12: ELA undersupply distribution: 0.7 smoothing constant on enrollment

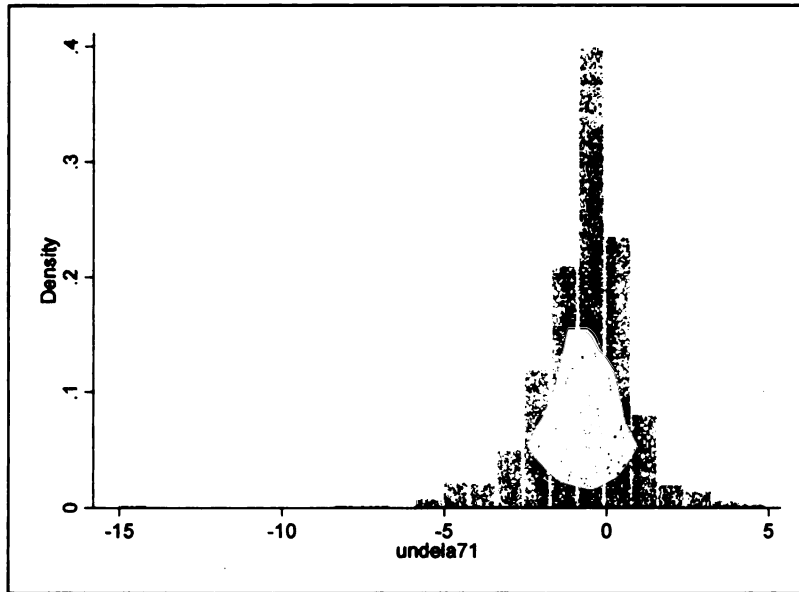


Figure C13: ELA undersupply distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, 1)$ and courses per FTE $\sim N(5, .5)$.

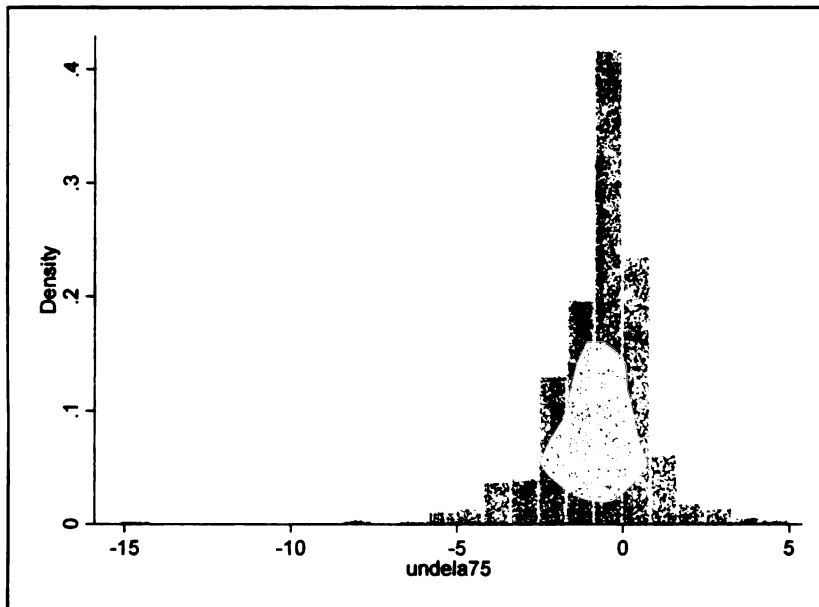


Figure C14: ELA undersupply distribution: 0.7 smoothing constant on enrollment; class size $\sim N(25, .5)$ and courses per FTE $\sim N(5, .5)$

Science

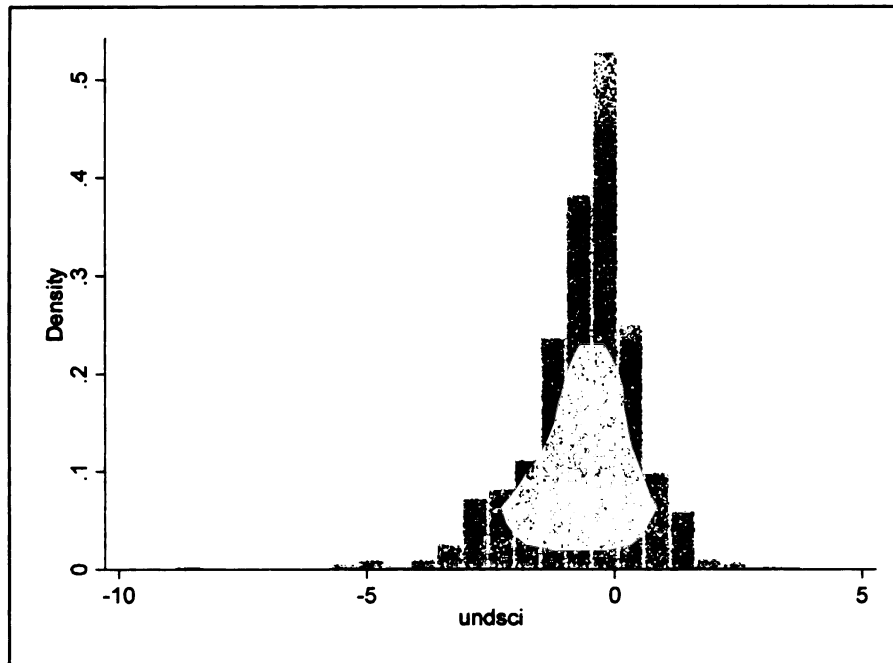


Figure C15: Science undersupply distribution: original formula (average enrollment)

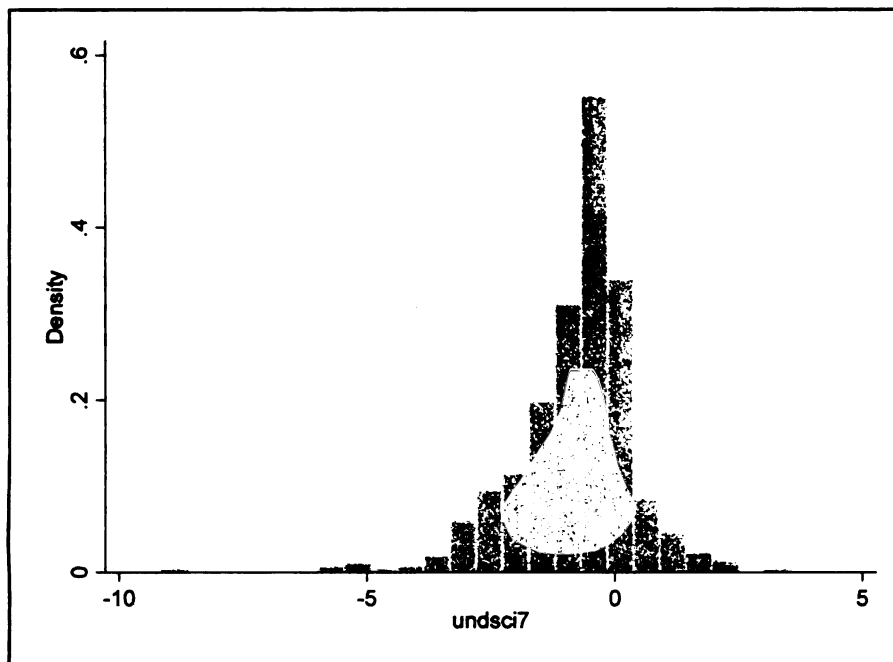


Figure C16: Science undersupply distribution: 0.7 smoothing constant on enrollment

