QUALITY TEACHERS WANTED: AN EXAMINATION OF STANDARDS-BASED
EVALUATION SYSTEMS AND SCHOOL STAFFING PRACTICES IN NORTH
CAROLINA MIDDLE SCHOOLS

Shanyce L. Campbell

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Public Policy

Chapel Hill
2014

Approved by:

Gary T. Henry

Douglas L. Lauen

Karolyn D. Tyson

Lora Cohen-Vogel

Steven W. Hemelt

ABSTRACT

SHANYCE L. CAMPBELL: Quality Teachers Wanted: An Examination of Standards-Based Evaluation Systems and School Staffing Practices in North Carolina Middle Schools
(Under the direction of Dr. Gary T. Henry)

A quality education has substantial returns such as higher wages, improved health, reduced incarceration rates and increased civic engagement. While there are long-term benefits to obtaining an education, research continues to highlight disparities in the educational opportunities students receive. Recent, policy efforts to improve education for all students has focused on teachers, because they are the most influential school-level factor to student success.

This dissertation examines the ways in which school leaders recruit, allocate, and evaluate beginning teachers across North Carolina's middle schools in efforts to improve student success. Understanding these dynamics can help policy makers create and implement policies geared towards improving the quality of teachers in classrooms. In the first chapter, I use administrative data to examine whether principals' evaluation ratings of teachers validly reflect teachers' effectiveness, as measured by value-added scores. This chapter also examines whether principals fairly evaluate teachers based on teachers' race, ethnicity, and gender. In the second chapter, I assess how principals evaluate teachers based classroom and school characteristics. Analyses suggest that there are marginal differences in evaluation ratings based on teachers' ascriptive characteristics – race, ethnicity, and gender. However, classroom and school characteristics has little to no relationship on teachers' evaluation ratings.

In the third chapter, I employ a qualitative case study design to understand how beginning teachers are staffed to schools and assigned to classrooms. Interviews with core

(English language arts, math, science, social studies) middle grade teachers and principals from a large district, show little evidence that beginning teachers hold specific preferences when applying for and accepting teaching positions. Conversely, school leaders do not intentionally seek out beginning teachers in staffing their schools. As it relates to how teachers are assigned to the classroom, school leaders use of a "spread the wealth" philosophy and strategy among ELA and math teachers to ensure all teachers teach a heterogeneous group of students.

This dissertation addresses an important but often overlooked area in education policy – equitable evaluation and assessment of beginning teachers. Much of the research on improving student achievement has focused on how best to measure teacher effectiveness – credentials or value-added estimates. As school leaders continue to improve the quality of education for students, the results presented here suggest several possible ways achieve this goal. In these analyses, classroom and school characteristics are not associated with evaluations ratings; however, there is an association between evaluation ratings and the race/ethnicity and gender of teachers, after controlling for teachers' performance. This underscores the need for North Carolina to develop a rigorous statewide training on the proper use of the evaluation rubric. Furthermore, the findings indicate that school leaders use an equity-based approach to ensure that no teacher is assigned to all high performing or low performing students. This strategy assumes that all teachers have the knowledge, skills, and dispositions to teach all students; however, if this hypothesis is inaccurate school leaders may want to reconsider the weight placed on equity-based teacher assignment practices.

This dissertation is dedicated to my late grandmother, A. Nadine Williamson, who spent over 25 years in service as an educator and all educators who tirelessly work to educate our children not simply school them.

ACKNOWLEDGMENTS

I am grateful to formally express my deep appreciation and gratitude to those who have helped me complete this journey. First and foremost, I would like to thank my co-author, the one who created my innermost being, who knit me together in my mother's womb, and fearfully and wonderfully made me in your image – my Abba. This journey could not and would not have been completed without your grace, mercy, joy, peace, assurance, and protection. Your works are wonderful and I know this very well (Psalms 139:13). In those dark places, where I felt like giving up you constantly reminded me that you started this journey and would complete it (Philippians 1:6). You were there many of days when this journey seemed unbearable, but you guided me through it your mighty right hand (Psalms 73:23, 139:10). Through the process of becoming Dr. Campbell, you have taught me to persevere, fight, listen, trust, and completely surrender my life's work to glorify you. Although, I am not all the way perfected, everyday I am learning to seek you for my daily bread, not leaning on my own limited understanding, and acknowledging that you are where my help comes from.

Thank you Gary T. Henry, my Chair, for your guidance and support even before becoming my chair. You provided several opportunities for me to grow as an independent scholar and challenged me to think outside of my box and explore new ideas. On a personal level, thank you for allowing me a safe space to share my struggles and my frustrations.

I thank each of my committee members, Doug Lauen, Karolyn Tyson, Steve Hemelt, and Lora Cohen-Vogel, for their contribution to the rigor of your research by sharing their expertise and by offering me encouragement and compassion. A special thank you to Karolyn, who served

not only as a committee member, but as a mentor. Thank you for pushing me intellectually, but doing so with such humility. Your candid advice and support definitely helped me persist on this journey.

Many thanks to all of the staff, faculty, and doctoral students in the Department of Public Policy for your instruction, support, and laughs through the doctoral program. Gail Corrado, what would I have done without you???!! You are an amazing mentor, advocate, and friend, who constantly assured me that 'I know enough, I'm good enough, and I am enough.' You have taught me to think outside of box of normalcy in order to instigate real change. You have given me an arsenal of strategies and ideas to use in advocating for marginalized populations. Christine Durrance, thank you for being a wonderful econ professor and allowing me the space to be completely vulnerable, lost, and confused. More importantly, thank you for allowing the space to vent and share my life issues.

Much thanks to my *original* cohort members – Kevin Bastian, Shelley Golden, Jade Marcus Jenkins, Jaehee Park, and Roderick Rose. Since our first day at math camp (excluding Rod because he was too smart for that nonsense ☺), we became not only a cohort, but a family. As a family we have had some great experiences such as seeing kids start walking and talking, marriages, and new jobs. Together, we have also experienced some not so good events such as deaths of loved ones and animals, and job/research frustrations. Thank you all of always helping me think through concepts, teaching me new STATA and SAS commands, and simply keeping me sane and encouraged to along this journey. Special shout out to my cohort sisters – Shelley and Jade. Shells, what an awesome sister I have in you. You always made sure I was fed and brought me into the entire "GoBo" family. While I'm on the GoBo's, Tom, thank you for being a great brother and fellow Aggie, who was not only concerned about my dissertation journey but

Valarie Ashby you taught me that my voice is valuable in a space that has tried many times to marginalize and silence my voice. Thank you for your candid words of encouragement early in this journey.

I am grateful for each of my therapist who helped get me through this process, with minimal mental health damage. A special thank you to Wanda Tyndall White for getting me to a place to mentally to take care of myself while sacrificing myself in service to others. Thank you for being a part of my healing with your candid, honest, and solution-oriented advice. It all helped carry me throughout this journey.

Thank you to Winnie Ayers, Captola Harris, Linda Holinshed, and Norma Monroe for being my "other mothers" who genuinely loved me, corrected my grammar, provided me access to scholarships for college.

Thank you to staff and faculty at the greatest higher ed institution in the world – North Carolina A& T State University – for teaching me that 'you lead by serving and you serve by leading.' Thank you Quiester Craig, my Dean, for your belief in me as a student, leader, and scholar. Thank you to Gwendolyn "Dr. HQ" Highsmith-Quick for being my "other mother" and mentor. Scott Simkins, thank you encouraging me to consider the Ph.D and economics…I didn't become an economist, but I got real close…

The terms family and friends are synonymous concepts to me so I will thank my capital F, Family. Thank you to the Williamson, Hudson, Berry, and Rhodes-Sankofa Waters families. To my mother, Ina Rhodine Hudson, thank you for life. Who would I be if we didn't share in life's trials and tribulations together? Who would I be if I couldn't be your "Precious Jewel?" You have taught me how to fight, how to preserve, how to overcome obstacles, how to love, and

how to forgive. All of which I used to get through this journey, so I honor you with this dissertation, because it is just as much yours as it is mine. I love you!

To my dad, Cartrell Hudson, thank you for always being a listening ear, even when you had no idea what I was talking about. I don't think I would have made it through this journey without your daily calls just "to hear my voice" and laugh hysterically about the simple things. Thank you for being an awesome dad, who wouldn't allow me to quit when I wanted to and who I could call wanting nothing at all. I also honor you with this dissertation. I love you!

To the Brooks family, thank you for allowing me to just be your sister and an auntie. You all gave me a way to escape the daily grind of disseratating to enjoy skype-versations (yea I just made up a word). Jackie Brooks, my sister, your support has always been endless, whether it was flying to Chicago to be by my side or just allowing me to vent. I love you my big, little sister!

Thank you to my aunt, Yvette Williamson, who has always been my advocate, my support, and my friend. Whether you realize it or not, your life has been such an example and encouragement me along this journey. You are such a wonderful example of a beautiful and resilient woman. Vette, you have taught me to endure, to love BIG, to hope for better, and to put action to my words. You have proved the adage that says, "a setback is just a setup for a comeback." In those days when I needed encouragement to push through setbacks, I was reminded of you. I am proud of all that you have achieved in the health field and honor you with this dissertation. Your Beacer loves you!

Tanesha Berry, my sister, thank you living out this journey with me. You were there before the idea of getting a Ph.D. and have been here through the entire process. You were there to celebrate and cry with me! In those dark times, when I was frustrated and wanted to give up on this journey, you always encouraged me to keep going. When I accomplished a milestone

along this process, you also encouraged me to keep going. Every card and gift that you sent to simply say 'I'm thinking of you' always assured me that no matter what I knew I had you in my corner. Thank you for doing life with me for over twenty years. You are such treasure and I honor you with this dissertation.

Billye and Malari Sankofa Waters, my sister and brother, thank you both for being my life vests throughout this process. Malari, you challenged me to think more deeply as an intellectual, but also allowed me to not think at all and just 'Be.' I'm am grateful to have such a smart, inquisitive, and hilarious brother. Billye…can I just stop there?! From the day you accosted me with a big, invasive hug, I knew you were different, strange, full of love, and permanent! SMH…LOL! I love you for being all those things and more…your love, prayers, sisterhood, etc. etc. are endless. I am forever indebted to you for being such a solid rock of support throughout this journey. Thank you for doing life with me and I honor you with this dissertation. Thank you also to Mary Rhodes, my other mother, who's constant prayers were necessary and heard.

Jessie Montana Cain, my sister and twinsie, I never knew that I could share so many of the same experiences with another person. I am grateful to you for helping me think through ideas, for being my working partner, and for being my escape from academia. How would I have made it through this journey without our 'sophistaratchet' escapades, late night random trips to Target or Wal-Mart just to cut up and act a fool, or SB-a-thongs? Thank you for allowing me into your Family. I honor you with this dissertation.

Thank you to Darnell and Danielle Cloud for being there with me before this process ever began. Darnell, thank you encouraging me years ago to consider the Ph.D. You gave me my first experience as a researcher and I wouldn't have become Dr. Campbell, without your help editing

my personal statements, offering advice, and assuring me that I would finish this process. You are such a wonderful mentor, friend, and brother and I honor you with this dissertation.

A special acknowledgement goes to my Summit church family. J.D. Greear, my lead pastor, thank you for being such a humble man of God. Thank you for always being a solid teacher – for teaching me so much about the Gospel, about who Jesus is, about living missionally, and pushing me to go All-In with my relationship with God. It didn't fall on deaf ears or rocky soil! A special thanks to all the members of my small group and the First Impressions 'tent ladies.' Thank you ladies for fervently praying for me through this journey, for helping me grow spiritually, and for keeping me accountable as I fully and completely decided to go all-in in my relationship with God. You ladies cried when I cried, celebrated with me, and encouraged me to remember that this dissertation is just a part of God's journey for my life. For all that I have become is owed to the prayers of my church family who righteously sought God on my behalf.

It was at the Summit Church that I met two special life-long sisters, Yolanda Reed (and family) and Yolanda Finney. Ladies, you made me know even more that God is real and that God loves me, because God sent y'all to me when I needed it the most – the last leg of this dissertation process. Thank you for challenging me to grow spiritually, for being a listening ear, for calling me in those midnight hours to pray with me. Yolanda R., thank you for your endless love! You gave me an example of what it looks like to trust God in living out your purpose. You opened your doors and allowed me to become an aunt to Keijuane and Zion, who allowed me to be a big kid. But also a sister to Chuck, my voice of reason and spiritual mentor. You showed me what a Godly marriage looks like in its most authentic form. YoYo Fin, thank you for your endless love as well. You fussed me out many times to push passed my "I'm over this" mentality

and finish this dissertation. Because of you, I understand how to trust God in the unknown and surrender my plans, ideas, and ways to Him.

To my Sleepover Sisters (S.O.S.) – Billye Sankofa Waters, Corliss Brown Thompson, Cassandra Davis, Shannon Lewis, Jessie Cain and Danielle Parker. As you all know the doctoral process can be isolating, stressful (even more so for scholars of color), and just down right difficult. Thank y'all for accepting me as your sister and pushing, encouraging, laughing, eating, shopping, conferencing, and working with me through this journey. I COULD NOT have gotten through this process without each of you, individually and collectively. I honor you all with this dissertation; it's really OUR dissertation.

Joseph Atatsi, thank you for doing life with me as I studied for the GRE, for reviewing my personal statement, for praying with me in times of celebration and frustration, and for offering your endless love and support throughout this journey. You have taught me how to rely on God, how to trust myself, and push past obstacles. Before I left Chicago, you gave me a card and wrote an acrostic poem, which has hung on my bedroom mirror for the past 6 years to always remind me that The Lord will be my **S**hield, **H**elper, **A**nchor, **N**ew Wine, **Y**outh, **C**omforter, and **E**difier. You were absolutely right! Thank you for being such a wonderful partner, friend, and brother. I honor you with this dissertation.

Dwayne Deloatch, thank you for walking alongside me through this journey. You were always there the listen to me and offer your advice through the difficult times. Thank you for always checking in on me to make sure I remained healthy – physically, mentally, and spiritually. Thank you for being such a wonderful partner, friend, and brother. I honor you with this dissertation.