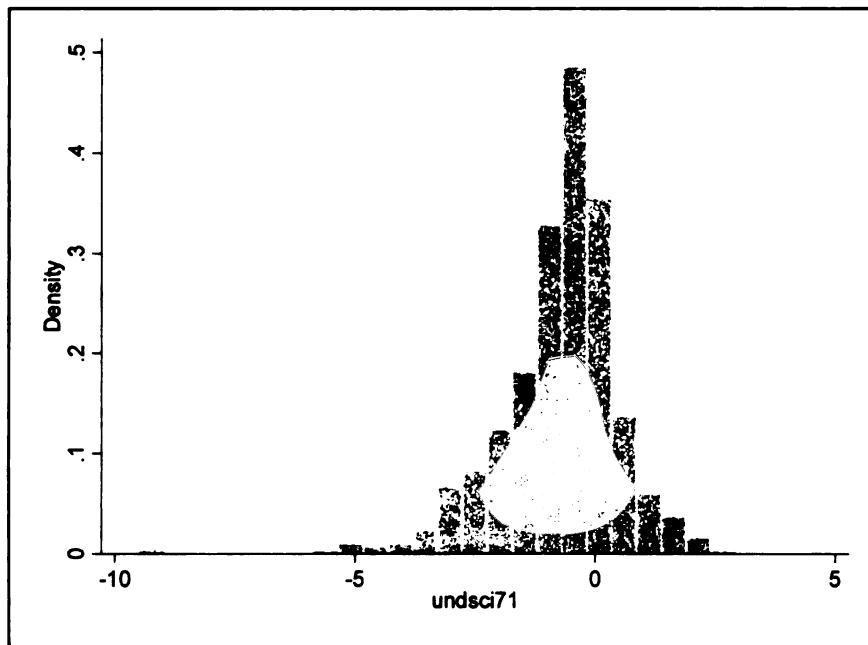


Figure C17: Science undersupply distribution: 0.7 smoothing constant on enrollment;
class size $\sim N(25, 1)$ and courses per FTE $\sim N(5, .5)$

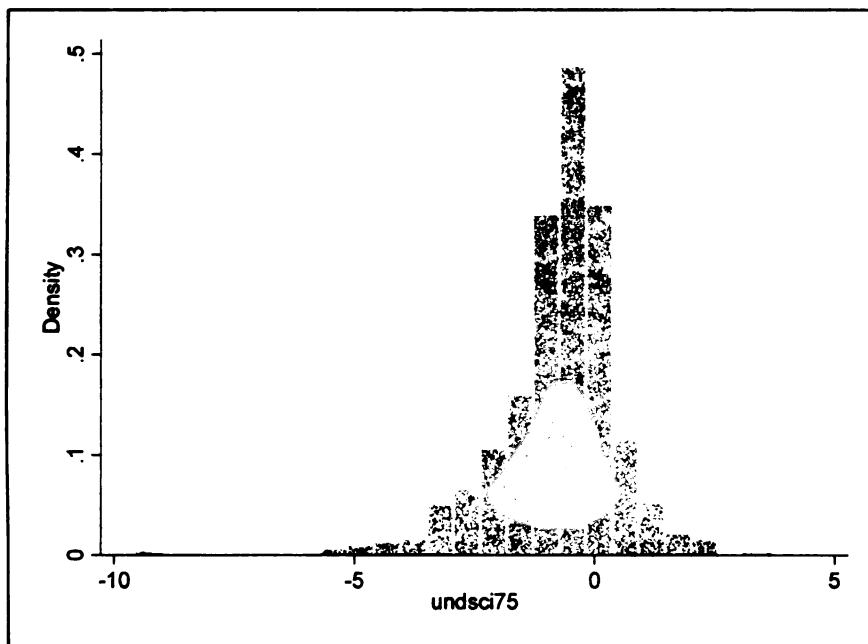


Figure C18: Science undersupply distribution: 0.7 smoothing constant on enrollment;
class size $\sim N(25, .5)$ and courses per FTE $\sim N(5, .5)$

**APPENDIX C: ADDITIONAL INFORMATION ON PROPENSITY SCORE METHODS:
DISCUSSION OF VARIOUS PROPENSITY SCORE METHODS UTILIZED**

TECHNICAL APPENDIX C:
ADDITIONAL INFORMATION ON PROPENSITY SCORE METHODS
DISCUSSION OF VARIOUS PROPENSITY SCORE METHODS UTILIZED

All propensity score matching strategies used in this dissertation involve the estimation of a propensity score as a first step, to identify the propensity to receive the treatment, and then matching based on that propensity score. There are a class of matching estimators that do not require the estimation of a propensity score at the outset (see Guo & Fraser, 2010, Chapter 6).

Estimating the Propensity Score

The propensity score is a balancing score, which estimates the probability of treatment given a set of observed covariates. Rosenbaum and Rubin (1983) were the first to suggest this term, as a solution to the problem of conditioning on all relevant covariates when there is a high dimensional vector of covariates. The propensity is usually estimated using a logit or probit model, although the linear probability model can be used and criticisms of using this model likely do not apply when the model is used for classification, rather than estimation of structural coefficients (Smith, 1997). In these analyses, the propensity score is estimated using a logit model.

Variable choice is critically important in the estimation of the propensity score. Omitting important variables in the matching can increase bias in the resulting estimates (Heckman, Ichimura, & Todd, 1997). It is necessary to specify a model that leads to estimated propensity scores that best represent the true propensity scores, and that balance the groups of observed covariates (Guo & Fraser, 2010). Variables included in the prediction equation must be unaffected by participation, and therefore should be either fixed over time or measured before participation (Caliendo & Kopeinig, 2005).

The literature suggests the use of stepwise logistic regression to select variables (Rosenbaum & Rubin, 1984; Hirano & Imbens, 2001). However, in the context of this study, the range of available variables was relatively limited. Therefore, in order to obtain balance between treatment and control groups on all available observed covariates, any variables that were deemed to be a) causally prior to treatment and b) relevant to the analysis were included in the prediction equation.

Kernel Matching

Kernel-based matching constructs matches using all individuals (or in this analysis, schools) in the potential control sample in such a way that it takes more information from those who are closer matches and downweights more distal observations. In this way, it uses comparatively more information than other matching algorithms (Guo & Fraser, 2010).

Kernel matching uses non-parametric regression. It is a method for estimating average treatment on the treated, using one-to-many matching. Instead of assuming strongly ignorable treatment assignment, kernel matching assumes only the outcome under the control condition is required to be independent of the treatment assignment (Guo & Fraser, 2010; Heckman Ichimura, & Todd, 1997; 1998).. This approach calculates the weighted average of the outcome variable for all nontreated cases and then compares that weighted average with the outcome of the treated case, where the difference between the two terms provides an estimate of the treatment effect for the treated (Guo & Fraser, 2010). A major advantage of this method is that the variance is lower, because more information is used. A drawback is that the observations used may be bad matches (Caliendo & Kopeinig, 2005). Kernel matching can be seen as a

weighted regression of the counterfactual outcome on an intercept with weights given by the kernel weights (Smith & Todd, 2005). These weights are based on the distance between each individual from the control group and the participant observation for which the counterfactual is estimated. As with the selection of the caliper in nearest neighbor matching, the selection of the bandwidth parameter represents an important trade-off between small variance and unbiased estimate (Pagan & Ullah, 1999).

Nearest Neighbor Matching: Nearest neighbor matching is the most straightforward of the matching estimators, where an individual from a comparison group is matched with a treated individual that is closest in terms of the propensity score (Caliendo & Kopenig, 2005). It is part of a class of matching methods known as greedy matching (D'Agostino, 1998; Guo & Fraser, 2010; Smith & Todd, 2005). Nearest neighbor matching can either be conducted with replacement, where each match is used more than once if necessary, or without replacement. There is a trade-off involved between the quality of the matches and the variance of the estimator. When replacement is allowed, the number of distinct non-participants to construct the counterfactual is reduced, and this increases the variance (Smith & Todd, 2005). It is also possible to use more than one neighbor in the matches, as was done in the analysis in Chapter 6. This also has the trade-off of reduced variance but potentially increased bias from poorer matches (Smith, 1997).

Nearest neighbor matching can also impose a caliper, or a restriction on the distance between the propensity scores for each observation. It is suggested that this caliper be a quarter of a standard deviation of the sample estimated propensity scores (Rosenbaum & Rubin, 1985).

The inherent dilemma in nearest neighbor matching, and more broadly, all of the greedy matching algorithms, is the tension between incomplete matching and inaccurate matching. Finding the best neighbor can be challenging; while trying to obtain a maximum number of exact matches, cases can be excluded, but when trying to maximize the number of cases, the matching is often more inexact (Parsons, 2001). It is often necessary to run different caliper sizes, checking the sensitivity of results to different calipers, and choosing the best method based on that (Guo & Fraser, 2010).

Another criticism of this method is that it requires a substantial common support region to function efficiently. Greedy matching can exclude cases on the upper end and lower end of the common support region, because they have no matches. Despite its limitations, greedy matching is that it allows for many types of multivariate analyses to evaluate causal effects, and is a flexible matching strategy (Guo & Fraser, 2010).

Stratification

The stratification method is another commonly used method of propensity score matching (Guo & Fraser, 2010). The main concept of stratification matching is to partition the common support region of the propensity score into a set of strata, and then calculate the impact of the treatment within each interval by taking the mean difference in outcomes between treatment and control conditions (Caliendo & Kopeinig, 2005). This method was developed initially by Rosenbaum & Rubin (1983), but has received much attention and use since then.

One of the main question that arises in stratification matching is the number of strata that are appropriate. Five strata are usually suggested as sufficient to reduce 95% of the bias associated with one single covariate (Cochrane & Chambers, 1965). All bias

under the unconfoundedness assumption is associated with the propensity score itself, so this suggests that under normality, the use of five strata removes most of the bias associated with those covariates (Imbens, 2004).

An important step in the implementation of stratification matching is to check the balance of all of the covariates within each stratum (Aakvik, 2001). This is a built-in function of Stata's `pscore` matching routine.

Weighting on Propensity Score

This is developed more fully in-text, as it was the main matching method used, so is discussed less here. Propensity scores can also be used as weights to obtain a balanced sample of treated and untreated cases (Rosenbaum, 1987; Hirano & Imbens, 2001; Hirano, Imbens, & Ridder, 2003; Imbens, 2004). These weights are used as sampling weights in a regression.

The goal of propensity score weighting is to reweight treated and control participants to make them representative of the population under study. There are different types of weights that can be used, including those to estimate the average treatment effect, the treatment on the treated and the treatment on the control. These were developed in Chapters 2 and 6 and will not be repeated here.

The way in which the propensity score itself is estimated is crucial when implementing weighted estimators (Freedman & Berk, 2008; Zhao, 2004). One optimal method to implement these estimators is to use them with regression adjustment (Hirano & Imbens, 2001), as was done in Chapter 6.

There are limitations of this method. This has been shown to be an optimal method only if study participants were independent and identically distributed, selection

was exogenous, and the selection equation was specified correctly (as discussed above) (Freedman & Berk, 2008). If this is not the case, weighting is likely to increase random error in the estimates, and possibly bias the standard errors downward (Freedman & Berk, 2008). It is sometimes recommended to simply fit a causal model without weights (Guo & Fraser, 2010).

Propensity Scores Used In Multilevel Models

Propensity Analyses in a Multilevel Context

In this analysis, the treatment of interest is a school-level treatment—high rates of school-level teacher retention. Therefore, the propensity score is estimated at the school-level, and then that score is introduced into the actual estimation equations at the school level as well (i.e. as a weight on schools in the weighting equation; as the matching for strata, etc.) There is work that has been done on the estimation of an individual propensity score using a multilevel model (Kim & Seltzer, 2007; Arpino & Mealli, 2008), but little (to none) done on the logistics and implications of doing a school-level propensity with an individual level outcome and how the models need to be modified

The use of propensity scores in a multilevel context is relatively new ground. As Arpino and Mealli state, “In the paper we explore the use of multilevel techniques for the specification of the propensity score for multilevel data. The issue has received little focus in the literature. For the best of our knowledge only Kim and Seltzer (2007) address the issue explicitly. They propose use of a multilevel model for estimation of the propensity score and then implementation of the matching algorithm within each cluster. If we impose the condition that treated and matched controls must belong to the same cluster, we then automatically achieve perfect balancing in all the observed and

unobserved cluster characteristics. This strategy is not likely to be feasible in those situations, representing the norm in social and economic observational studies, where we have relatively few units within each cluster. In these cases, in fact, it is likely that in several clusters it is difficult to find for each treated unit a good matched control belonging to the same cluster.”

This suggests that the strategy is to estimate the propensity score in a multilevel framework, and then estimate the treatment effect for each cluster, which is essentially what I am doing when I use the `atts` command with the `bootstrap (cluster)` command. This is not the best application for my question, when I conceive of the propensity score as being at level 2. In this case, it does not make sense to estimate the treatment effect within each cluster (i.e. school) because the schools themselves are applied to treatment or control. I was able to use the weighting strategy to implement these models, by weighting the level-2 observations (i.e. schools) by their propensity. The results of this are below.

Table A3 below shows the estimated effects of high school-level teacher retention on student mobility, with estimates from the weighted models, as well as from the original unweighted multilevel models. When weighted by propensity to have high teacher retention using the EOTM, students in schools with high teacher retention are .83 times as likely to change or leave the school, although this effect does not meet the threshold for statistical significance. The treatment effect for the treated, similarly, is not statistically significant. However, when looking at the effect of high school-level teacher retention on the control schools (treatment effect on the control), students in schools with high teacher retention are .71 times as likely to change schools as students in schools with

lower retention, and this effect statistically significant (although somewhat marginally so, with t-ratio of -1.820, p-value of .069).

Table A3: Estimated Effect of High School-Level Teacher Retention on Student Mobility

Model	Coefficient	Odds ratio	Std error	t-ratio	pvalue
Weighted by propensity (EOTM)	-0.18	.83	0.18	0.995	.321
Weighted by propensity (treatment effect for the treated)	-.27	.76	.20	1.368	.172
Weighted by propensity (treatment effect for the control)	-0.34	.71	0.19	1.820	.069
Unweighted, with covariates	-0.42	0.65	0.19	2.202	0.028
Unweighted, only prior mobility	-0.06	0.55	0.17	3.543	0.001
Unweighted, bivariate	-1.06	0.35	0.17	6.098	.000

This obviously yields a different finding than when investigating the relationship between school-level teacher retention and student mobility using multilevel modeling with longitudinal data and covariates (the unweighted models). In those models, students in schools with high teacher retention were significantly less likely to leave or change schools, with odd of .65 in the full model, .55 in the abbreviated model, and .35 in the bivariate model.

Given the concerns with the weighting method, my second strategy was to attempt to investigate this using the stratification method. All schools were stratified according to their propensity score, and then separated into strata, and checked for balance. For each school, all students nested within that school are then attached to that

school (and hence, to that strata). Therefore, each strata has a number of schools and students whose stratification membership is defined by the school propensity score.

The table below reports the findings from the stratification method, with bootstrapped standard errors on the ATT.

ATT estimation with the Stratification method Bootstrapped standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
76302	19702	-0.087	0.004	-22.097

Within each strata, the average treatment on the treated (i.e. those students who were in schools with similar propensities to have high teacher retention when the schools actually had high teacher retention) was -0.087, which means that, when matched by the school's propensity to have high teacher retention, students in schools with high teacher retention were significantly less likely to change schools than those in low teacher retention schools, by approximately a 10% difference in probability.. In other words, higher teacher retention "causes" a decrease in student mobility.