Thank you to all of my study participants, known and unknown. This work could not have been done without actual individuals. As researchers, especially quantitative researchers, we often forget that there are people behind the numbers. Teachers, principals, students, and parents you are not forgotten and are constantly in my prayers. Be encouraged that true education and true education reform is possible. I honor you with this dissertation.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter I

STANDARDS-BASED TEACHER EVALUATION SYSTEMS: RATINGS, PERFORMANCE
MEASURES AND THE VARIATION BY RACE, ETHNICITY, AND GENDER

In response to federal pressure, states have developed standards-based evaluation systems that
combine principal ratings, value-added scores and in some cases, other measures of teachers'
performance, such as student surveys. School leaders and district administrators use these
evaluation systems for low and high stakes decisions (Murphy, Hallinger, & Heck, 2013). As
educational leaders and policymakers continue to use standards-based evaluations of teachers as
a strategy to reform public education, understanding the validity and fairness of standards-based
evaluation systems provide important information about the integrity of the evaluation system
and its potential to improve teacher performance.

In this manuscript, we examine the extent to which principals' ratings appear to validly
reflect teachers' performance as measured by value-added scores and whether principals fairly
evaluate teachers in North Carolina's public middle schools. The data come from North Carolina
public school (NCPS) administrative data and includes middle school teachers in their first five
years of teaching during the 2010-2011 school year.

Our results suggest that principals give higher ratings to teachers who make larger
student gains. We also find that after controlling for teachers' concurrent value-added scores,
principals rate White male English language arts (ELA) teachers systematically lower than White

female ELA teachers. In addition, principals rate Black female math teachers systematically lower than White female math teachers, after controlling for concurrent value-added scores.

Introduction

In recent years, federal and state education reforms have placed increased importance on improving teacher quality. Although teachers exert the largest influence on student learning of all school-based factors, there is significant variation in the effectiveness of teachers both between and within school settings (Rockoff, 2004; Hanushek et al., 2005; Henry et al. 2014a; Henry et al. 2014b). However, until recently the evaluation of teachers' performance appears to have been more-or-less perfunctory. In response to federal pressure, primarily from the Race to the Top funding competition, states have developed standards-based evaluation systems that combine principal ratings, value-added scores and in some cases, other measures of teachers' performance, such as student surveys.

According to a September 2013 report by the National Center on Teacher Quality, only 11 states and the District of Columbia require statewide implementation of an evaluation system. The remaining states either have no statewide specifications (2 states); states provide a model where districts can opt out (10 states) or districts have the flexibility to design their own system (27 states) (Doherty & Jacobs, 2013). Additionally, 44 states and the District of Columbia require classroom observations; however, only twenty-seven require the use of student test growth data a measure in the evaluation system (Doherty & Jacobs, 2013). States also use the evaluation system for varying purposes including professional development, dismissal, tenure, and compensation.

Traditional teacher evaluations, which were usually based on informal classroom observations conducted by principals (hereafter 'principal evaluations of teachers'), are criticized

for several reasons including principals' tendency to inflate ratings, conflicting motivations and objectives for the evaluations, and desire to maintain positive relationships with teachers (Bol, 2011; Levy & Williams, 2004). However, current education reforms sought to counter the perceived arbitrariness of the principal evaluations of teachers by instituting standards-based evaluation systems.

Standards-based teacher evaluation systems have incorporated two new ideas: (1) the use of rubric-based standards on which principals evaluate teachers according to their performance in key areas related to student learning and (2) the inclusion of value-added measures in the evaluation of teachers. However, little is known about the extent to which principal ratings using standards-based evaluation systems are closely related to objective measures of teachers' performance or if the ratings might compromise the validity and fairness of teachers' evaluations. For these evaluation systems to have the desired impact of improving and reducing the variability of the performance of teachers, the evaluations must reflect teachers' performance and be fair – showing no consistent or uniform, i.e. systematic bias, differences toward any racial, ethnic or gendered groups.

In this manuscript, we examine the extent to which principals' ratings appear to validly reflect teachers' performance, as measured by value-added scores, and whether principals fairly evaluate teachers in North Carolina's public middle schools. We define fairness as the lack of systematic associations between teachers' race, ethnicity or gender and the principals' ratings of teacher performance after controlling for an objective measure of teachers' performance. As educational leaders and policymakers continue to use standards-based evaluations of teachers as a strategy to reform public education, understanding the validity and fairness of standards-based

evaluation systems provide important information about the integrity of the evaluation system and its potential to affect teacher performance positively.

More generally, our findings inform the teacher quality literature by providing preliminary evidence that standards-based evaluation systems, at least as currently implemented in North Carolina, may not remove subjective influences on principals' ratings of teachers. This manuscript also contributes to the empirical literature on teacher evaluation systems by incorporating objective measures of teachers' effectiveness in the examination of systematic differences in principals' evaluation ratings. Additionally, the methods used include within school-fixed effects to account for observable and unobservable time-invariant characteristics of schools and school leaders, thus strengthening the validity of the systematic patterns presented in the findings.

In the next section, we review the literature on teacher evaluations, followed by an overview of North Carolina's teacher evaluation system. We then describe the methodology including the data and analytic plan. Finally, the results and discussion are presented.

Education Literature on Evaluation Systems

Similar to many industries, evaluations in the field of education serve two main functions: improving performance of teachers and personnel decision-making by school leaders (Arvey & Murphy, 1998; Castilla, 2008; Scriven, 1967). As a performance tool, evaluations may provide valuable sources of information for educational administrators and teachers to improve instructional quality. The use of the evaluation as a performance tool is especially important for beginning teachers who often require additional guidance regarding instructional practices, classroom management, and other aspects of teaching than experienced teachers (Peterson, 2004). Recent research using rigorous designs strongly suggests that using standards-based evaluation for developmental purposes can increase the achievement gains of students taught by

4

teachers who participated in the evaluation process, even veteran teachers (Taylor & Tyler, 2012). However, for higher stakes purposes such as performance incentives, the research to-date finds no link between incentives and teachers' behaviors and performance (Springer et al., 2012; Yuan et al., 2012).

Educational administrators also use evaluations in their decision-making efforts to continue employment, confer tenure, or determine performance pay awards.

To examine the role of evaluation systems within the field of education, we synthesize three areas of literature: principals' evaluations of teachers; standards-based evaluation systems; and the relationship between evaluations and teacher characteristics.

Principal Evaluations and Teacher Effectiveness

The evaluation of teachers is a long held practice and arguably one of the most important responsibilities of principals (Marzano, Frontier, & Livingston, 2011). However, the validity and reliability of principal evaluations of teachers remains a source of controversy. Principal evaluations of teachers are those in which principals enter the classroom to informally observe teachers and often without specific observation criteria. Many researchers and practitioners call into question principals' knowledge and skills necessary to appropriately evaluate teachers. Common concerns center on principals' lack of instructional leadership such as their limited content knowledge and understanding of the classroom environment to evaluate teachers' performance (Soar, Medley, & Coker, 1983; Stodolsky, 1984; Epstein, 1985; Darling-Hammond, Wise, & Pease, 1983).

To date, the majority of studies on principal evaluations focus on the relationship between teacher evaluation ratings and student achievement. Early studies examining the effect of principal evaluation ratings of teachers and teachers' effectiveness, measured by student achievement gains, found no significant relationship between these two measures (see Medley &

5

Coker, 1987 for a list of eleven studies). According to Medley and Coker (1987), these studies suffered from threats to internal validity. However, after accounting for these threats, Medley and Coker found consistent results and concluded, "research provides no support whatever for the widely held belief that the average principal is a good judge of teacher performance" (p.245).

In these prior studies, student achievement gains were calculated using norm-referenced tests, which are designed to produce a relative ranking of students compared to their peers. This test does not attempt to measure gains in student learning in terms of the objectives that have been set for them or a teacher's ability to provide quality instruction. A criterion-referenced test (CRT) measures how well students learned the content expected in the standards. This test provides a more accurate measure of student learning, because it directly relates to teachers' instructional practices, which principals evaluate during classroom observations. Using CRTs, Manatt and Daniels (1990) replicated Medley and Coker's (1987) study and found that principals are able to accurately evaluate teacher effectiveness. The authors suggest that use of CRTs, extensive and effective principal training on the evaluation instrument, and more advanced methods explained the contrasting results from prior studies.

Researchers have also measured principal evaluation ratings using other instruments other than classroom observations. Examining 360° feedback surveys from students, teachers, and principals, Wilkerson and colleagues (2000) found that principals' ratings of teachers were positively correlated with student achievement in math and ELA; however, there was no statistical correlation in reading. A more recent study, using a principal evaluation survey of teachers, found that principals were able to distinguish between high- and low-performing teachers, but had difficulty identifying average performing teachers (Jacob & Lefgren, 2008).

Taken together, there is some evidence that principals are able to identify effective teachers, especially at the extremes of performance. The research evidence also suggests that principal training is a source of variation in the relationship between evaluation ratings and student achievement. Moreover, this may indicate the need for a systematic evaluation process, which is the basis of standards-based evaluation systems.

Standards-Based Evaluations and Teacher Effectiveness

During in the 1990s, accountability pressures began to shift school leaders' interest from principal evaluations of teachers to standards-based evaluations of teachers. Standards-based evaluations of teachers are based on a comprehensive framework or rubric that defines the expectations for teachers' performance and are conducted by school leaders, usually principals or assistant principals and in some cases independent observers. As a system, the evaluation requires a thorough collection of evidence that includes observations and artifacts (i.e. lesson plans, student work, etc.). Much of the research on standards-based teacher evaluation employed Danielson's Framework for Teaching. The Framework for Teaching is a set of standards that measure instructional practices related to improving student learning (Milanowski, 2004). Similar to research on principal evaluations of teachers, the underlying working hypothesis for research on standards-based evaluation systems is that the ratings for teachers should have a positive and significant relationship with teachers' ability to raise students' test scores. These studies, implemented in several districts across the U.S., were conducted primarily by a group of researchers from the Consortium for Policy Research in Education (CPRE) (Milanowski, Kimball, & White, 2004; White, 2004; Milanowski, 2004; Kimball, White, Milanowski, & Borman, 2004). The findings from CPREs collection of studies provide mixed evidence for the predictive validity of standards-based teacher evaluation systems.

In two studies of the Cincinnati Public School district, teachers' evaluation ratings were positively correlated with their students' achievement gains (Holtzapple, 2003; Milanowski, 2004). The magnitude of the correlations between composite evaluation ratings and student achievement gains ranged from 0.26 in science to 0.43 in math. While Milanowski's study examined evaluation ratings using the entire four-domain Framework for Teaching rubric, Holtzapple (2003) examined a single domain arguably most related to student achievement – Teaching for Learning. Holtzapple (2003) also found that students taught by the lowest rated teachers (i.e. unsatisfactory or basic rating) performed lower across two years on reading, science, and social studies test than predicted. Only students taught by teachers who received distinguished ratings made positive gains across all subjects and years. However, a study on Washoe County School's evaluation system in Nevada found no statistically significant results between evaluation ratings and students' math achievement or across some grade levels (Kimball, White, Milanowski, & Borman, 2004).

Studies on standards-based evaluation suggest a large variation in the relationship between evaluation ratings and student achievement gains across grade levels and subjects. This variation may be explained by the interaction between evaluators' motivation, skill, and context (Kimball & Milanowski, 2009) or systemic issues such as misalignment between standards-based evaluation rubrics and state standards by subject (Gallagher, 2004).

It is not clear, however, whether these findings are generalized to various geographic locations, secondary grades, or using rigorous specifications. For instance, CPRE's studies examined single districts or schools within two regions of the U.S. – the Midwest and West and focused primarily on elementary school grades. The reliance on single districts and elementary grades present external validity concerns about the consistency of the findings across grades,

regionally, and statewide. The studies also used different types of tests to examine student gains. For example, Kimball et al. (2004) used the Terra Nova, a norm referenced test, as a pretest for fifth grade students and the state criterion-referenced test as the posttest. Finally, the studies employed two-level hierarchical linear models to account for the nesting of students within classrooms, but did not account for the nesting of classrooms within schools.

Evaluations and Teacher Characteristics

Much of the research on evaluations, both principal and standards-based, have focused on two primary points of inquiry: the effect of teacher evaluation ratings on student achievement and the identification of effective teachers. This research has important implications for the broader domain of teacher quality; however, the impact of these evaluations on teachers is missing from the conversation. While it is important that teachers improve student outcomes, it is equally important that teachers be evaluated fairly, conditional on their performance. For these evaluation systems to be considered fair, the ratings for teachers should be systematically related to teachers' performance, not ascriptive characteristics such as gender, ethnicity and race.

To our knowledge, there is only one study that directly examines the relationship between evaluation ratings and teacher characteristics. Jacob and Walsh examined the relationship between evaluation ratings and teacher characteristics associated with performance such as educational credentials, experience and absences as well as race/ethnicity-gender characteristics (Jacob & Walsh, 2011). They found that compared to first year teachers, those with 10 years of experience were rated 0.55 standard deviations higher on evaluations. Teachers from selective colleges and those who majored in education also received higher ratings than their peers. In terms of ascriptive characteristics, White female teachers received higher ratings than all other racial/ethnic-gendered groups.

While Jacob and Walsh's study appears to be the only examination of teacher characteristics, other studies have examined the relationships between ethnic and gender congruence (e.g., principals and teachers of the same racial or ethnic group) on working conditions and evaluation ratings. The findings suggest that principals provide racially congruent teachers with more intangible benefits (Grissom and Keiser, 2011). In addition, gender congruence is also shown to influence working conditions and teacher retention. Specifically, male teachers supervised by a female principal were less satisfied with their jobs and more likely to leave their current school than female teachers (Grissom, Nicholson-Crotty, & Keiser, 2012).

While these studies were not directly related to how teachers are evaluated, they offer empirical support for the hypothesis that congruence or incongruence in terms of racial/ethnic and gender identification of the principal and teacher may affect evaluations of teachers.

North Carolina's Teacher Evaluation System

In this manuscript, we focus on the standards-based evaluation system being implemented statewide in North Carolina. This teacher evaluation system grew out of education reform efforts of the late 1970s to improve the quality of education (Stacey, Holdzkom, & Kuligowski, 1989; Ellett & Garland, 1987). In 1978, the General Assembly of North Carolina developed a statewide evidence-based evaluation system, known as the Teacher Performance Assessment Instrument (TPAI). All teachers were required to participate in the evaluation process; however, requirements varied based on teachers' license status (i.e. career teachers vs. probationary). The evaluation system was later revised (name changed to TPAI-R) in response to the implementation of the School-Based Management and Accountability Program of 1995 and the Excellent Schools Act of 1997.

Recognizing current challenges to ensure all students have the knowledge and skills to succeed in the 21st century; North Carolina's Department of Public Instruction, in partnership

with the Mid-continent Research for Education and Learning, developed a standards-based evaluation system – North Carolina Educator Evaluation System (NCEES) to replace the TPAI-R. The NCEES simultaneously serves dual purposes: professional development and personnel decision-making (e.g., promotion) of teachers. The NCEES was implemented in three phases. The first phase occurred during the 2008-2009 school year and included 13 voluntary districts across North Carolina. During the 2009-2010 school year, an additional 39 districts voluntarily participated in phase two of the evaluation process. By the 2010-2011 school year, the final online evaluation system was launched and included the remaining 63 districts.

During the NCEES process teachers receive four classroom observations throughout the school year, three by a school leader and one by a peer evaluator. In order to serve as a peer evaluator, teachers must complete the training on the North Carolina Teacher Evaluation Process. Until teachers are granted career status (commonly referred to as "tenure"), they are required to participate in the evaluation process. Career status teachers fully participate in NCEES (i.e. receive a summative evaluation) during their license renewal year.

The NCEES rating system is based on the North Carolina Professional Teaching Standards created by the North Carolina Professional Teaching Standards Commission. The rubric includes five standards and twenty-five elements that describe the knowledge, skills and dispositions of an effective teacher. The five standards are: (Standard 1) Demonstrate leadership; (Standard 2) Establish a respectful environment for a diverse population of students; (Standard 3) Know the content they teach; (Standard 4) Facilitate learning for their students; and (Standard 5) Reflect on their practice (SBE, n.d.). Within each standard, teachers are evaluated on a set of elements, which are subdivided into descriptors. Each descriptor contains language that describes the performance responsibilities at each level of the rubric: Distinguished, Accomplished,

Proficient, Developing, and Not Demonstrated. Evaluators assign ratings for individual

descriptors within an element. The overall rating for each element is based on the lowest rating

received across all descriptors. As an example, Figure 1 contains an element and descriptors for

Standard 1.

As illustrated in Figure 1, this fictional teacher would receive a rating of "Developing" on

the "Teachers lead in the classroom" element, despite receiving higher ratings on other

descriptors within this element, because "Developing" is the lowest rating in which all

descriptors are observed/marked.

The evaluation process includes seven components: training, orientation, teacher self-

assessment, pre-observation conference, classroom observations, post-observation conference,

and a summary evaluation conference. At the end of the school year, the principal conducts

summary evaluations to determine teachers' formal summative rating on each standard. The

rating is not a simple average of observation scores. Principals use multiple evidences including

classroom observations and artifacts (lesson plans, student work, service on committees, etc.) to

determine the final ratings for each standard.

Data and Sample

The data for this manuscript come from North Carolina Department of Public Instruction

(NCDPI) administrative data on students, teachers, classrooms, and school characteristics using

personnel, school report card and classroom roster files. We supplement this data with evaluation

rating data from NCDPI and use unique individual teacher identifiers to merge the two datasets.

Complete evaluation ratings are available for 45,900 NCPS teachers across all grade levels and

subjects for the 2010-2011 school year. However, our sample only includes beginning ELA,

math, Algebra I and science teachers in North Carolina public middle schools during the 2010-

2011 school year. Our final sample consists of 2,451 unique middle school teachers across the four subjects identified.

We focus on beginning teachers for two primary reasons. First, there is a preponderance of evidence that effectiveness increases for teachers in their five years of teaching (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Clotfelter, Ladd, & Vigdor, 2007, 2010; Harris & Sass, 2011; Henry, Bastian, & Fortner, 2011; Henry, Fortner, & Bastian, 2012). Therefore, it is suspected that more variation in evaluation ratings will exist among beginning teachers. Second, unlike career status teachers, all beginning teachers are required to undergo the full NCEES process; as such makeup a larger portion of those evaluated.

The analysis is limited to middle school teachers for conceptual and statistical reasons. Middle school is often an under-researched, yet crucial point in the academic future of students. Teachers in this grade level must ensure students are prepared for rigorous high school material. Failing to do this increases the risk that students will drop out of school (Bridgeland, Dilulio, & Morison, 2006; Rumberger, 2001). Examining the integrity of the evaluation system for middle schools teachers is critical in ensuring that principals' assessment of teachers' performance is substantiated.

Examining middle schools also allows for a more nuanced examination of evaluation ratings across multiple subject areas that are not possible in self-contained elementary classrooms. [1] The isolation by subject matter helps inform larger policy issues and practices. For example, if there are differences across subjects on Content Knowledge – Standard 3, teacher education programs may want to examine the quality of instruction within the specialized

---

[1] This assumes that ELA, math, Algebra I and science teachers are mutually exclusive, which is not the case in all schools across North Carolina. For example, if a teacher teaches math and Algebra, of the three required observations the principal evaluate the teacher solely on Algebra I performance, vice versa or a combination of the two courses.

content areas. Finally, beginning teachers in middle grades teach multiple classes and are accountable for the test performance of more students per year than elementary grade teachers. The increased number of classes and students taught provides more statistical power in the measurement models discussed later.

Dependent Variables and Covariates

As previously discussed, during the end of the school year principals provide teachers with a summary evaluation rating on each of the five standards. The final ratings are converted into a 5-point scale, where a rating of 1 indicates "Not Demonstrated" and a rating of 5 indicates "Distinguished". For this manuscript, we create a unit-weighted composite rating from the five summative ratings and use this as a dependent variable along with each of the five individual ratings.

Covariates

The focal variable of interest for the analysis of validity is the teacher effectiveness measure estimated by teachers' value-added scores. For the analysis of fairness, the focal variables of interest are teachers' race, ethnicity and gender. To isolate the effect of these socio-demographic characteristics on evaluation ratings, we also include teachers' effectiveness, experience, preparation, and principal characteristics covariates.

Teacher Effectiveness. In this study, teacher effectiveness is measured by individual teacher scores using value-added models (TVA). TVA is a statistical technique that estimates the contribution of a teacher in a given subject, grade, and year by isolating the effects of the teacher from other factors, such as family, student, and school characteristics. The defining characteristic of TVA models is that students' prior year test scores are included in the model along with the current scores to determine the "value" or amount of learning attributed to a particular in a given year.

Debate surrounding the use of value-added models is a major topic of discussion as policymakers and educational leaders seek to make high-stakes decisions using these estimates. On one hand, scholars argue that value-added measures are unreliable over time, limited to tested subjects and specific grades, do not account for teacher or student sorting, and do not effectively indicate teacher quality or effectiveness (Harris, 2009; Hill, 2009). On the other hand, scholars argue that value-added estimates explain variation in student test score gains, predict future teacher performance, and provide meaningful information at a low cost (Nye, Konstantopoulous, & Hedges, 2004; Gordon, Kane, & Staiger, 2006; Harris, 2009). Despite this debate, most agree that value-added models provide the best available objective measure of teacher effectiveness at this time.

Socio-demographics. Race, ethnicity and gender are used as key teacher characteristics of interest in this manuscript. We use three racial and ethnic categories of teachers: Black, White, and nonBlack teachers of color. NonBlack teachers of color are combined due to sample size limitations and include Hispanic Americans, Asian Americans, Native Americans, Multiracial/multiethnic Americans, and unspecified racial and ethnic groups. Traditional approaches in addressing racial stratification include race/ethnicity and gender as individual variables in analytic models. However, this approach ignores the intersectionality of race or ethnicity and gender (Crenshaw, 1989, 1991; Giddings, 1984). In a study on the intersectionality of race, ethnicity and gender in the labor market, Browne and Misra (2003) state, "race is 'gendered' and gender is 'racialized,' so that race and gender fuse to create unique experiences and opportunities for all groups" (p. 488). To provide a more meaningful examination of teacher evaluation systems and the policy implications, we incorporate this theoretical framework and include six race/ethnicity-by-gender groups of teachers: Black females, Black males, White

females, White males, nonBlack females of color, and nonBlack males of color. All groups were dummy coded 1 or 0. White females, the modal group, are the reference category for all analyses. We provide results with the three race/ethnicity groups; however, the six race/ethnicity-gender groups is our preferred specification.

Experience and Preparation. Teachers' experience is based on the total years of actual teaching and does not include credit from related work experience gained from non-education industries. In recent years, the type of preparation teachers enter the classroom with has become increasingly diverse (Henry et al., 2014a). As a result, beginning teachers may exhibit differences in their content knowledge, skills, dispositions and other important aspects of teaching. These differences may have important implications for teachers' evaluation ratings. For example, in-state public undergraduate prepared teachers might receive high ratings due to their pre-service course work and student teaching directly related to the North Carolina Professional Teaching Standards than out-of-state undergraduate prepared teachers.

Five preparation portals, which prior research has shown to have different effects on student achievement gains (Henry et al. 2014a; 2014b), are included as controls for this analysis: in-state public undergraduate prepared (reference group), Teach for America (TFA), out-of-state undergraduate prepared, lateral entry, and all other portals. All other portals include in-state and out of state graduate prepared, in-state private undergraduate and graduate programs, visiting international faculty, licensure only, and unclassifiable. These preparation portals are grouped together because they individually represent a small proportion of beginning teachers and the differences in their value-added estimates of effectiveness were relatively small. While teachers who entered the classroom via TFA preparation also represents a small proportion of the beginning teacher workforce, they are examined separately because studies show that TFA

teachers are more effective in secondary grades and STEM subjects (Henry et al., 2014a; Xu, Hannaway, & Taylor, 2011).

Principal Characteristics

    Socio-demographics. Similar to the teacher socio-demographic controls, we include six race/ethnicity-by-gender groupings of principals: Black females, Black males, White females, White males, nonBlack females of color, and nonBlack males of color. All groups were dummy coded 1 or 0. White females, the modal group, are the reference category. To examine the influence of race/ethnicity and gender congruence on evaluation ratings, we create three dichotomous congruence variables: racial congruence, gender congruence, and race/ethnicity – gender congruence.

    Experience. A continuous variable of years of experience as a principal is included as a control.

Analytic Approach

    The goal of this study is twofold: to examine the validity and fairness of NCEES. Because the goals are distinct, we employed two separate analytic approaches. To examine the validity, we examine the relationship between teacher evaluation ratings on the NCEES with a more objective measure of teacher effectiveness – TVA. We do not expect these measures to perfectly correlate with another because they do in fact measure different aspects of teacher or instructional quality. For instance, the evaluation ratings measure knowledge, skills and dispositions that are not well represented in TVA estimates such as teachers assuming leadership roles in the school, profession and community; establishing a respectful classroom environment; and reflecting on teaching practices. However, we do expect a positive relationship between the two measures since both should assess instructional quality. More concretely, we would expect that teachers with high TVA estimates also receive high evaluation ratings, especially on

standards most directly related to student learning such as knowing the content and facilitating student learning.

To examine this relationship, a two-step approach that includes a measurement and an analysis model is employed. In the measurement model, we derive individual teacher value-added estimates for a given subject using a three-level hierarchical linear model (Henry, Bastian, & Fortner, 2011 and Henry, Fortner, & Bastian, 2012). The hierarchical linear model accounts for the nesting of students within teachers' classrooms that are nested within schools. The TVA estimates, include a robust set of covariates such as students' prior test scores and other student, classroom, and school characteristics to adjust for factors, which are arguably outside of teachers' control, but affect student achievement gains (see Table 1.18 for a list of the covariates).

Teacher characteristics are omitted from the value-added estimates because of the possible correlation with student performance and evaluation ratings. For each teacher, we generate TVA estimates across classrooms and subjects. The reduced form equation for the estimation of the TVA is:

$$Y_{ijk} = \varphi_{000} + \beta_1 Y_{ijk(t-1)} + \beta_2 X_{ijk} + \gamma_{01} C_{jk} + \varphi_{001} S + \varepsilon_{ijk} + \rho_{0jk} + \phi_k \quad (1)$$

where

Y_{ijk} is the current test performance for student i taught by teacher j in school k;

$Y_{ijk(t-1)}$ is the prior test performance for student i, taught by teacher j in school k;

X_{ijk} is a vector of individual characteristics for student i taught by teacher j in school k;

C_{jk} is a vector of the teacher (classroom)-level characteristics for teacher j in school k;

S_k is a vector of school level characteristics common to all students and teachers in school k;

ε_{ijk} is the individual error term of student i taught by teacher j in school k;

ρ_{0jk} is the error between teachers within schools for teacher j in school k and yields shrunken empirical Bayes residuals that are used as the measure of teacher effectiveness for the analysis models; and

ϕ_k is the error between schools.

The analysis model utilizes ordinary least squares (OLS) regression and includes individual teacher value-added estimates from the measurement model. The equation is as follows:

$$ER_{jk} = \beta_0 + \beta_1 \rho_{0jk} + \epsilon_{jk} \qquad (2)$$

where

ER_{jk} is the composite or individual evaluation ratings of teacher j in school k and

ρ_{0jk} is the measure of teacher j's effectiveness in school k.

The second goal of this manuscript is to explore whether NCEES is a fair instrument to evaluate teachers' performance. That is, are there systematic differences in teacher evaluations based on teachers' race, ethnicity or gender? We employ an OLS regression model to estimate the relationship between our focal variables and evaluation ratings (see Equation 3). We also used cluster-adjusted standard errors at the school-level to account for the clustering of teachers within schools. The equation used to estimate the effects of race/ethnicity and gender on evaluation ratings is as follows:

$$ER_{jk} = \beta_0 + \beta_1 Race/Ethnicity_{jk} + \beta_2 Gender_{jk} + \tau_{jk} \qquad (3)$$

where

ER_{jk} is the composite or individual evaluation ratings of teacher j in school k $\qquad \tau_{jk}$ is the disturbance term that represents unexplained variation in evaluation ratings

As noted by Oppler and colleagues (1992), differences in evaluation ratings based on race, ethnicity, or gender do not necessarily imply evaluator bias; rather there may be differences in the actual performance of the members of those groups who are evaluated. In other words, performance may mediate the relationship between race/ethnicity and gender variables and evaluation ratings. After controlling for teachers' objective effectiveness, if race/ethnicity and gender coefficients are statistically significant this provides evidence consistent with systematic bias in the evaluation ratings. However, caution must be taken in interpreting the results because the effects are not causal, rather they describe a relationship between evaluation ratings and teacher characteristics after controlling for the best available objective measure of teacher performance.