However, I became concerned that the multilevel structure of the data was not being appropriately taken into account, and therefore used bootstrap standard errors, with schools identified as the resampling clusters. When using bootstrap standard errors with clustering accounted for, the average treatment on the treated is in the right direction (i.e. those who received the treatment were less likely to change schools), but the t-statistic is not very large, which suggests that this effect has borderline statistical significance.

This highlights one of the challenges of the multilevel framework—how to account for the clustered nature of the data without restricting the estimation to a within-school estimation. The *effect* is solely between schools; all students in each school

experience the same school-level propensity of high teacher retention. The question of interest is how students in schools with similar propensities differ in terms of mobility, not how students within each school differ in terms of their mobility decisions. By the same token, the clustered nature of the data cannot simply be ignored, as the observations clearly have correlated errors. When using the bootstrapping command in Stata, the bootstrapped standard errors are for the ATT—which is the weighted average of block-specific treatment effects, which are generated within each block as the average outcomes of treatment and controls within the same block. When the cluster (schcode) restriction is imposed, the bootstrapping is done within each school, which is inappropriate, as the standard error should be estimated on the treatment effect, which will not be within schools, but between schools.

APPENDIX D: SENSITIVITY ANALYSES

APPENDIX D: SENSITIVITY ANALYSES

As this is an applied study using observational data, I determined it to be critical to employ sensitivity analyses with each of the research methods, in order to help to quantify my results in terms of how sensitive they might be to hidden bias. There are various methods to approach sensitivity analyses. The most familiar framework is to quantify the possible estimates that would result under a variety of possible conditions. The purpose there is to introduce measures of potential bias into the analysis and see what the resulting effect would be, and whether or not that effect will lose its significance. An extension of this is quantifying the robustness of the inference to violations of the assumptions by attempting to identify the exactly how much an assumption must be violated to invalidate an inference (Frank & Sykes et al., 2008; Diprete & Gangl, 2004; Gastwirth, Krieger, & Rosenbaum, 1998; Rosenbaum, 2002). Sensitivity analyses are used in both contexts in this study. This technical appendix provides some additional detail regarding these analyses that was not included in the original chapters.

Sensitivity Analyses to Quantify the Robustness of the Inference

To express robustness that simultaneously accounts for the relationship between a confounding variable and the predictor of interest and between the confounding variable and the outcome in terms relatively accessible to social scientists, I utilize Frank and colleagues (Frank 2000; Pan & Frank, 2004) impact threshold for a confounding variable. As mentioned above, there has been work to extend sensitivity analysis into quantifying the robustness of the inference. The impact threshold for a confounding variable approach was developed in response to that work, to address two limitations (as

outlined by Frank & Sykes et al, 2008. The first is that a confounding variable, by definition, has a relationship with the predictor of interest *and* with the outcome, but the sensitivity analyses named above treat each component separately. These type of approaches that focus only on one component and ignore the dependences between the two components that is created by the presence of a treatment effect. When there is a treatment effect, covariates that are correlated with the predictor of interest are also more likely to be correlated with the outcome, which will inflate the magnitude of the impacts of the covariates on the effects (Frank, 2000) and will skew the distribution of impacts (Pan & Frank, 2003).

Secondly, most sensitivity analyses are expressed in ways that are not intuitive to social scientists, such as concordance or non-parametric statistics, instead of in the general linear model, which is an intuitive part of social scientist's repertoire of skills.

Therefore, the sensitivity analyses employed in this work with the teacher retention models (Chapter 5) and with the undersupply models (Chapter 7), as well as with the weighted propensity score regressions (Chapter 6), utilize this idea of the impact threshold for a confounding variable. Frank (2000) begins by defining the *impact* of a confounding variable on an estimated regression coefficient as $r_{v \bullet y} \times r_{v \bullet t}$, where $r_{v \bullet y}$ is the correlation between a covariate, v , and the outcome, y ; and $r_{v \bullet t}$ is the correlation between v and t , a predictor of interest.¹ The product $r_{v \bullet y} \times r_{v \bullet t}$ captures both the relationship between the confounding variable and the outcome and between the confounding variable and the treatment. Moreover it is through the *impact* that multiple regression adjusts for

¹ This work is taken extensively from the work of Frank (2000), and is drawn from the technical appendix in Frank & Sykes et al, (2008). The derivations are not mine.

covariates as in the following expression for a correlation between t and y , partialling for v :

$$r_{t \cdot y | v} = \frac{r_{t \cdot y} - r_{v \cdot y} \times r_{v \cdot t}}{\sqrt{1 - r_{v \cdot y}^2} \sqrt{1 - r_{v \cdot t}^2}} \quad (1)$$

Equation (4) shows that any reduction in the partial correlation must be attributed to $r_{v \cdot y} \times r_{v \cdot t}$, because the correlations in the denominator will serve only to *increase* $r_{t \cdot y | v}$ relative to $r_{t \cdot y}$.

To obtain the impact necessary to invalidate an inference, $r^\#$ is defined as a quantitative threshold for making inferences from a correlation. For example, $r^\#$ can be defined by a correlation of a specific magnitude (e.g., an effect size). Here $r^\#$ is defined by statistical significance. Although statistical significance is not sufficient for causal inference (Wilkinson & Task Force on Statistical Inference, 1999), statistical significance is often the first threshold in a two-step procedure for making causal inferences, “where first the likelihood of an effect (small p value) is established before discussing how impressive it is” (Wainer & Robinson, 2003, page 25). Most social scientists are uncomfortable making causal inferences if their estimated effect (or something more extreme) could have occurred more than a small percentage (e.g., 5%) of the time by the chance of sampling when in fact the null hypothesis is true.

Given the definition of $r^\#$, it is clear that the inference from $r_{x \cdot y}$ is invalid if $r_{x \cdot y | v} < r^\#$. Therefore to obtain the impact necessary to invalidate the inference, set $r_{x \cdot y | v} < r^\#$ in (4). Next, to give maximal credence to the challenge that the inference is invalid

because of exclusion of a confounding variable, maximize (4) with respect to *impact*.

Frank (2000) shows that this maximum occurs when $r_{v \cdot y} = r_{v \cdot t}$ (given the constraint:

$impact = r_{v \cdot y} \times r_{v \cdot t}$). Given this maximum, *impact* can be substituted for $r_{v \cdot y} \times r_{v \cdot t}$, $r_{v \cdot y}^2$

and $r_{v \cdot t}^2$ in (4) and solving for *impact* yields that $r_{x \cdot y | v} < r^{\#}$ if

$$impact > (r - r^{\#}) / (1 - |r^{\#}|). \quad (2)$$

Therefore the quantity $(r - r^{\#}) / (1 - |r^{\#}|)$ defines the impact threshold for a confounding variable; if there is a confounding variable with *impact* greater than $(r - r^{\#}) / (1 - |r^{\#}|)$ then the relationship between the treatment and outcome, given the confound ($r_{x \cdot y | v}$), would fall below the threshold ($r^{\#}$) for making a causal inference.²

Critically, because the impact is defined by correlation coefficients ($impact = r_{v \cdot y} \times r_{v \cdot t}$), it can be readily understood by social scientists comfortable with correlation and the general linear model. This makes it an ideal complement to our use of propensity weighting which is applied to a general linear model.

Multilevel Sensitivity Analyses

The models employed in Chapter 5 and Chapter 7 are multilevel models, with students nested within schools. This is a common situation with education data. In order to extend the sensitivity analyses described above into a multilevel framework, I follow Seltzer, Kim, & Frank (2006) and their extension of the ITCV into a multilevel context.

² The expressions can be easily adapted to focus on one component correlation when researchers have specific prior beliefs about the strength of the other correlation. The expressions can also be modified to account for the presence of other covariates in the model. See Frank (2000).