To examine whether the naïve association, presented in Equation 3, weakens after adjusting for teachers' effectiveness we include the TVA estimates from Equation 1 along with other teacher covariates. As previously discussed, there may be initial differences in ratings across race/ethnicity and gender characteristics; however, this is possibly due to differences in teacher performance that may be correlated with membership in the racial/ethnic or gender categories. Put crudely, compared to White female teachers, Black female teachers may receive lower ratings as a group because they perform worse, not because of their race/ethnicity or gender. The equation used to estimate the effects these adjustments on evaluation ratings is as follows:

$$\text{ER}_{jk} = \beta_0 + \beta_1 Race/Ethnicity_{jk} + \beta_2 Gender_{jk} + \beta_3 \rho_{0jk} + \beta_4 Years\ of\ Experience_{jk}$$
$$+ \beta_5 Portal_{jk} + \omega_{jk} \quad (4)$$

After controlling for teacher performance by including the TVA and correlates of performance such as years of experience, there is still the possibility that systematic differences

remain. Including correlates such as years of experience is especially important if some of the race/ethnicity and gender groups have less experience and experience is associated with performance in ways that are not captured by the teacher's TVA score. We also include teachers' preparation portals to capture the quality of teachers' preparation. This is important if some of the race/ethnicity and gender groups are disproportionately prepared by one type of program and preparation programs are associated with teachers' performance.

To further examine these differences, we explore whether race/ethnicity and gender characteristics and experience of principals influence how they rate teachers. We include principals' race/gender groups, racial congruence, gender congruence, race/ethnicity – gender congruence and experience covariates in the analysis. This model is our preferred model. The equation is as follows:

$$\text{ER}_{jk} = \beta_0 + \beta_1 Race/Ethnicity_{jk} + \beta_2 Gender_{jk} + \beta_3 \rho_{0jk} + \beta_4 Years\ of\ Experience_{jk}$$
$$+ \beta_5 Portal_{jk} + \beta_6 Principal_{pk} + \omega_{jk} \quad (5)$$

where

$Principal_{pk}$ is a vector of principal characteristics, which includes race/ethnicity, gender, years of experience, and race and gender congruence[2] with the teacher of principal p in school k.

The covariate adjusted estimates specified to this point attempt to isolate the effect of teacher characteristics on evaluation ratings. Although this study does not make causal claims, there is still a potential threat to internal validity due to the nonrandom assignment of teachers

---

[2] The variance inflation factor (VIF) was used in each of the models to check for possible multicollinearity, especially among the three congruence variables – same race, same gender, and same race and gender. The presence of multicollinearity may make the test of significance overly conservative and create unstable coefficient estimates. The results from the VIF do not indicate the presence of multicollinearity among the variables (mean VIFs ranged from 1.62 to 1.92); therefore, all congruence variables are included in the model.

and unobserved school factors. More specifically, if principals' rating decisions are correlated with unobservable characteristics of teachers or the rating system, then our estimates of the relationship between evaluation ratings and teacher race/ethnicity and gender characteristics will be bias. To address this endogeneity issue, we include several covariates in our regression models. While conducting a randomized control trial is viewed as the gold standard in eliminating endogeneity threats, a rich set of covariates is shown to reduce bias comparable to randomized control trial (Shadish, Clark, & Steiner, 2010). However, unobserved characteristics of the school and principal that are correlated with teacher characteristics and evaluation ratings main may still be present; therefore, we also employ school/principal fixed effects to address the possible heterogeneity in ratings between principals and/or schools. For example, principals may vary in their ability to evaluate teachers using the NCEES. They may also vary in how they understand and operationalize the NCEES rubric. This variation may be due to observed or unobserved characteristics associated with schools and school leaders. The school/principal - fixed effects specification accounts for these observed and unobserved characteristics by controlling for time-invariant characteristics. It is important to note that even with the use of covariates and school-fixed effects estimates will still suffer from omitted variable bias.

There are two major limitations of the school-fixed effects specification. First, estimates are only obtained in schools where there is variation among teachers based on gender, ethnicity, and race; for example, schools staffed with only female teachers are excluded from the estimates as are schools staffed exclusively by White teachers. This exclusion may reduce the sample size making it difficult to detect an effect. Second, the subsample of schools that can be included may differ from the population of all North Carolina public middle schools. The school-fixed effects specification is as follows:

$$\text{ER}_{jk} = \beta_0 + \beta_1 Race/Ethnicity_{jk} + \beta_2 Gender_{jk} + \beta_3 \rho_{0jk} + \beta_4 Years\ of\ Experience_{jk}$$

$$+ \beta_5 Portal_{jk} + \beta_6 Principal_{pk} + \mu_k + \upsilon_{jk} \qquad (6)$$

where

> μ_k is the school-specific error component that is constant for every teacher in schools
>
> υ_{jk} is idiosyncratic error that varies across teachers within schools and
>
> *Principal* represents race and/or gender congruence variables.

In the fixed effects specification, principals' race/ethnicity-gender and experience are excluded from the model because these are time-invariant characteristics.

Using the intersectionality framework, we conduct analyses similar to those estimated from Equations 4, 5 and 6 using interaction terms to account for the intersection of race/ethnicity and gender (Dubrow, 2008). Here, we use what Weldon (2008) refers to as an "intersection-only" approach where the main effects of race/ethnicity and gender are not included in the models to focus on the interaction effects.

Results

Table 1.1 shows descriptive statistics from the sample of teachers in ELA, math, Algebra I and science. On average, teachers across subjects are rated proficient or slightly higher on the NCEES. In terms of race/ethnicity-gender, the majority of teachers in the models are White women, which is representative of teachers statewide. However, White women are underrepresented in science in comparison to the other subjects. In general, men are heavily concentrated in STEM related subjects. Black women are uniformly represented across subject areas. With the exception of science teachers, teachers' value-added estimates are below average, which is consistent with the evidence about the effectiveness of beginning teachers. The distribution of years of teaching experience was spread out across the five experience variables with the lowest proportion of teachers being those with one year of teaching experience. With the

exception of science teachers, in-state public undergraduate programs prepare the majority of teachers in models. Thirty-eight percent of science teachers are prepared via a lateral entry program. In addition, a larger percentage of TFA teachers teach science.

The sample schools are staffed with mostly White principals of which, women made up a slightly larger percentage than men. It is not surprising then that a large number of teachers are the same race or ethnicity as their principal – 64% to 68%. Additionally, there is a sizable range in the percentage of teachers that are the same race/ethnicity and gender as their principal across the subjects – 30% to 36%. On average, the principals in the sample have a little over 5 years of experience as a principal in North Carolina public schools.

Turning to the first goal in this manuscript, which examines the validity of the NCEES, the findings indicate a positive and statistically significant relationship between TVA estimates and teacher evaluation ratings, among ELA, math and science teachers (Table 1.2). The magnitude of the effect was small ranging between 0.09 and 0.15 points. The ratings for Algebra I teachers are not statistically significantly correlated with TVA estimates, which may be due to the small sample of teachers in the model; however, the direction of the estimate is positive. As an additional check, we examine the pairwise correlations between the TVA estimates and teachers' evaluation ratings for each subject. Tables 1.3-1.6 show positive and significant correlations between TVA estimates and composite ratings in all subjects except Algebra I, which is consistent with the findings from the regression analysis. As expected, the correlations between TVA and the composite ratings are modest at best, ranging from 0.16 in ELA to 0.28 in math. However, we see strong positive and significant correlations among the five standards.

The second goal of this study is to explore whether principals' evaluate teachers fairly using the NCEES. Overall, the findings suggest that conditional on TVA, systematic differences

based on teacher characteristics vary across subjects; although, the magnitude of the effects are small.

ELA. The unadjusted findings for ELA teachers show that White men are rated systematically lower than White women (see Table 1.7). After adjusting for TVA and other covariates, the coefficient on White male teachers is reduced marginally and remains statistically negative. This provides some evidence that TVA partially mediates the effect between teachers' race/ethnicity-gender and evaluation ratings. Once principal characteristics are added (Model C) on Table 1.7 also shows that nonBlack men of color are rated higher than White women by 0.53 points.

There is a monotonic relationship between the evaluation ratings and teachers' years of teaching experience. That is, evaluation ratings are higher as years of teaching experience increase. There are no significant findings for the preparation portals by which ELA teachers enter the classroom. After including the full set of controls shown in Model C on Table 1.7, the effect of teacher characteristics on evaluation ratings persist. Moreover, the principal characteristics indicate that White male principals rate ELA teachers higher than White female principals do, however, there is no significant finding regarding the racial, ethnic or gender congruence between the principal and ELA teachers.

The findings from the school-fixed effects model in Table 1.11 indicate that the within schools there is no systematic difference in ratings for White male teachers and nonBlack male teachers of color compared with White female teachers. However, the results for teachers' years of teaching experience are consistent with the OLS model.

Math. As shown in Table 1.8, Model A indicates that among math teachers, Black women receive lower ratings by about 0.19 points compared to White women. After adjusting

for teacher effectiveness and other covariates, the coefficient on Black female teachers is marginally reduced to 0.16 points and remains statistically negative. The reduction in the effect provides some evidence that TVA partially mediates the effect between teachers' race/ethnicity-gender and evaluation ratings.

Consistent with the ELA results, there is a positive, significant monotonic relationship between the evaluation ratings and years of teaching experience. In the models that include principal characteristics, compared to in-state public undergraduate prepared math teachers; those prepared through a lateral entry program, receive significantly lower ratings. Similar to the ELA findings, Model C on Table 1.8 shows that White male principals provide higher teacher evaluation ratings than White female principals; however, nonBlack female principals of color provide lower teacher evaluations ratings. Again, there is no significant finding regarding the racial, ethnic or gender congruence between the principal and math teachers. The results between the OLS and school-fixed effects models are similar. However, within schools lateral entry teacher are rated no differently than in-state public undergraduate prepared.

Algebra I. Table 1.9 shows that Black male teachers receive higher ratings than White female teachers (Model A). TVA is not significantly related to the ratings for Algebra 1 teachers. After adjusting for teacher effectiveness, these effects are reduced, but they remain statistically significant, suggesting that the inclusion of other teacher covariates partially mediate the relationship. After adjusting for principal characteristics, there is no longer a statistically significant difference in ratings between Black male and White female teachers. Algebra I teachers with 2 years of teaching experience receive higher ratings than teachers in their first year of teaching. Similar to the result shown in the ELA and math models, White male principals provide higher ratings than White female principals. There is no relationship between evaluation

ratings and Algebra I teachers' preparation portal. The school-fixed effects results indicate that there are no systematic differences in ratings based on teacher characteristics. The lack of within school findings may be a consequence of the reduced sample size.

Science. Table 1.10 shows that there is no statistically significant relationship between evaluation ratings and race/ethnic-gender teacher groups. Consistent with the ELA and math OLS models, teachers' experience has a positive relationship on evaluation ratings, although the coefficients do not monotonically increase. Compared to in-state public undergraduate prepared, teachers prepared by an out-of-state undergraduate preparation program and through a lateral entry program receive lower ratings, -0.30 and -0.22 points, respectively. However, these effects do not persist within schools. Conditional on the school-fixed effects, there are no systematic differences in ratings based on teacher characteristics, which may be the result of the reduced sample size.

Based on the findings to this point it appears that principals rate some racial/ethnic-gender groups of math and ELA teachers differently than others. However, this may be a function of differing group traits or dispositions among the standards. For example, male teachers may be rated higher on Standard 1: Teachers demonstrate leadership – given perceptions that men are "naturally" more effective leaders and supported by the general consensus that women face more barriers to becoming leaders (Eagly & Karau, 2002). Conversely, men may receive lower ratings on Standard 2: Teachers establish a respectful environment for a diverse population of students. For this standard, teachers are rated based on creating nurturing relationships, which social norms suggest is a trait of women more so than men; despite research that suggests nurturing is not a trait exclusive to women (Levit, 1998). If teachers from one racial/ethnic-gender group are rated higher or lower than others across all five

27

standards, there is strong evidence that principals may exhibit discriminatory rating behaviors. To examine this hypothesis we conduct additional analysis of the relationship between evaluation ratings and teacher characteristics, using each standard as a separate dependent variable. The results from this analysis are shown in Tables 1.12 and 1.13.

As you will recall, findings from Table 1.7 indicate that White male ELA teachers are rated lower and nonBlack males of color receive higher ratings than White female ELA teachers. To understand whether a single standard is driving the negative effect among White men, we ran our OLS regression with the full set of teacher and principal covariates. Table 1.12 shows that while the effect of White males is negative, these results are only significant on three of the five standards: Standard 1: Demonstrates leadership, Standard 2: Establishes a respectful classroom environment, and Standard 4: Facilitates student learning. NonBlack men of color receive significantly higher ratings than White women on three of the five standards: Standards 1: Demonstrates leadership, Standard 4: Facilitates student learning, and Standard 5: Reflects on teaching practices. Interestingly, White men and nonblack men of color receive opposing ratings on standards one and four (Demonstrates leadership and Facilitates student learning) when compared with White women. Additionally, White male principals provide statistically higher ratings on all standards except Standard 4: Knows the content than White female principals.

Recall that among math teachers, Black women receive lower ratings than White women. Table 1.13 shows that with the exception of Standard 4: Knows the content; Black women receive lower than their White female counterparts. OLS results also suggest that White male principals provide higher ratings than White female principals and nonBlack female principals of color provide significantly lower ratings than their White female counterparts. Table 1.13 shows that the results for White male and nonBlack female principals persist across all five standards.

28

Additional Specifications

Further examining the findings presented, we ran additional analyses modifying the

dependent variable and adjusting for district-level characteristics not captured in the OLS and

school-fixed effects models. Analyses are conducted using the full set of covariates previously

mentioned.

Variations of the outcome

Composite Rating of Standards 3 and 4. Standard 3: Knows the content and Standard 4:

Facilitates student learning are arguably more closely related to teacher value added estimates,

because of the "direct" effect of these knowledge, skills and dispositions on student learning.

Despite the similar correlations between all of the standards, we reassessed our analysis using a

unit-weighted composite rating score of these standards 3 and 4 as the dependent variable. The

results for each subject, shown in Tables 1.14 and 1.15, are consistent with the findings from the

composite ratings from all five standards.

Exceeds Proficient. We create a dichotomous dependent variable to examine whether the

relationship between evaluation ratings and teacher characteristics are motivated by   differences

at the upper or lower tails of the ratings distribution. Admittedly, there is little variation in

evaluation ratings across the standards. For these models, the dichotomous variable takes on a

value of 1 if the teacher receives a rating of distinguished or accomplished and 0 if the teacher

receives a rating of proficient, developing or not demonstrated.

To examine the relationship between the evaluation ratings and teacher characteristics,

we estimate odds ratios from logistic regressions controlling for teacher effectiveness, experience

and preparation and principal characteristics. The equation used to estimate the probability of

receiving a rating above proficiency is as follows:

$$P(Exceeds\ Proficient = 1)$$
$$= \frac{\exp{(\beta_0 + \beta_1 Race\_Ehnicity\_Gender + \beta_2 TVA + \beta_3 Teaching\ Experience + \beta_4 Portal + \beta_5 Principal}}{1 + \exp(\beta_0 + \beta_1 Race\_Ehnicity\_Gender + \beta_2 TVA + \beta_3 Teaching\ Experience + \beta_4 Portal + \beta_5 Princip}$$

The estimates shown in Table 1.16 indicate that the relationship between ratings and teacher characteristics is driven in part by the upper tail of the distribution. For example, the odds of receiving a rating above proficient are 6.88 times higher for nonBlack male ELA teachers than White female ELA teachers. The results persist after including school-fixed effects in the specification.[3]

District-Fixed Effects

Although the evaluation system is a statewide initiative, there may be some observed and unobserved time-invariant differences among districts. Because NCEES was implemented in three phases it is possible that districts in the first phase are more familiar with the rating system and thus use this tool more accurately or effectively. Additionally, North Carolina places the responsibility of training principals on local districts and schools; therefore, differences in the quality, duration and rigor may vary by district. To account for these unobserved district-level factors that might be correlated with evaluation ratings and teacher characteristics, we employ district-fixed effects. The results from the district-fixed effects models are consistent with the OLS effects.

Discussion

In this study, we extend the limited literature on principals' assessment of teachers by examining two important questions: whether North Carolina's teacher evaluation system is a valid instrument to assess teacher quality and whether the current system fairly assesses teacher performance. Our findings provide some evidence that the principal ratings of teachers are possibly valid as additional measures of teacher effectiveness. Simply put, principals give higher

---

[3] Due to the limited variation within schools, we are unable estimate Algebra I and Science models.

ratings to teachers who make larger student test score gains, with positive correlations between 0.09 and 0.15 in all subject areas except Algebra I. However, these effects sizes are much smaller than prior literature (Holtzapple, 2003; Milanowski, 2004). The findings provide some confidence that North Carolina's evaluation system may be a moderately valid tool for assessing teachers' performance. However, the question remains whether the standards-based evaluation systems fairly evaluates beginning teachers. While one may not expect to find rater bias within a more formalized, standards-based evaluation system; the results show some evidence that principals rate teachers differently based on race, ethnicity and gender. However, the magnitude of the effects is small. We find that White male ELA teachers and Black female math teachers lower ratings when compared with White women in those subject areas, which is consistent with findings from Jacob and Lefgren (2005). However, nonBlack male ELA teachers are receive higher ratings than White women.

While there is marginal evidence that some teacher groups within a given subject are not rated similar to their White females peers, overall, there is no statistically significant difference in ratings for the majority of the racial/ethnic-gender groups across subjects. For instance, among Algebra I and science teachers there were no racial/ethnic-gender differences in principals' evaluation of these teachers. Moreover, among math teachers there was only a negative statistically significant difference between Black and White women. The findings also show that systematic biases are not attributable to ratings by principals from a different racial, ethnic and/or gender group. It is possible that principals provide intangible benefits, such as encouragement, advocacy, and autonomy to teachers who share the same race/ethnicity and/or gender throughout the school year as found by Grissom and Keiser (2011); however, when evaluating teachers these benefits based on ascriptive characteristics are not evident.

While this study adds to the paucity of research on standards-based evaluation systems, the limitations of this study must also be considered. First, the use of a cross-sectional design using only data from the first year the evaluation system was fully implemented, allows us to only make inferences regarding the relationship between evaluation ratings and teacher characteristics, not causality. For example, being a Black female math teacher does not *cause* principals to them lower, there is simply a relationship between these teacher characteristics and evaluation ratings. Including additional years of evaluation rating data still will not allow us to make causal inferences; however, it is more likely to suggest a cause-and-effect relationship by the very nature of examining multiple time points. From a statistical perspective, a longitudinal study would also increase statistical power, which may be warranted, especially for the Algebra I and science models. Second, this study only examines middle grades. Expanding this analysis to the secondary level may provide useful information regarding the reliability of the NCESS across grades and subjects. Finally, the models used may not include all relevant variables in the estimation equations resulting in omitted variable bias. This bias will affect the interpretations that are made regarding the effect of evaluation ratings and teachers' racial/ethnic and gender. For example, this study does not include school contextual factors, which may influence evaluation ratings.

Considering these limitations, future research should examine the fairness of the evaluation system using more sophisticated quasi-experimental designs to causally determine whether racial/gender discrimination behaviors influence teachers' evaluations. A more in-depth analysis may be warranted to understand how policies designed to improve teacher performance are influenced by race, ethnicity and gender. For instance, future research could examine the mechanisms associated with evaluation ratings for male teachers, such as the expectations of

male teachers (i.e. disciplinarians) within the current female-dominated teaching profession. Studies could also examine gender differences in instructional practices.

Careful consideration is made in discussing the implications of the findings given the relatively small effect sizes and insignificance of the key variables of interest. From a statistical perspective, we know that with a sufficient sample size, any of the key variables of interest in our models can reach statistical significance. Furthermore, 1 in 20 times the results will be positive despite there being an actual relationship between evaluation ratings and teacher characteristics. This raises the issue of whether the effects are practically significant despite statistically significance or insignificance. Litschge, Vaughn and McCrea (2010) argues, "small effect sizes can have substantial practical value. This is particularly the case if a treatment is relatively inexpensive, is easy to execute, is politically feasible, and can be employed on a large scale, thereby affecting large numbers of individuals" (p. 22). With respect to our study, the practical value of addressing the possibility of systematic biases in evaluation ratings based on ascriptive teacher characteristics is important for school leaders and policy makers.

From a policy perspective, the preliminary results point to ways in which implementation may be improved upon. For example, North Carolina currently allows districts to develop their own evaluation training for raters, which can significantly vary in quality and effectiveness and lead to unwarranted biases. One possible solution is the development of a rigorous statewide training conducted by the NCDPI. In addition, to the "how to" of the evaluation rubric, the training could include topics related to diversity, equity, and fairness in evaluating a heterogeneous group of teachers. The inclusion of these topics within the training may prompt raters to better understand their own reactions to differences and consciously reflect on their own biases when rating teachers. Another solution to mitigate the potential bias is the use of multiple

raters. This solution may be more costly; however, it may improve the fairness of teacher evaluation ratings.

Table 1.1 Descriptive statistics for middle school teachers, 2010-2011

Variables	Reading Model	Math Model	Algebra I Model	Science Model
Effectiveness				
Teacher Value Added	-0.089 (0.940)	-0.102 (1.027)	-0.083 (0.917)	0.062 (0.906)
NCEES ratings				
Composite ratings	3.337 (0.551)	3.310 (0.555)	3.426 (0.584)	3.348 (0.498)
Std1: Demonstrate leadership	3.306 (0.649)	3.305 (0.652)	3.391 (0.688)	3.359 (0.604)
Std2: Establish positive environment	3.305 (0.614)	3.293 (0.628)	3.391 (0.667)	3.328 (0.591)
Std3: Content knowledge	3.321 (0.614)	3.268 (0.624)	3.464 (0.696)	3.350 (0.597)
Std4: Facilitate learning	3.415 (0.650)	3.349 (0.640)	3.435 (0.639)	3.359 (0.584)
Std5: Reflect on practices	3.336 (0.642)	3.334 (0.659)	3.449 (0.684)	3.343 (0.600)
Teacher socio-demographics				
Male	0.141	0.254	0.254	0.338
Black teacher	0.155	0.153	0.102	0.174
White teacher	0.808	0.798	0.832	0.780
Nonblack teacher of color	0.037	0.049	0.065	0.046
Black female	0.129	0.116	0.094	0.125
Black male	0.024	0.036	0.007	0.049
White female	0.690	0.589	0.601	0.498
White male	0.108	0.201	0.225	0.277
Nonblack female teacher of color	0.029	0.034	0.043	0.036
Nonblack male teacher of color	0.007	0.015	0.022	0.009
Experience and Preparation				
Zero years of teaching experience	0.221	0.246	0.210	0.207
One year of teaching experience	0.160	0.184	0.152	0.149
Two years of teaching experience	0.259	0.235	0.188	0.231
Three years of teaching experience	0.224	0.223	0.319	0.267
Four years of teaching experience	0.136	0.111	0.130	0.146
In-state undergrad prepared	0.312	0.354	0.391	0.261
Out of state undergrad prepared	0.208	0.204	0.188	0.198
Teach for America prepared	0.036	0.041	0.036	0.061
Lateral entry	0.225	0.229	0.261	0.377
Other preparation	0.220	0.172	0.123	0.103
Principal Characteristics				
Black female principal	0.170	0.178	0.174	0.140
Black male principal	0.113	0.114	0.116	0.159
White female principal	0.383	0.356	0.355	0.354
White male principal	0.324	0.334	0.355	0.323
Nonblack female principal of color	0.003	0.002	0.000	0.000
Nonblack male principal of color	0.004	0.006	0.000	0.015
Years of experience	5.421	5.115	5.062	5.065
Same race/ethnicity	0.668	0.682	0.638	0.647
Same gender	0.536	0.529	0.493	0.480
Same race/ethnicity & gender	0.356	0.358	0.297	0.313

Note: Standard deviations are in parentheses.

Table 1.2 Relationship between Teacher Value Added Estimates and Principals' Evaluation Ratings

Composite Evaluation Rating	Reading Model	Math Model	Algebra I Model	Science Model
Teacher Value Added	0.096	0.149	0.060	0.092
	(0.018)*	(0.015)*	(0.047)	(0.028)*
Constant	3.345	3.325	3.431	3.342
	(0.022)*	(0.024)*	(0.055)*	(0.031)*
R^2	0.03	0.08	0.01	0.03
N	1,006	978	138	329

Notes: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. * indicates significance at the $p < 0.05$ level.

Table 1.3 Pearson's Correlation of TVA estimates and Evaluation Ratings in ELA

	Teacher value-added	Composite evaluation rating	Standard 1 Demonstrate leadership	Standard 2 Respectful classroom environment	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Teaching
Teacher value added	1.0000						
Composite evaluation rating	0.1644*	1.0000					
Standard 1 - Demonstrate leadership	0.1126*	0.8610*	1.0000				
Standard 2 - Respectful classroom environment	0.1410*	0.8863*	0.7034*	1.0000			
Standard 3 - Content Knowledge	0.1562*	0.8670*	0.6863*	0.7393*	1.0000		
Standard 4 - Facilitate Learning	0.1453*	0.8631*	0.6400*	0.7224*	0.6809*	1.0000	
Standard 5 - Reflect on Teaching	0.1606*	0.8714*	0.7076*	0.6983*	0.6746*	0.7034*	1.0000

* indicates significance at the p<0.05 level.

Table 1.4 Pearson's Correlation of TVA estimates and Evaluation Ratings in Math

	Teacher value-added	Composite evaluation rating	Standard 1 Demonstrate leadership	Standard 2 Respectful classroom environment	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Teaching
Teacher value added	1.0000						
Composite evaluation rating	0.2762*	1.0000					
Standard 1 - Demonstrate leadership	0.2463*	0.8516*	1.0000				
Standard 2 - Respectful classroom environment	0.2473*	0.8917*	0.6919*	1.0000			
Standard 3 - Content Knowledge	0.2485*	0.8577*	0.6552*	0.7510*	1.0000		
Standard 4 - Facilitate Learning	0.2193*	0.8597*	0.6381*	0.7109*	0.6686*	1.0000	
Standard 5 - Reflect on Teaching	0.2365*	0.8744*	0.6995*	0.7184*	0.6541*	0.7092*	1.0000

* indicates significance at the p<0.05 level.

Table 1.5 Pearson's Correlation of TVA estimates and Evaluation Ratings in Algebra I

	Teacher value-added	Composite evaluation rating	Standard 1 Demonstrate leadership	Standard 2 Respectful classroom environment	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Teaching
Teacher value added	1.0000						
Composite evaluation rating	0.0938	1.0000					
Standard 1 - Demonstrate leadership	0.0508	0.8493*	1.0000				
Standard 2 - Respectful classroom environment	0.1142	0.8768*	0.6663*	1.0000			
Standard 3 - Content Knowledge	0.0666	0.8928*	0.7159*	0.7548*	1.0000		
Standard 4 - Facilitate Learning	0.1249	0.8375*	0.6064*	0.6603*	0.6926*	1.0000	
Standard 5 - Reflect on Teaching	0.0537	0.8722*	0.6781*	0.7160*	0.6937*	0.6853*	1.0000

* indicates significance at the p<0.05 level.

Table 1.6 Pearson's Correlation of TVA estimates and Evaluation Ratings in Science

	Teacher value added	Composite evaluation rating	Standard 1 Demonstrate leadership	Standard 2 Respectful classroom environment	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Teaching
Teacher value added	1.0000						
Composite evaluation rating	0.1670*	1.0000					
Standard 1 - Demonstrate leadership	0.1722*	0.8344*	1.0000				
Standard 2 - Respectful classroom environment	0.1497*	0.8765*	0.6855*	1.0000			
Standard 3 - Content Knowledge	0.1628*	0.8266*	0.5815*	0.6938*	1.0000		
Standard 4 - Facilitate Learning	0.1168*	0.8260*	0.6027*	0.6478*	0.6107*	1.0000	
Standard 5 - Reflect on Teaching	0.0970	0.8227*	0.6178*	0.6438*	0.5745*	0.6047*	1.0000

* indicates significance at the p<0.05 level.