Frank's approach described above was originally developed for use in settings with linear models with non-nested data (Seltzer, Kim, & Frank, 2006).

Seltzer, Kim, & Frank suggest that it is often beneficial to recast multilevel models into cluster-level regression analyses and using OLS estimates of level-1 intercepts or slopes as outcomes. They find that when data are balanced, the OLS regression with cluster-level predictors will produce estimates of fixed effects and their standard errors that are identical to those produced by a REML/EB estimation approach that takes into account the clustered nature of the data. If the data are not balanced, they suggest using a Weighted Least Squares cluster-level regression, using weights based on the REML estimates of the variance components (Seltzer, Kim & Frank, 2006). These weights are constructed as:

$$Weight = \frac{1}{\tau + V_j} \text{ where } V_j = \frac{\sigma^2}{n_j} \quad (3)$$

To apply this framework to the sensitivity analyses in Chapter 5, the following steps were taken:

- 1) An unconditional multilevel model was estimated, with the outcome variable as student mobility. The purpose of this model is to produce estimates of τ and σ^2 . However, since the outcome variable was dichotomous, there is no variance component (σ^2) produced in the unconditional model. The variance of a proportion is calculated by $np(1-p)$, where p is the proportion of students in each school who changed schools.
- 2) These estimates were used to generate precision weights, with the following formula:

$$\text{Precision} = \text{Weight} = \frac{1}{\tau + V_j} \text{ where } V_j = \frac{\sigma^2}{n_j} \text{ and where } \sigma^2 = np(1-p)$$

- 3) These weights were used as sampling weights in the cluster-level regressions. All student-level predictors were aggregated to the cluster-level (i.e. the school level). To obtain the R^2 estimates necessary for calculating the impact threshold for a confounding variable, the treatment is regressed on the predictors of interest, generating an R^2 value. Then the outcome is regressed on the same predictors, without the treatment. These R^2 values are used to calculate the impact.

In chapter 7, a similar method was utilized. The unconditional model estimated in step 1 above did produce both an estimate for τ and σ^2 . These were used to calculate precision weights. This was done for each outcome—mathematics achievement, English language arts achievement, and science achievement, and these weights were used in the cluster-level regressions as described above.

Sensitivity Analyses with Propensity Score Matching

Before discussing the strategy used to evaluate the sensitivity of the inference under the stratification and nearest neighbor matching estimates, I will discuss the propensity score analysis with propensity scores used as weights.

The theory behind sensitivity analysis in this situation is the same as above. Given a set of predictors and an outcome of interest, with each school weighted by its propensity to receive the treatment, how large would an unobserved confound need to be in order to invalidate the inference? Therefore, precision weights were constructed as follows:

- 1) The weights obtained from the multilevel regression predicting student mobility obtained in Chapter 5 were utilized again.
- 2) As the propensity weights are also weights, the product of the precision weight and the propensity weight was calculated.
- 3) The full models were ran, weighted by propensity only and also weighted by the product of propensity and precision. No noticeable differences were detected.
- 4) The impact thresholds were calculated for each propensity weight—the EOTM, the TOT, and the TOC—in the usual format as described above.

To conduct sensitivity analyses on the stratification and nearest neighbor matching, a different conceptual framework was applied. The question in sensitivity analyses in propensity score matching relates to hidden bias—the presence of an unobserved variable that would affect both the likelihood to receive treatment and the likelihood to have increased outcomes. The goal of these sensitivity analyses is to determine how strongly an unmeasured covariate must influence *the selection process* in order to undermine the implications of the matching analysis (Rosenbaum, 2002; Caliendo & Kopeinig, 2005; Aakvik, 2001; DiPrete & Gangl, 2004; Caliendo, Hujer, & Thomsen, 2005).

The approach is as follows³:

Assume that the probability of participation in treatment to be:

$$P(x) = P(D_i = 1 \mid x_i) = F(\beta x_i + \gamma_0) \quad (5)$$

Where

x_i = observed characteristics for individual i

³ Please note: I rely heavily on Rosenbaum, 2002, and a discussion of Rosenbaum found in Caliendo & Kopeinig, 2005, to represent these ideas. I will not cite repeatedly throughout this section. In particular, I want to note that the derivations are not my own but are simply reproduced here for reference and to appropriately acknowledge the intellectual underpinnings of the method I applied.

u_i = unobserved variable

γ = effect of u_i on the participation decision.

If the study is free of hidden bias, γ will be zero, which indicates the participation probability will be solely determined by the observed characteristics (x_i). If there is hidden bias, two individuals (or in this case, schools) with the same observed covariates (x) will have *differing* chances of receiving treatment. Assume that there is a matched pair of schools i and j and also assume that F is the logistics distribution. The odds that

schools receive treatment are then

$$\frac{\frac{P(x_i)}{1 - P(x_i)}}{\frac{P(x_j)}{1 - P(x_j)}} \text{ and } \frac{P(x_j)}{(1 - P(x_j))} \quad (6)$$

This yields an odds ratio that is given by:

$$\frac{P(x_i)(1 - P(x_j))}{P(x_j)(1 - P(x_i))} = \frac{\exp(\beta x_j + \gamma u_j)}{\exp(\beta x_i + \gamma u_i)} = \exp[\gamma(u_i - u_j)] \quad (7)$$

Both schools differ in their odds of receiving treatment by a factor that includes the parameter γ and the difference in the unobserved covariates (u). If $u_i = u_j$ (there are no differences on the unobserved covariates) or if $\gamma = 0$ (the unobserved variables have no impact on the probability of participating), the odds ratio will be one, which implies that there is no hidden or unobserved selection bias.

The goal of the sensitivity analyses is to analyze how influence about the treatment effect might be altered by different values of γ and $(u_i - u_j)$.

To conduct these analyses, it is necessary to assume that the unobserved covariate is a dummy variable. This implies the following bounds on the odds-ratio that either of the two matched schools will receive treatment (Rosenbaum, 2002):

$$\frac{1}{e^\gamma} \leq \frac{P(x_i)(1 - P(x_j))}{P(x_j)(1 - P(x_i))} \leq e^\gamma \quad (8)$$

Both matched schools have the same probability of receiving treatment only if $e^\gamma=1$. If $e^\gamma=2$, then schools that appear to be the same in terms of their observed covariates could still differ in their odds of receiving treatment by as much as a factor of 2. This makes e^γ a measure of the degree of departure from a study free of hidden bias (Rosenbaum, 2002).

To test this, the non-parametric Mantel-Haenszel (MH, 1959) test-statistic is used. This test statistic compares the successful number of persons in the treatment group against the same expected number given the treatment effect is zero. In other words, it is used to test for no treatment effect within different strata. The test statistic is:

$Q_{MH} = (Y_{1s} - E(Y_{1s})) / \text{Var}(Y_{1s})$ where Y_{1s} = the number of successful participants, Y_{0s} is the number of successful non-participants, and Y_s is the number of total successes in stratum s . This test statistic follows a chi-square distribution with one degree of freedom and is given by:

$$Q_{mh} = \frac{U^2}{\text{Var}(U)} = \frac{[\sum_{s=1}^S (Y_{1s} - \frac{N_{1s}Y_s}{N_s})]^2}{\sum_{s=1}^S \frac{N_{1s}N_{0s}Y_s(N_s - Y_s)}{N_s^2(N_s - 1)}} \quad (9)$$

In order to utilize this test statistic, the treatment and control group need to be as equal as possible, which is accomplished by the matching procedure. Therefore, I can discuss the possible influences of $e^\gamma > 1$. The test statistic Q_{mh} can be bounded by two known distributions (Rosenbaum, 2002). If $e^\gamma = 1$, the bounds are equal to the baseline scenario in which there is no hidden bias. As e^γ increases, the bounds move apart, and therefore reflect uncertainty about the test-statistics if there is unobserved selection bias. To quantify this, there are two variables: Q_{MH}^+ assumes an overestimation of the treatment effect and Q_{MH}^- assumes an underestimation of the treatment effect. The two bounds are given by:

$$Q_{MH}^{+(-)} = \frac{[\sum_{s=1}^S Y_{1s} - \tilde{E}_s^{+(-)}]^2}{\sum_{s=1}^S Var(\tilde{E}_s^{+(-)})} \quad (10)$$

Where \tilde{E}_s and $Var(\tilde{E}_s)$ are the large sample approximations to the expectation and variance of the number of successful participants when u is binary and for given γ (Caliendo & Kopeinig, 2005).

BIBLIOGRAPHY

Bibliography

- Aakvik, A. (2001). Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics*, 63(1), 115-143.
- Achieve, Inc. (2008). *Closing the expectations gap 2008: An annual 5-state report on the alignment of high school policies with the demands of college and careers*. Retrieved from <http://www.achieve.org/files/50-state-2008-final02-25-08.pdf>
- Achieve, Inc. (2009). *Will raising high school graduation requirements cause more students to drop out?* Retrieved on June 2, 2009 from: <http://www.achieve.org/node/599>
- Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.
- American Association of State Colleges and Universities. (2006). *High school coursework: Policy trends and implications for higher education*. Retrieved from <http://www.aascu.org/media/pm/pdf/v3n7.pdf>
- Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous predictors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics*, 19, 2-16.
- Arnold, C. L, Choy, S. P, & Bobbitt, S. A. (1993). *Modeling teacher supply and demand, with commentary* (NCES 93-461). Washington, DC: National Center for Education Statistics.
- Arpino, B., & Mealli, F. (2008). *The specification of the propensity score in multilevel observational studies* (Working Paper No. 6). Milan: Carlo F. Dondena Centre for Research on Social Dynamics.
- Balfanz, R., & West, T. C. (2009). *Raising graduation rates: A series of data briefs*. The Everyone Graduates Center. Retrieved from <http://www.every1graduates.org/PDFs/StateProgressReport.pdf>
- Barnes, G., Crowe, E., & Schaefer, B. (2007). *The cost of teacher turnover in five districts: A pilot study*. New York: National Commission on Teaching and America's Future.
- Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal*, 2, 358-377.

- Becker, S. O., & Caliendo, M. (2007). *Mhbounds: Sensitivity analysis for average treatment effects* (IZA Discussion Paper No. 2542). Bonn, Germany: Institute for the Study of Labor.
- Bidwell, C. (1965). The school as a formal organization. In J. March (Ed.), *Handbook of Organizations* (pp. 973-1002). Chicago: Rand McNally.
- Boe, E. E., & Gilford, D. M. (1992). *Teacher supply, demand, and quality: Policy issues, models, and data bases*. Washington DC: National Academies Press.
- Borman, G., & Dowling, N. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, 78, 367-409.
- Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2006). *Analyzing the determinants of the matching of public school teachers to jobs*. Retrieved from http://www.teacherpolicyresearch.org/portals/1/pdfs/Matching_of_Public_School_Teachers_to_Jobs.pdf
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.
- Bryk, A., Lee, V., & Smith, J. (1990). High school organization and its effects on teachers and students: An interpretive summary of the research. In W. H. Clune and J. F. Witte (Eds.), *Choice and control in American education, Vol. 1: The theory of choice and control in education* (pp. 135-226). New York: Falmer Press.
- Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.
- Burns, T., & Stalker, G. (1961). *The management of innovation*. London: Tavistock.
- Caliendo, M., & Kopeinig, S. (2005). *Some practical guidance for the implementation of propensity score matching* (IZA Discussion Paper No. 1588). Bonn, Germany: Institute for the Study of Labor.
- Caliendo, M., Hujer, R., & Thomsen, S. (2005). *The employment effects of job creation schemes in Germany: A microeconomic evaluation* (IZA Discussion Paper No. 1512). Bonn, Germany: Institute for the Study of Labor.
- California Department of Finance (2006). *California Public K-12 enrollment and high school graduate projections by county, 2006 series*. Sacramento: California Department of Finance, Demographic Research Unit.
- Campbell, H. S. (1997). School enrolment in an extended demographic model. *Journal of Economic and Social Measurement*, 23, 27-44.

- Carroll, S., Reichardt, R., & Guarino, C. (2000). *The distribution of teachers among California's schools and districts*. Santa Monica, CA: RAND Corporation.
- The Cherry Commission. (2004). *Final report of the Lt. Governor's commission on higher education & economic growth*. Retrieved from <http://www.cherrycommission.org/docs/finalReport/CherryReportFULL.pdf>
- Cochrane, W., & Chambers, S. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society, Series A*, 128, 234-266.
- Coleman, J., & Hoffer, L. (1987). *Public and private schools: The impact of communities*. New York: Basic Books.
- Cook, T., Shadish, W., & Wong, V. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27, 724-750.
- Croninger, R. G., & Lee, V. E. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *Teachers College Record*, 103, 548-581.
- Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high schools. *American Sociological Review*, 74, 709-730.
- D'Agostino, R. B., Jr. (1998). Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17, 2265-2281.
- Darling-Hammond, L. (1984). *Beyond the commission report: The coming crisis in teaching*. Santa Monica, CA: RAND Corporation.
- Darling-Hammond, L. (2000). *Solving the dilemmas of teacher supply, demand, and standards: How we can ensure a competent, caring, and qualified teacher for every child*. New York: National Commission on Teaching and America's Future.
- Dee, T. S., & Jacob, B. A. (2007). The impact of high school exit exams on educational attainment and labor market outcomes. In A. Gamoran (Ed.), *Standards-based reform and the poverty gap: Lessons for "No Child Left Behind"* (pp. 154-199). Washington, DC: Brookings Institution Press.
- DiPrete, T., & Gangl, M. (2004). *Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments*. Retrieved from http://www.wjh.harvard.edu/~cwinship/cfa_papers/HBprop_021204.pdf

- Dolton, P., & van der Klaauw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics*, 81, 543-552.
- Dounay, J. (2006). *Alignment of high school graduation requirements and state-set college admissions requirements*. Denver, CO: Education Commission of the States. Retrieved on June 2, 2009 from: <http://www.ecs.org/clearinghouse/68/60/6860.pdf>. Update June 12, 2009.
- Durkheim, E. (1961). *Moral education: A study in the theory and application of the sociology of education* (E.K. Wilson & H. Schnurer, Trans.) New York: Free Press.
- Ehrenberg, R., & Smith, R. (1997). *Modern labor economics: Theory and public policy* (6th ed.). Reading, MA: Addison-Wesley.
- Figlio, D. (2010, January 20). *Competitive effects of means-tested school vouchers*. Presentation at the Economics of Education Project Visiting Scholars Colloquium at Michigan State University.
- Finn, J. D. & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27, 557-577.
- Frank, K. (2000). Impact of a confounding variable on the inference of a regression coefficient. *Sociological Methods and Research*, 29, 147-194.
- Frank, K. A., & Min, K-S. (2007). Indices of robustness for sample representation. *Sociological Methodology*, 37, 349-392.
- Frank, K., Muller, C., Schiller, K., Riegle-Crumb, C., Mueller, A., Crosnoe, R., et al. (2008). The social dynamics of mathematics coursetaking in high school. *American Journal of Sociology*, 113, 1645-1696.
- Frank, K., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A., et al. (2008). Does NBPTS Certification Affect the Number of Colleagues a Teacher Helps with Instructional Matters? *Education, Evaluation and Policy Analysis*, 30, 3-30.
- Frank, K., Duong, M., Maroulis, S., Kelcey, B. (2010). *Quantifying discourse about causal inferences from randomized experiments and observational studies in educational research*. Manuscript submitted for publication.
- Freedman, D. A., & Berk, R. A. (2008). Weighting regressions by propensity scores. *Evaluation Review*, 32, 392-409.
- Futernick, K. (2007). *A possible dream: Retaining California's teachers so all students learn*. Sacramento: California State University.

- Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis* 22, 241-254.
- Gastwirth, J. L., Krieger, A. M., & Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, 85, 907-920.
- Goodman, J. (2010). Skills, schools and credit constraints: Evidence from Massachusetts. *Education Finance and Policy*, 5, 36-53.
- Grant, G. (1988). *The world we created at Hamilton High*. Cambridge, MA: Harvard University Press.
- Grissmer, D., & Kirby, S. (1997). Teacher turnover and teacher quality. *Teachers College Record*, 99, 45-56.
- Grodsky, E., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971-2004. *Educational Policy*, 23, 589-614.
- Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76, 173-208.
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Los Angeles: Sage Publications.
- Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19, 141-164.
- Hanushek, E. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and other investigations of class size effects. *Educational Evaluation and Policy Analysis*, 21, 143-163.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39, 326-354.
- Harris, D. N & Adams, S. (2007). Understanding the level and causes of teacher turnover: A comparison with other professions. *Economics of Education Review*, 26, 325-337.
- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature*, 33, 1829-1878.

- Haviland, A., Nagin, D. S., & Rosenbaum, P. R. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods, 12*, 247-267.
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64*, 605-654.
- Heckman, J., Ichimura, H., & Todd, P. E. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies, 65*, 261-294.
- Heckman, J. (2005). The scientific model of causality. *Sociological Methodology, 35*, 1-99.
- Hedges, L. V., Laine, R., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Education Researcher, 23*, 5-14.
- Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. *Health Services & Outcomes Research Methodology, 2*, 259-278.
- Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*, 1161-1189.
- Hom, P., & Griffeth, R. (1995). *Employee turnover*. Cincinnati: South-Western Publishing.
- Hussar, W., & Bailey, T. (2008). *Projection of education statistics to 2017* (NCES 2008-078). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Imazeki, J. (2005). Teacher salaries and teacher attrition. *Economics of Education Review, 24*, 431-449.
- Imbens, G. W. (2004). Nonparametric estimation of the average treatment effects under exogeneity: A review. *The Review of Economics and Statistics, 86*, 4-29.
- Ingersoll, R. (1995). *Teacher supply, teacher qualifications, and teacher turnover*. Washington, DC: National Center for Education Statistics.
- Ingersoll, R. (2000). *Turnover among mathematics and science teachers in the U.S.* Washington, DC: National Commission on Mathematics and Science Teaching for the 21st Century.

- Ingersoll, R. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38, 499-534.
- Ingersoll, R. (2003). *Is there really a teacher shortage?* Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Ingersoll, R., & Perda, D. (2009). *The mathematics and science teacher shortage: Fact and myth* (CPRE Research Report #RR62). The Center for Policy Research in Education (CPRE). Retrieved from http://www.cpre.org/images/stories/cpre_pdfs/math%20science%20shortage%20paper%20march%202009%20final.pdf
- Institute for Education Sciences. (n.d.) *Statewide Longitudinal Data Systems Grant Program: Overview*. Retrieved from <http://nces.ed.gov/programs/slds/>
- Jacob, B. (2001). Getting tough? The impact of mandatory high school graduation exams on student achievement and dropout rates. *Educational Evaluation and Policy Analysis*, 23, 99-122.
- Johnson, M. K., Crosnoe, R., & Elder, G. H. (2001). Student attachment and academic engagement: The role of ethnicity. *Sociology of Education*, 74, 318-340.
- Kalleberg, A., & Mastekaasa, A. (1998). Organizational size, layoffs, and quits in Norway. *Social Forces*, 76, 1243-1273.
- Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22, 523-539.
- Kanter, R. (1977). *Men and women of the corporation*. New York: Basic.
- Keesler, V. A., Wyse, A., & Jones, N. (2008a). *Beyond compliance: Calculating within school teacher (in FTEs) supply and demand to meet the Michigan Merit Curriculum* (Technical Report No. 2). Lansing: State of Michigan.
- Keesler, V. A., Wyse, A. & Jones, N. (2008b). *Calculating the ability of within-school teacher supply to meet the demands of new requirements: The example of the Michigan Merit Curriculum* (REL Technical Brief). Retrieved from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/techbrief/tr_00508.pdf
- Kennedy, P. (2003). *A guide to econometrics* (5th ed). Cambridge: MIT Press.
- Kim, J. & Seltzer, M. (2007). *Causal inference in multilevel settings in which selection processes vary across schools* (CSE Technical Report 708). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Center for the Study of Evaluation (CSE).

- Kruger, A. B. (1999). Experimental estimates of educational production functions. *Quarterly Journal of Economics*, 114, 497-532.
- Kruger, A. B. (2000). *Understanding the magnitude and effect of class size on student achievement*. In L. Mishel & R. Roghstein (eds.), *The class size policy debate* (pp.7-34). Washington, DC: Economic Policy Institute,
- Lankford, M., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis*, 26, 217-236.
- Lee, V. (Ed.). (2002). *Reforming Chicago's high schools: Research perspectives on school and system level change*. Chicago: Consortium on Chicago School Research.
- Lee, V., Bryk, A., & Smith, J. (1993). The organization of effective secondary schools. *Review of Research in Education*, 19, 171-267.
- Lee, V., & Smith, J. (1997). High school size: which works best and for whom? *Educational Evaluation and Policy Analysis*, 19, 205-227.
- Levine, P. B., & Zimmerman, D. J. (1995). The benefit of additional high school math and science classes for young women: Evidence from longitudinal data. *Journal of Business and Economic Statistics*, 13, 137-149.
- Likert, R. (1967). *The human organization*. New York: McGraw-Hill.
- Lillard, D. R., & DeCicca, P. P. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20, 459-473.
- Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, 80(3), 44-70.
- Lortie, D. (1975). *School teacher*. Chicago: University of Chicago Press.
- Lucas, S. R., & Good, A. D. (2001). Race, class, and tournament track mobility. *Sociology of Education*, 74, 139-156.
- Lynn, F. B., & Keesler, V. A. (2008, October). *Technical report 3.1: Teacher licensure and endorsement in the State of Michigan*. Lansing: State of Michigan.
- Lynn, F. B., Keesler, V. A., Diemer, M., Guarino, C., Jones, N., Wang, Q., et al. (2007). *Beyond compliance: Descriptive characteristics of public school teachers in Michigan* (Technical Report No. 1). Lansing: State of Michigan.

- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Marvel, J., Lyter, D., Peltola, P., Strizek, G., Morton, B., & Rowland, R. (2007). Teacher attrition and mobility: results from the 2004-2005 Teacher follow up survey (NCES 2007-307). Washington, DC: National Center for Education Statistics, U.S. Department of Education..
- Michigan Department of Education. (2006). *Improving outcomes for high school students*. Retrieved from <http://www.michigan.gov/mde/0,1607,7-140-38924---,00.html>
- Michigan Consortium for Educational Research. (2010). *The impact of the Michigan Merit Curriculum and Michigan Promise Scholarship on student outcomes* [Grant application submitted to IES Evaluation of State and Local Programs].
- Michigan's Race to the Top Application. (2010). Retrieved from http://www.michigan.gov/mde/0,1607,7-140-37818_53083-236662--,00.html
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, 81, 247-258.
- Milanowski, A., & Odden, A. (2007). *A new approach to the cost of teacher turnover* (Working Paper 13). Seattle, WA: Center on Reinventing Public Education.
- Morgan, S., & Harding, D., (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research*, 35, 3-60.
- Mueller, C., & Price, J. (1990). Economic, psychological, and sociological determinants of voluntary turnover. *Journal of Behavioral Economics*, 19, 321-335.
- Muller, C. (2001). The role of caring in the teacher-student relationship for at-risk students. *Sociological Inquiry*, 71, 241-255.
- Murphy, P., DeArmond, M., & Guin, K. (2003). A national crisis or localized problems? Getting perspective on the scope and scale of teacher shortage. *Education Policy Analysis Archives*, 11(23). Retrieved from <http://epaa.asu.edu/epaa/v11n23/>
- National Commission on Teaching and America's Future. (2002, August). *Unraveling the teacher shortage problem: Teacher retention is the key*. A Symposium of The National Commission on Teaching and America's Future and NCTAF State Partners, Washington, DC.