Table 1.7 Relationship between teacher evaluation ratings and teacher characteristics in ELA

Dependent Variable: Composite Evaluation Rating	Model A	Model B	Model C
Black women teachers	-0.043	-0.085	-0.050
	(0.051)	(0.052)	(0.057)
Black men teachers	-0.136	-0.138	-0.136
	(0.111)	(0.103)	(0.103)
White men teachers	-0.172	-0.160	-0.144
	(0.059)*	(0.053)*	(0.054)*
NonBlack women teachers of color	0.025	-0.069	-0.005
	(0.108)	(0.099)	(0.115)
NonBlack men teachers of color	0.296	0.370	0.533
	(0.240)	(0.155)*	(0.186)*
Teacher value added		0.070	0.074
		(0.018)*	(0.018)*
One year of teaching experience		0.264	0.235
		(0.054)*	(0.055)*
Two years of teaching experience		0.407	0.403
		(0.047)*	(0.047)*
Three years of teaching experience		0.500	0.473
		(0.050)*	(0.050)*
Four years of teaching experience		0.606	0.581
		(0.062)*	(0.066)*
Out of state undergrad prepared		-0.033	-0.026
		(0.050)	(0.054)
Teach for America prepared		-0.048	-0.041
		(0.071)	(0.072)
Lateral Entry prepared		-0.025	-0.054
		(0.049)	(0.051)
Other preparation		-0.047	-0.061
		(0.045)	(0.048)
Black women principals			0.033
			(0.071)
Black men principals			0.044
			(0.078)
White men principals			0.161
			(0.061)*
NonBlack women principals of color			-0.036
			(0.241)
NonBlack men principals of color			0.034
			(0.463)
Principals years of experience			0.008
			(0.005)
Same race/ethnicity			0.036
			(0.065)
Same gender			-0.016
			(0.075)
Same race/ethnicity & gender			0.048
			(0.083)
Constant	3.361	3.056	2.943
	(0.025)*	(0.044)*	(0.086)*
R^2	0.01	0.18	0.19
N	1,006	1,006	925

Notes: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. * indicates significance at the $p<0.05$ level

Table 1.8 Relationship between teacher evaluation ratings and teacher characteristics in Math

Dependent Variable: Composite Evaluation Ratings	Model A	Model B	Model C
Black women teachers	-0.190	-0.171	-0.162
	(0.056)*	(0.055)*	(0.059)*
Black men teachers	-0.072	-0.115	-0.030
	(0.092)	(0.088)	(0.094)
White men teachers	-0.076	-0.072	-0.084
	(0.046)	(0.043)	(0.045)
NonBlack women teachers of color	0.127	0.113	0.156
	(0.086)	(0.083)	(0.097)
NonBlack men teachers of color	-0.039	-0.190	-0.231
	(0.146)	(0.145)	(0.136)
Teacher value added		0.122	0.120
		(0.015)*	(0.015)*
One year of teaching experience		0.242	0.218
		(0.049)*	(0.052)*
Two years of teaching experience		0.397	0.378
		(0.049)*	(0.048)*
Three years of teaching experience		0.458	0.430
		(0.049)*	(0.051)*
Four years of teaching experience		0.532	0.514
		(0.063)*	(0.068)*
Out of state undergrad prepared		-0.090	-0.082
		(0.047)	(0.048)
Teach for America prepared		0.119	0.165
		(0.076)	(0.098)
Lateral Entry prepared		-0.088	-0.095
		(0.046)	(0.048)*
Other preparation		-0.081	-0.092
		(0.046)	(0.048)
Black women principals			0.025
			(0.076)
Black men principals			0.088
			(0.091)
White men principals			0.161
			(0.057)*
NonBlack women principals of color			-0.328
			(0.077)*
NonBlack men principals of color			0.083
			(0.206)
Principals years of experience			0.010
			(0.004)*
Same race/ethnicity			0.077
			(0.066)
Same gender			0.019
			(0.067)
Same race/ethnicity & gender			-0.026
			(0.073)
Constant	3.346	3.108	2.959
	(0.031)*	(0.044)*	(0.080)*
R^2	0.02	0.20	0.22
N	978	978	904

Notes: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses.*indicates significance at the $p<0.05$ level.

Table 1.9 Relationship between teacher evaluation ratings and teacher characteristics in Algebra I

Dependent Variable: Composite Evaluation Rating	Model A	Model B	Model C
Black women teachers	0.010	0.039	0.040
	(0.205)	(0.186)	(0.220)
Black men teachers	0.533	0.331	0.525
	(0.073)*	(0.150)*	(0.304)
White men teachers	-0.176	-0.132	-0.171
	(0.113)	(0.112)	(0.132)
NonBlack women teachers of color	-0.067	-0.123	-0.070
	(0.172)	(0.171)	(0.256)
NonBlack men teachers of color	-0.133	-0.311	-0.309
	(0.288)	(0.325)	(0.357)
Teacher value added		0.056	0.036
		(0.054)	(0.061)
One year of teaching experience		0.321	0.265
		(0.141)*	(0.182)
Two years of teaching experience		0.614	0.564
		(0.182)*	(0.198)*
Three years of teaching experience		0.562	0.492
		(0.129)*	(0.137)*
Four years of teaching experience		0.530	0.413
		(0.158)*	(0.176)*
Out of state undergrad prepared		-0.009	0.033
		(0.152)	(0.160)
Teach for America prepared		-0.009	-0.050
		(0.116)	(0.152)
Lateral Entry prepared		0.025	0.066
		(0.148)	(0.147)
Other preparation		-0.085	-0.071
		(0.144)	(0.156)
Black women principals			0.229
			(0.245)
Black men principals			0.199
			(0.221)
White men principals			0.341
			(0.142)*
Principals years of experience			-0.013
			(0.015)
Same race/ethnicity			0.038
			(0.276)
Same gender			-0.118
			(0.175)
Same race/ethnicity & gender			0.086
			(0.225)
Constant	3.467	3.059	3.003
	(0.073)*	(0.120)*	(0.297)*
R^2	0.02	0.19	0.24
N	138	138	129

Notes: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses. * indicates significance at the $p<0.05$ level.

Table 1.10 Relationship between teacher evaluation ratings and teacher characteristics in Science

Dependent Variable: Composite Evaluation Rating	Model A	Model B	Model C
Black women teachers	-0.113	-0.139	-0.171
	(0.085)	(0.087)	(0.090)
Black men teachers	-0.031	0.039	-0.001
	(0.117)	(0.101)	(0.130)
White men teachers	-0.112	-0.111	-0.103
	(0.062)	(0.059)	(0.064)
NonBlack women teachers of color	-0.273	-0.143	-0.163
	(0.166)	(0.153)	(0.177)
NonBlack men teachers of color	-0.206	-0.170	-0.067
	(0.288)	(0.374)	(0.381)
Teacher value added		0.069	0.071
		(0.027)*	(0.029)*
One year of teaching experience		0.227	0.209
		(0.082)*	(0.086)*
Two years of teaching experience		0.214	0.176
		(0.077)*	(0.082)*
Three years of teaching experience		0.423	0.380
		(0.081)*	(0.087)*
Four years of teaching experience		0.476	0.461
		(0.100)*	(0.110)*
Out of state undergrad prepared		-0.149	-0.160
		(0.079)	(0.086)
Teach for America prepared		-0.188	-0.303
		(0.086)*	(0.107)*
Lateral Entry prepared		-0.213	-0.222
		(0.076)*	(0.084)*
Other preparation		-0.150	-0.164
		(0.092)	(0.096)
Black women principals			0.068
			(0.111)
Black men principals			0.088
			(0.105)
White men principals			0.035
			(0.080)
NonBlack men principals of color			-0.090
			(0.329)
Principals years of experience			0.006
			(0.007)
Same race/ethnicity			0.044
			(0.083)
Same gender			0.114
			(0.100)
Same race/ethnicity & gender			-0.039
			(0.115)
Constant	3.406	3.267	3.174
	(0.040)*	(0.074)*	(0.125)*
R^2	0.02	0.16	0.18
N	329	329	307

Notes: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses. * indicates significance at the $p<0.05$ level.

Table 1.11 Relationship between teacher evaluation ratings and teacher characteristics using school-fixed effects

	ELA Model	Math Model	Algebra I Model	Science Model
Black women teachers	0.051	-0.181	-0.704	-0.004
	(0.095)	(0.078)*	(1.076)	(0.212)
Black men teachers	-0.130	-0.111	--	0.019
	(0.188)	(0.135)	--	(0.287)
White men teachers	-0.103	-0.073	0.169	-0.115
	(0.066)	(0.062)	(0.318)	(0.128)
NonBlack women teachers of color	0.079	0.017	-0.776	0.020
	(0.170)	(0.104)	(0.898)	(0.281)
NonBlack men teachers of color	0.285	-0.419	--	0.354
	(0.438)	(0.227)	--	(0.264)
Teacher value added	0.062	0.103	0.038	0.065
	(0.025)*	(0.021)*	(0.130)	(0.044)
One year of teaching experience	0.217	0.296	0.642	-0.133
	(0.083)*	(0.076)*	(0.643)	(0.262)
Two years of teaching experience	0.369	0.354	0.156	-0.084
	(0.065)*	(0.065)*	(0.584)	(0.211)
Three years of teaching experience	0.494	0.463	0.359	0.172
	(0.073)*	(0.071)*	(0.777)	(0.178)
Four years of teaching experience	0.523	0.437	-0.378	0.100
	(0.097)*	(0.101)*	(0.565)	(0.221)
Out of state undergrad prepared	0.040	-0.026	0.374	-0.061
	(0.078)	(0.060)	(0.791)	(0.248)
Teach for America prepared	0.116	0.294	--	0.073
	(0.179)	(0.186)	--	(0.293)
Lateral Entry prepared	0.037	-0.046	-0.036	-0.144
	(0.077)	(0.060)	(0.562)	(0.192)
Other preparation	-0.053	-0.048	0.297	-0.239
	(0.073)	(0.064)	(0.515)	(0.205)
Same race/ethnicity	0.075	0.008	--	-0.022
	(0.103)	(0.083)	--	(0.190)
Same gender	0.019	-0.000	0.715	-0.073
	(0.102)	(0.095)	(0.493)	(0.234)
Same race/ethnicity & gender	0.082	0.008	-0.707	-0.026
	(0.115)	(0.103)	(0.782)	(0.247)
Constant	2.946	3.090	3.074	3.495
	(0.106)*	(0.083)*	(0.513)*	(0.216)*
R^2	0.71	0.78	0.97	0.97

Notes: Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts.
* indicates significance at the p<0.05 level.

Table 1.12. Relationship between teacher evaluation ratings and teacher characteristics - ELA, by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Black women teachers	-0.071	-0.031	-0.040	-0.061	-0.049
	(0.067)	(0.063)	(0.062)	(0.071)	(0.066)
Black men teachers	-0.082	-0.078	-0.161	-0.214	-0.144
	(0.117)	(0.118)	(0.115)	(0.115)	(0.111)
White men teachers	-0.133	-0.154	-0.114	-0.194	-0.122
	(0.064)*	(0.068)*	(0.061)	(0.063)*	(0.063)
NonBlack women teachers of color	-0.063	-0.052	-0.034	0.035	0.091
	(0.120)	(0.130)	(0.145)	(0.153)	(0.136)
NonBlack men teachers of color	0.795	0.296	0.436	0.356	0.781
	(0.271)*	(0.156)	(0.247)	(0.123)*	(0.291)*
Teacher value added	0.060	0.070	0.081	0.076	0.082
	(0.021)*	(0.020)*	(0.021)*	(0.021)*	(0.022)*
One year of teaching experience	0.175	0.229	0.246	0.265	0.257
	(0.067)*	(0.062)*	(0.057)*	(0.068)*	(0.068)*
Two years of teaching experience	0.398	0.370	0.391	0.434	0.423
	(0.058)*	(0.054)*	(0.056)*	(0.058)*	(0.055)*
Three years of teaching experience	0.469	0.421	0.507	0.489	0.481
	(0.059)*	(0.061)*	(0.058)*	(0.059)*	(0.059)*
Four years of teaching experience	0.527	0.578	0.583	0.630	0.588
	(0.076)*	(0.074)*	(0.074)*	(0.075)*	(0.076)*
Out of state undergrad prepared	-0.027	-0.012	-0.026	-0.074	0.010
	(0.061)	(0.060)	(0.062)	(0.062)	(0.065)
Teach for America prepared	0.064	-0.067	-0.094	-0.133	0.026
	(0.089)	(0.090)	(0.083)	(0.106)	(0.117)
Lateral Entry prepared	-0.067	-0.041	-0.051	-0.060	-0.052
	(0.061)	(0.057)	(0.059)	(0.060)	(0.057)
Other preparation	-0.078	-0.095	-0.018	-0.057	-0.059
	(0.054)	(0.057)	(0.056)	(0.059)	(0.053)
Black women principals	-0.027	0.052	0.052	0.033	0.056
	(0.081)	(0.077)	(0.079)	(0.083)	(0.086)
Black men principals	0.040	0.038	-0.043	0.096	0.091

	(0.090)	(0.089)	(0.088)	(0.093)	(0.083)
White men principals	0.146	0.155	0.106	0.206	0.191
	(0.071)*	(0.072)*	(0.067)	(0.070)*	(0.069)*
NonBlack women principals of color	0.027	0.018	-0.034	-0.164	-0.030
	(0.261)	(0.239)	(0.223)	(0.243)	(0.259)
NonBlack men principals of color	-0.010	-0.307	-0.106	0.390	0.202
	(0.664)	(0.505)	(0.641)	(0.344)	(0.231)
Principals years of experience	0.002	0.012	0.008	0.010	0.008
	(0.005)	(0.006)*	(0.005)	(0.006)	(0.005)
Same race/ethnicity	0.084	0.050	-0.003	0.019	0.031
	(0.075)	(0.075)	(0.069)	(0.079)	(0.074)
Same gender	0.051	-0.070	-0.049	0.015	-0.029
	(0.086)	(0.083)	(0.083)	(0.090)	(0.089)
Same race/ethnicity & gender	0.038	0.044	0.063	0.026	0.069
	(0.096)	(0.092)	(0.091)	(0.100)	(0.099)
Constant	2.918	2.941	2.972	2.992	2.890
	(0.098)*	(0.099)*	(0.095)*	(0.105)*	(0.094)*
R^2	0.13	0.15	0.15	0.16	0.15
N	925	925	925	925	925

Note: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. *indicates significance at the $p<0.05$ level.

Table 1.13. Relationship between teacher evaluation ratings and teacher characteristics – Math, by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Black women teachers	-0.162	-0.177	-0.135	-0.170	-0.164
	(0.067)*	(0.067)*	(0.076)	(0.071)*	(0.070)*
Black men teachers	-0.026	-0.101	-0.155	0.030	0.099
	(0.118)	(0.113)	(0.112)	(0.115)	(0.119)
White men teachers	-0.097	-0.097	-0.114	-0.059	-0.055
	(0.054)	(0.051)	(0.051)*	(0.056)	(0.055)
NonBlack women teachers of color	0.065	0.237	0.162	0.146	0.169
	(0.122)	(0.113)*	(0.100)	(0.114)	(0.119)
NonBlack men teachers of color	-0.161	-0.235	-0.249	-0.270	-0.239
	(0.125)	(0.160)	(0.122)*	(0.159)	(0.161)
Teacher value added	0.125	0.122	0.123	0.110	0.121
	(0.019)*	(0.018)*	(0.018)*	(0.019)*	(0.019)*
One year of teaching experience	0.203	0.258	0.173	0.208	0.248
	(0.063)*	(0.063)*	(0.062)*	(0.061)*	(0.058)*
Two years of teaching experience	0.350	0.406	0.348	0.368	0.417
	(0.061)*	(0.059)*	(0.056)*	(0.057)*	(0.053)*
Three years of teaching experience	0.396	0.467	0.452	0.360	0.476
	(0.062)*	(0.061)*	(0.062)*	(0.057)*	(0.059)*
Four years of teaching experience	0.540	0.473	0.493	0.477	0.588
	(0.084)*	(0.076)*	(0.073)*	(0.082)*	(0.088)*
	(0.057)	(0.055)*	(0.059)	(0.058)	(0.055)*
Teach for America prepared	0.274	0.134	0.033	0.202	0.181
	(0.132)*	(0.094)	(0.102)	(0.125)	(0.138)
Lateral Entry prepared	-0.088	-0.095	-0.052	-0.080	-0.159
	(0.061)	(0.054)	(0.054)	(0.057)	(0.059)*
Other preparation	-0.116	-0.077	-0.022	-0.077	-0.166
	(0.055)*	(0.059)	(0.055)	(0.059)	(0.058)*
Black women principals	-0.039	0.042	0.034	0.031	0.058
	(0.085)	(0.082)	(0.089)	(0.084)	(0.084)

Black men principals	0.135	0.125	-0.027	0.026	0.180
	(0.098)	(0.101)	(0.096)	(0.105)	(0.103)
White men principals	0.175	0.187	0.128	0.175	0.139
	(0.063)*	(0.063)*	(0.062)*	(0.065)*	(0.066)*
NonBlack women principals of color	-0.255	-0.360	-0.362	-0.379	-0.284
	(0.097)*	(0.109)*	(0.082)*	(0.090)*	(0.100)*
NonBlack men principals of color	0.106	-0.053	0.102	0.011	0.247
	(0.364)	(0.237)	(0.271)	(0.162)	(0.129)
Principals years of experience	0.009	0.011	0.016	0.011	0.004
	(0.006)	(0.005)*	(0.005)*	(0.005)*	(0.006)
Same race/ethnicity	0.123	0.096	0.023	0.017	0.129
	(0.075)	(0.075)	(0.083)	(0.077)	(0.078)
Same gender	0.043	0.050	-0.001	-0.021	0.023
	(0.084)	(0.073)	(0.076)	(0.079)	(0.081)
Same race/ethnicity & gender	-0.087	-0.032	-0.018	0.034	-0.025
	(0.093)	(0.082)	(0.087)	(0.084)	(0.087)
Constant	2.964	2.889	2.946	3.039	2.957
	(0.101)*	(0.090)*	(0.091)*	(0.088)*	(0.097)*
R^2	0.17	0.19	0.18	0.15	0.18
N	904	904	904	904	904

Note: Each column is an OLS regression. Standard errors are robust to clustering at the school level in parentheses.*indicates significance at the $p<0.05$ level.

Table 1.14 Relationship between teacher evaluation ratings and teacher characteristics – OLS Models

Standards 3 & 4 Composite Rating	ELA Model	Math Model	Algebra I Model	Science Model
Black women teachers	-0.051	-0.153	0.045	-0.161
	(0.061)	(0.067)*	(0.201)	(0.103)
Black men teachers	-0.188	-0.062	0.560	0.057
	(0.101)	(0.099)	(0.319)	(0.144)
White men teachers	-0.154	-0.087	-0.236	-0.081
	(0.055)*	(0.048)	(0.152)	(0.072)
NonBlack women teachers of color	0.000	0.154	-0.059	-0.144
	(0.135)	(0.103)	(0.309)	(0.191)
NonBlack men teachers of color	0.396	-0.260	-0.367	0.027
	(0.157)*	(0.128)*	(0.365)	(0.444)
Teacher value added	0.079	0.116	0.040	0.072
	(0.019)*	(0.016)*	(0.075)	(0.036)*
One year of teaching experience	0.256	0.190	0.338	0.205
	(0.057)*	(0.055)*	(0.185)	(0.097)*
Two years of teaching experience	0.412	0.358	0.554	0.141
	(0.050)*	(0.051)*	(0.200)*	(0.084)
Three years of teaching experience	0.498	0.406	0.399	0.389
	(0.052)*	(0.054)*	(0.141)*	(0.095)*
Four years of teaching experience	0.606	0.485	0.454	0.365
	(0.068)*	(0.069)*	(0.167)*	(0.115)*
Out of state undergrad prepared	-0.050	-0.046	0.080	-0.148
	(0.057)	(0.053)	(0.176)	(0.094)
Teach for America prepared	-0.114	0.117	-0.147	-0.339
	(0.073)	(0.098)	(0.221)	(0.123)*
Lateral Entry prepared	-0.055	-0.066	0.020	-0.188
	(0.055)	(0.050)	(0.161)	(0.086)*
Other preparation	-0.038	-0.050	-0.106	-0.160
	(0.052)	(0.050)	(0.164)	(0.104)
Black women principals	0.042	0.033	0.176	0.040
	(0.074)	(0.082)	(0.235)	(0.124)
Black men principals	0.026	-0.001	0.155	0.048

	(0.083)	(0.096)	(0.218)	(0.108)
White men principals	0.156	0.151	0.277	0.061
	(0.063)*	(0.059)*	(0.155)	(0.084)
NonBlack women principals	-0.099	-0.370		
	(0.231)	(0.080)*		
NonBlack men principals	0.142	0.057	--	-0.081
	(0.490)	(0.183)	--	(0.321)
Principal's years of experience	0.009	0.013	-0.013	0.006
	(0.006)	(0.005)*	(0.016)	(0.007)
Same race/ethnicity	0.008	0.020	0.058	-0.058
	(0.067)	(0.073)	(0.286)	(0.094)
Same gender	-0.017	-0.011	-0.083	0.036
	(0.078)	(0.071)	(0.192)	(0.112)
Same race/ethnicity & gender	0.044	0.008	-0.007	0.074
	(0.087)	(0.077)	(0.246)	(0.127)
Constant	2.982	2.993	3.112	3.245
	(0.091)*	(0.081)*	(0.284)*	(0.137)*
R^2	0.18	0.19	0.21	0.15
N	925	904	129	307

Notes: Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. *indicates significance at the $p<0.05$ level.

Table 1.15 Relationship between teacher evaluation ratings and teacher characteristics – School Fixed Effects Models

Standards 3 & 4 Composite Rating	ELA Model	Math Model	Algebra I Model	Science Model
Black women teachers	0.066	-0.169	-1.203	0.057
	(0.096)	(0.083)*	(1.312)	(0.281)
Black men teachers	-0.176	-0.116	--	0.243
	(0.191)	(0.137)	--	(0.384)
White men teachers	-0.087	-0.084	0.223	-0.166
	(0.073)	(0.068)	(0.450)	(0.188)
NonBlack women teachers of color	0.105	0.015	-0.900	-0.055
	(0.234)	(0.121)	(1.331)	(0.433)
NonBlack men teachers of color	0.249	-0.434	--	0.403
	(0.436)	(0.217)*	--	(0.325)
Teacher value added	0.057	0.093	0.111	0.051
	(0.026)*	(0.022)*	(0.212)	(0.069)
One year of teaching experience	0.240	0.251	0.617	-0.398
	(0.083)*	(0.080)*	(0.819)	(0.376)
Two years of teaching experience	0.363	0.329	0.068	-0.182
	(0.071)*	(0.068)*	(0.730)	(0.224)
Three years of teaching experience	0.513	0.456	0.059	0.118
	(0.081)*	(0.077)*	(0.870)	(0.206)
Four years of teaching experience	0.545	0.385	-0.506	-0.078
	(0.105)*	(0.106)*	(0.832)	(0.289)
Out of state undergrad prepared	0.014	-0.001	0.300	-0.006
	(0.091)	(0.071)	(1.012)	(0.286)
Teach for America prepared	0.069	0.258	--	0.140
	(0.189)	(0.181)	--	(0.306)
Lateral Entry prepared	0.061	-0.024	0.050	-0.071
	(0.088)	(0.068)	(0.846)	(0.162)
Other preparation	-0.053	-0.028	0.282	-0.292
	(0.084)	(0.067)	(0.670)	(0.274)

Same race/ethnicity	0.072	-0.001	--	-0.187
	(0.110)	(0.092)	--	(0.265)
Same gender	0.044	0.030	1.286	-0.192
	(0.106)	(0.109)	(0.684)	(0.362)
Same race/ethnicity & gender	0.056	-0.018	-1.216	0.070
	(0.122)	(0.117)	(0.974)	(0.357)
Constant	2.963	3.094	3.138	3.700
	(0.114)*	(0.082)*	(0.763)*	(0.279)*
R^2	0.68	0.75	0.96	0.95

Notes: Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. *indicates significance at the $p<0.05$ level.

Table 1.16 Odds ratios for being rated above proficient

Dependent Variable: Exceeds Proficiency =1	ELA Model	Math Model	Algebra I Model	Science Model
Black women teachers	0.845	0.598	1.072	0.369*
	(0.238)	(0.188)	(0.855)	(0.164)
Black men teachers	0.745	1.348	0.593	1.326
	(0.329)	(0.592)	(0.338)	(0.992)
White men teachers	0.765	0.759	0.683	0.556
	(0.187)	(0.173)	(0.684)	(0.182)
NonBlack women teachers of color	1.111	2.946*	0.384	0.164
	(0.533)	(1.292)	(0.595)	(0.188)
NonBlack men teachers of color	6.880*	0.631	6.882*	1.181
	(5.117)	(0.386)	(6.520)	(1.887)
Teacher value added	1.229*	1.628*	0.933	1.166
	(0.095)	(0.135)	(0.240)	(0.155)
One year of teaching experience	2.626*	2.535*	6.882*	3.327*
	(0.765)	(0.759)	(6.520)	(1.626)
Two years of teaching experience	4.459*	4.735*	6.696*	1.785
	(1.150)	(1.296)	(5.841)	(0.886)
Three years of teaching experience	6.082*	5.572*	7.236*	6.061*
	(1.629)	(1.508)	(5.525)	(2.789)
Four years of teaching experience	8.234*	8.083*	5.539	6.215*
	(2.494)	(2.496)	(4.959)	(3.132)
Out of state undergrad prepared	0.955	0.906	2.060	0.340*
	(0.215)	(0.195)	(1.357)	(0.143)
Teach for America prepared	0.743	1.619	0.402	0.137*
	(0.383)	(0.884)	(0.587)	(0.106)
Lateral Entry prepared	0.813	0.717	1.816	0.450*
	(0.170)	(0.158)	(1.010)	(0.156)
Other preparation	1.035	0.760	0.872	0.492
	(0.211)	(0.174)	(0.625)	(0.245)
Black women principals	0.898	1.726	2.737	1.761
	(0.318)	(0.527)	(2.401)	(0.833)
Black men principals	1.184	1.969	3.674	1.057
	(0.421)	(0.906)	(3.471)	(0.575)
White men principals	1.211	2.263*	3.702*	1.086
	(0.339)	(0.596)	(2.216)	(0.416)
NonBlack women principals	0.741	--	--	--
	(0.803)	--	--	--
NonBlack men principals	4.989	1.241	--	0.923
	(7.585)	(1.096)	--	(1.267)
Principal's years of experience	1.046*	1.056*	0.965	1.027
	(0.024)	(0.019)	(0.060)	(0.034)
Same race/ethnicity	1.623	2.280*	2.296	1.185
	(0.462)	(0.771)	(1.941)	(0.573)
Same gender	1.106	1.532	0.974	1.549
	(0.356)	(0.493)	(0.763)	(0.812)

Same race/ethnicity & gender	0.670	0.615	0.469	0.673
	(0.252)	(0.223)	(0.438)	(0.407)
Observations	925	902	128	307

Notes: ELA=English language arts; z-scores are shown in the parentheses; * indicates significance at the $p<0.05$ level.