- Newman, B. M., Lohman, B. J., Newman, P. R., Myers, M. C., & Smith, V. L. (2000). Experiences of urban youth navigating the transition to ninth grade. *Youth and Society*, 31, 387-416.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5, 14-37.
- Pagan, A., & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge, UK: Cambridge University Press.
- Pan, W., & Frank, K. A. (2004). A probability index of the robustness of a causal inference. *Journal of Educational and Behavioral Statistics*, 28, 315-337.
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review*, 29, 297-318.
- Parsons, L. S. (2001). *Reducing bias in a propensity score matched-pair sample using greedy matching techniques* (SAS SUGI paper 214-26). Proceedings of the 26th annual SAS Users' Group International Conference, SAS Institute, Cary, NC.
- Porter, L., Lawler, E., & Hackman, J. (1975). *Behavior in organizations*. New York: McGraw-Hill.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Reichardt, R. (2006). *In search of quality: Teacher retention*. Hartford: Connecticut Center for School Change.
- Reichardt, R. (2008). Teacher attrition and teacher diversity in Colorado. In J. Paone, J. Whitcomb, T. Rose, & R. Reichardt (Eds.), *Shining the light II: State of teacher quality, attrition, and diversity in Colorado* (pp. 36-49). Denver, CO: The Alliance for Quality Teaching.
- Rice, J. A. (2007). *Mathematical statistics and data analysis* (3rd ed.). Belmont, CA: Thomson Higher Education.
- Riegle-Crumb, C. (2006). The path through math: Course-taking trajectories and student performance at the intersection of gender and race/ethnicity. *American Journal of Education*, 113, 101-122.
- Robins, J. M., Hernan, M., & Brumback, B. (2000). Marginal structural models and causal inferences in epidemiology. *Epidemiology*, 11, 550-560.

- Robins, J. M., & Ronitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90, 122-129.
- Robins, J. M., Rotnitzky, A., & Scharfstein, D. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In E. Halloran & D. Berry, D. (Eds.), *Statistical models in epidemiology, the environment, and clinical trials* (pp. 1-92) New York: Springer.
- Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (1998). Academic and emotional functioning in early adolescence: longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology*, 10, 321-352.
- Rose, H. & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, 86, 497-513.
- Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82, 387-394.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.) New York: Springer.
- Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1809-1814). New York: Wiley.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using sub-classification on the propensity score. *Journal of the American Statistical Association*, 79, 516-524.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, 39, 33-38.
- Rosenfeld, L. B., Richman, J. M., & Bowen, G. L. (2000). Social support networks and school outcomes: the centrality of the teacher. *Child and Adolescent Social Work Journal*, 17, 205-226.
- Rosenholtz, S. (1989). *Teacher's workplace: The social organization of schools*. New York: Longman.
- Rumberger, R., & Larson, K. (1998). Student mobility and the increased risk of high school drop out. *American Journal of Education*, 107, 1-35.

- Rumberger, R., & Thomas, S. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73, 39-67.
- Sadler, P. M., & Tai, R. H. (2007). The two high-school pillars supporting college science. *Science*, 317(5837), 457-458.
- Santiago, P. (2002). *Teacher demand and supply: Improving teaching quality and addressing teacher shortages*. (OECD Education Working Papers No. 1). Paris: OECD.
- Schneider, B., Swanson, C. B., & Riegle-Crumb, C. (1998). Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education* 2, 25-53.
- Sells, L. W. (1973). High school mathematics as the critical filter in the job market. In R. T. Thomas (Ed.), *Developing opportunities for minorities in graduate education* (pp. 37-39). Berkeley: University of California Press.
- Seltzer, M., Kim, J., & Frank, K. (2006). *Studying the sensitivity of inferences to possible unmeasured confounding variables in multisite evaluations* (CSE Technical Report 701). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Center for the Study of Evaluation (CSE).
- Shen, J. (1997). Teacher retention and attrition in public schools: Evidence from SASS91. *Journal of Educational Research*, 91, 81-88.
- Simpkins, S., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42, 70-83.
- Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*, 27, 325-353.
- Smith, J. (2000). A critical survey of empirical methods for evaluating active labor market policies. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 136(3), 1-22.
- Smith, T., & Ingersoll, R. (2004). Reducing teacher turnover: What are the components of effective induction? *American Educational Research Journal*, 41, 687-714.
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125, 325-353.

- Stanton-Salazar, R. D., & Dornbusch, S. M. (1995). Social capital and the reproduction of inequality: Information networks among Mexican-origin high school students. *Sociology of Education*, 68, 116-135.
- Stevenson, D. L., Schiller, K. S., & Schneider, B. (1994). Sequences of opportunities for learning. *Sociology of Education*, 67, 184-198.
- Stinebrickner, T. (1998). An empirical investigation of teacher attrition. *Economics of Education Review*, 17, 127-136.
- Stinebrickner, T. R. (2001). A dynamic model of teacher labor supply. *Journal of Labor Economics*, 19, 196-230.
- Stinebrickner, T. (2002). An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers leave. *Journal of Human Resources*, 37, 192-216.
- Stockard, J., & Lehman, M. (2004). Influences on the satisfaction and retention of 1st-year teachers: The importance of effective school management. *Educational Administration Quarterly*, 40, 742-771.
- Swanson, C., & Schneider, B. (1999). Students on the move: residential and educational mobility in America's schools. *Sociology of Education*, 72, 54-67.
- Sweeney, S. H., & Middleton, E. J. (2005). Multiregional cohort enrolment projections: Matching methods to enrolment policies. *Population, Space, and Place*, 11, 361-379.
- Teachman, J., Paasch, K., & Carver, K. (1996). Social capital and dropping out of school early. *Journal of Marriage and Family*, 58, 773-783.
- Temple, J., & Reynolds, A. (1997). Student mobility and achievement: Longitudinal findings from an urban cohort. *Journal of School Psychology*, 37, 355-377.
- Turner, A. N., & Lawrence, P. R. (1964). *Industrial jobs and the worker*. Cambridge, MA: Harvard University Press.
- Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher*, 32, 22-30.
- Waller, W. (1932). *The sociology of teaching*. New York: Wiley.
- Walton, R. (1980). Establishing and maintaining high commitment work systems. In J. Kimberly & R. Miles (Eds.), *The organization life cycle* (pp. 208-290). San Francisco: Jossey-Bass.

- Warren, J. R., Jenkins, K., & Kulick, R. (2006). High school exit examinations and state level completion and GED rates, 1975-2002. *Educational Evaluation and Policy Analysis*, 28, 131-152.
- Watkins, T. D., Jr. (2003). *The Michigan definition for identifying highly qualified teachers*. Lansing: Michigan Department of Education.
- Whitener, S., Gruber, K., Lynch, H., Tingos, K., Perona, M., & Fondelier, S. (1997). *Characteristics of stayers, movers, and leavers: Results from the Teacher Follow-Up Survey: 1994-1995* (NCES 97-450). Washington, DC: National Center for Education Statistics.
- Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *The Review of Economics and Statistics*, 86, 91-107.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03063 6561