Table 1.17 Relationship between teacher evaluation ratings and teacher characteristics using district-fixed effects

	Reading Model	Math Model	Algebra I Model	Science Model
Black women teachers	0.039	-0.142	0.185	-0.113
	(0.063)	(0.061)*	(0.279)	(0.094)
Black men teachers	-0.068	-0.025	0.791	-0.024
	(0.124)	(0.100)	(0.383)*	(0.164)
White men teachers	-0.136	-0.059	-0.108	-0.108
	(0.054)*	(0.047)	(0.177)	(0.069)
NonBlack women teachers of color	0.044	0.112	0.123	-0.111
	(0.123)	(0.089)	(0.308)	(0.189)
NonBlack men teachers of color	0.365	-0.199	0.761	-0.094
	(0.278)	(0.162)	(0.351)*	(0.338)
Teacher value added	0.070	0.112	0.029	0.060
	(0.018)*	(0.014)*	(0.070)	(0.027)*
One year of teaching experience	0.243	0.180	0.479	0.187
	(0.059)*	(0.052)*	(0.220)*	(0.096)
Two years of teaching experience	0.368	0.316	0.344	0.191
	(0.048)*	(0.044)*	(0.248)	(0.091)*
Three years of teaching experience	0.450	0.403	0.574	0.285
	(0.048)*	(0.049)*	(0.184)*	(0.089)*
Four years of teaching experience	0.545	0.436	0.265	0.418
	(0.070)*	(0.068)*	(0.257)	(0.093)*
Out of state undergrad prepared	-0.001	-0.051	-0.102	-0.079
	(0.052)	(0.045)	(0.234)	(0.095)
Teach for America prepared	0.105	0.338	0.617	0.050
	(0.088)	(0.123)*	(0.410)	(0.143)
Lateral Entry prepared	-0.052	-0.031	0.075	-0.197
	(0.054)	(0.045)	(0.206)	(0.083)*
Other preparation	-0.062	-0.072	-0.152	-0.142
	(0.049)	(0.047)	(0.185)	(0.112)
Black women principals	0.050	-0.017	0.188	0.081
	(0.079)	(0.083)	(0.302)	(0.126)
Black men principals	0.036	0.054	0.126	-0.052
	(0.086)	(0.086)	(0.310)	(0.107)
White men principals	0.157	0.069	0.173	-0.029
	(0.067)*	(0.060)	(0.245)	(0.082)
NonBlack women principals of color	-0.090	-0.328	--	--
	(0.356)	(0.133)*	--	--
NonBlack men principals of color	-0.080	0.006	--	0.108
	(0.492)	(0.223)	--	(0.337)
Principals years of experience	0.005	0.009	-0.004	0.002
	(0.006)	(0.005)	(0.020)	(0.007)
Same race/ethnicity	0.046	0.058	0.423	0.085
	(0.066)	(0.065)	(0.366)	(0.090)
Same gender	0.013	-0.007	0.009	-0.013
	(0.079)	(0.063)	(0.201)	(0.109)
Same race/ethnicity & gender	0.042	0.006	-0.014	-0.005
	(0.084)	(0.072)	(0.294)	(0.117)

Constant	2.930	3.023	2.706	3.232
	(0.091)*	(0.077)*	(0.388)*	(0.139)*
R^2	0.37	0.42	0.64	0.57
Districts	102	103	45	77

Notes: Standard errors are robust to clustering at the school level in parentheses. ELA=English language arts. * indicates significance at the $p<0.05$ level.

Table 1.18 Covariates used in the teacher value-added measurement model

Student Covariates	Classroom Covariates	School Covariates
1) Prior test scores (mathematics and reading) 2) Classmates' prior test scores 3) Days absent 4) Structural mobility 5) Within year mobility 6) Other between year mobility 7) Race or ethnicity 8) Gender 9) Participation in the free or reduced price lunch program, proxy for economic disadvantage 10) Gifted status 11) Disability status 12) Currently receives English as a second language services 13) Previously received English as a second language services 14) Overage for grade 15) Underage for grade 16) Advanced curriculum 17) Remedial curriculum	1) Class size 2) Heterogeneity of prior achievement within the class	1) School size (ADM) 2) School size squared 3) Violent acts per 1,000 students 4) Suspensions per 100 students 5) Total per-pupil expenditures 6) Average district teacher supplement 7) School's racial or ethnic composition 8) School's concentration of economic disadvantage

References: Chapter I

Ambady, N., Rosenthal, R., (1993). Half a minute: predicting teacher effectiveness from thin slices of nonverbal behavior and facial attractiveness. *Journal of Personality and Social Psychology*, 64, 431–441.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology*, 49(1), 141-168.

Balfanz, R. (2009). Putting middle grades students on the graduation path. Retrieved from National Middle School Association website: www.nmsa./org/portals/0/pdf/research/Research_from_the_Field/Policy_Brief_ Balfanz.pdf

Browne I., & Misra, J. (2003). The intersection of gender and race in the labor market. *Annual Review of Sociology*, 29, 487-513.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *Accounting Review*, 86(5), 1549-1575.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education*, *1*(2), 176-216.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, 97(1), 65-108.

Castilla, E. (2008). Gender, race, and meritocracy in organizational careers. *American Journal of Sociology*, 113(6), 1479-1526.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Clotfelter, C. T., Ladd, H.F., & Vigdor, J.L. (2007). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. Working Paper #11. Washington, DC: CALDER.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. University of Chicago Legal Form, 139-167.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241-1299.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of educational research*, *53*(3), 285-328.

Dubrow, J.K. (2008). How can we account for intersectionality in quantitative analysis of survey data? Empirical illustration for Central and Eastern Europe. ASK, 17, 85-100.

Eagly, A.H. & Karau, S.J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573-598.

Ellett, C. D., & Garland, J. S. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to 'state-of-the-art'systems?. *Journal of Personnel Evaluation in Education*, *1*(1), 69-92.

Epstein, J. L. (1985). A question of merit: Principals' and parents' evaluations of teachers. *Educational Researcher*, *14*(7), 3-10.

Gallagher, H. A. (2004). Vaughn elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.

Giddings, P. (1984). When and where I enter: The impact of Black women on race and sex in America. NY: William Murrow.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying Effective Teachers Using Performance on the Job. Discussion Paper 2006-01. *Brookings Institution*.

Grissom, J. A., & Keiser, L. R. (2011). A supervisor like me: Race, representation, and the satisfaction and turnover decisions of public sector employees. *Journal of Policy Analysis and Management*, 30(3), 557-580.

Grissom, J. A., Nicholson-Crotty, J., & Keiser, L. (2012). Does My Boss's Gender Matter? Explaining Job Satisfaction and Employee Turnover in the Public Sector. *Journal of Public Administration Research and Theory*, 22(4), 649-673.

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

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

Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education*, *4*(4), 319-350.

Henry, G.T., Bastian, K.C., Smith, A.A. (2012). Scholarships to Recruit the "Best and Brightest" Into Teaching: Who Is Recruited, Where Do They Teach, How Effective Are They, and How Long Do They Stay? *Educational Researcher*. 41(3), 83-92

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and Leavers Early-Career Teacher Effectiveness and Attrition. *Educational Researcher*, 40(6), 271-280.

Henry, G. T., Fortner, C. K., & Bastian, K. C. (2012). The Effects of Experience and Attrition for Novice High-School Science and Mathematics Teachers. *Science*, 335(6072), 1118-1121.

Henry, G. T., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Purtell, K. M., Thompson, C. L., & Zulli, R. A. (2014a). Teacher preparation policies and their effects on student achievement. *Education Finance and Policy*.

Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014b). The Effects of Teacher Entry Portals on Student Achievement. *Journal of Teacher Education*, *65*(1), 7-23.

Hill, H. C. (2009). Evaluating value-added models: A validity argument approach.*Journal of Policy Analysis and Management*, *28*(4), 700-709.

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education*, 17(3), 207-219.

Jacob, B. A., & Walsh, E. (2011). What's in a rating? *Economics of Education Review*, 30(3), 434-448.

Jacob, B., & Lefgren, L. (2005). Principals as agents: Subjective performance measures in education. NBER Working Paper no. 11463, National Bureau of Economic Research, Cambridge, MA.

_____(2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.

Levit, N. (1998). The gender line: Men, women, and the law. New York: New York University Press.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of management*, *30*(6), 881-905.

Litschge, C. M., Vaughn, M. G., & McCrea, C. (2010). The empirical status of treatments for children and youth

with conduct problems: An overview of meta-analytic studies. Research on Social Work Practice, 20(1), 21- 35.

Manatt, R. P., & Daniels, B. (1990). Relationships between principals' ratings of teacher performance and student achievement. *Journal of Personnel Evaluation in Education*, *4*(2), 189-201.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). Effective supervision: Supporting the art and science of teaching. Heinle & Heinle Pub.

Medley, D.M. & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 80(4), 242-247.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.

Milanowski, A.T, Kimball, S.M., & White, B. (2004). The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites. (CPRE-UW Working Paper Series, TC-04-01). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research, Consortium for Policy Research in Education.

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via Teacher Evaluation The Case of the Missing Clothes?. *Educational Researcher*, *42*(6), 349-354.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects?. *Educational evaluation and policy analysis*, *26*(3), 237-257.

Oppler S.H., Campbell, J.P., Pulakos, E.D., & Borman, W.C. (1992). Three approaches to the investigation of subgroup bias in performance measurement: review, results, and conclusions. *Journal of Political Economic,* 104(5), 201-217.

Peterson, K. (2004). Research on school teacher evaluation. NASSP Bulletin 88(639), 60-79.

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

SBE Highlights - North Carolina Public Schools. (n.d.). Retrieved from http://www.ncpublicschools.org/sbe-archives/highlights/2008/06highlights

Scriven, M. S. (1967). The methodology of evaluation (Perspectives of Curriculum Evaluation, and AERA monograph Series on Curriculum Evaluation, No. 1). *Chicago: Rand NcNally*.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334-1344.

Soar, R., Medley,D. & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan*, 65(4), 239.

Springer, M. G., Pane, J. F., Le, V. N., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. (2012). Team Pay for Performance Experimental Evidence From the Round Rock Pilot Project on Team Incentives. *Educational Evaluation and Policy Analysis*, *34*(4), 367-390.

Stacey, D. C., Holdzkom, D., & Kuligowski, B. (1989). Effectiveness of the North Carolina teacher performance appraisal system. *Journal of personnel evaluation in education*, *3*(1), 79-106.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13(9), 11-18.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review*, *102*(7), 3628-3651.

Weldon, L.S. (2008). Intersectionality in Goertz, G. and Mazur, A. (Eds.). Politics, gender, and concepts: Theory and methodology (193-218). New York: Cambridge University Press.

White, B. (2004). The Relationship Between Teacher Evaluation Scores and Student Achievement: Evidence from Coventry, RI. (CPRE-UW Working Paper Series TC-04-04). Chapter presented at the annual meeting of the American Educational Research Association, San Diego, CA. Madison: University of Wisconsin-Madison, Wisconsin Center Education Research, Consortium for Policy Research in Education.

Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360 feedback® for teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 179-192.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach for America in high school. *Journal of Policy Analysis and Management,30*(3), 447-469.

Yuan, K., Le, V. N., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2013). Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices Results From Three Randomized Studies. *Educational Evaluation and Policy Analysis*, *35*(1), 3-22.

Chapter II

THE MAKE UP TEST: AN EXAMINATION OF THE RELATIONSHIP BETWEEN
CLASSROOM AND SCHOOL CHARACTERISTICS AND TEACHER PERFORMANCE

Since 2009, states have increased their use of standards-based evaluations of teachers to inform

policy and practice to improve teacher quality. In order to accurately inform policy and practice

evaluation systems must be valid and fair. Much of the research on SBEs examines the predictive

validity of evaluation ratings on student test score gains. However, we know little about the

fairness of evaluation systems in education. In this study, I examine whether principals fairly

evaluate teachers in North Carolina's public middle schools, conditional on an objective measure

of teacher effectiveness.

Using data North Carolina administrative data, the results suggest little or no evidence of

systematic bias in the evaluation of middle school teachers based on classroom and school

characteristics. The two most salient findings are that principals give lower ratings to teachers

with larger class sizes and higher ratings to teachers in schools with high total per pupil

expenditures, on average. These findings account for differences in teacher's objective

performance, as measured by their value-added estimates.

The findings provide preliminary evidence that standards-based evaluation systems, at

least as currently implemented in North Carolina, may not completely remove subjective

influences on principals' ratings of teachers. However, the degree of systematic differences may

be considered alarmingly large or comfortingly small.

Introduction

Heightened attention to the quality and effectiveness of teachers has permeated the current generation of education reforms. One such reform involves standards-based evaluations of teachers. Since 2009, states have increased their use of standards-based evaluations of teachers to inform policy and practice to improve teacher quality. However, states vary in structure and purpose of the evaluation system. According to a September 2013 report by the National Center on Teacher Quality, only 11 states and the District of Columbia require statewide implementation of an evaluation system. The remaining states either have no statewide specifications (2 states); states provide a model where districts can opt out (10 states) or districts are given the flexibility to design their own system (27 states) (Doherty & Jacobs, 2013). Additionally, 44 states and the District of Columbia require classroom observations; however, only twenty-seven require the use of student test growth data a measure in the evaluation system (Doherty & Jacobs, 2013). States also use the evaluation system for varying purposes including professional development, dismissal, tenure, and compensation.

In order to accurately inform policy and practice evaluation systems must be valid and fair, regardless of different structures and purposes. Much of the research on standards-based evaluation examines the predictive validity of evaluation ratings on student test score gains. However, we know little about the fairness of evaluation systems in education. In this study, I examine whether principals fairly evaluate teachers in North Carolina's public middle schools, conditional on an objective measure of teacher effectiveness. Fairness is defined as the lack of systematic associations between classroom and school characteristics and the principals' ratings of teacher performance after controlling for an objective measure of teachers' performance. For example, for a standards-based evaluation system to be fair teachers in schools with high concentrations of economically disadvantaged students and/or schools with a high concentration

63

of students of color must receive ratings based on their actual performance. In addition, their ratings should not be systematically different from teachers in schools with few economically disadvantaged students or students of color.

Motivated in part by one of North Carolina's major goals, which is to be "fair to persons being evaluated (SBE, n.d., p.1)," this study focuses on the fairness of North Carolina's evaluation system, by examining whether systematic differences exist in principals' ratings of teachers based on classroom and school working environments. Within schools, teachers and principals attempt to carry out their roles and responsibilities based on individual knowledge, skills and disposition; however, their performance is not devoid of the influence of classroom and school contextual factors. Simply put, context matters in how teachers and principals perform their jobs (Thrupp & Lupton, 2006). The hypotheses for this study rest on the assumption that systematic differences should not exist during the evaluation process of teachers. The research questions the study examines are:

1. Are there systematic differences in teachers' evaluation ratings based on classroom composition after controlling for teacher characteristics, including measures of teachers' value-added?

2. Do any systematic differences that are found in evaluation ratings based on classroom compositional effects remain when examining within school variance only (school-fixed effects)?

3. Are there systematic differences in evaluation ratings based on school composition after controlling for teacher and classroom characteristics?

While there are several limitations of the analysis, which are described later in this essay, the results suggest little or no evidence of systematic differences in the evaluation of teachers based on classroom and school characteristics. In addition, conditional on teacher effectiveness, principals appear to give lower ratings to teachers with larger class sizes. Teachers in schools

64

with high total per pupil expenditures, on average, receive higher ratings. More generally, the findings provide preliminary evidence that standards-based evaluation systems, at least as currently implemented in North Carolina, may not completely remove subjective influences on principals' ratings of teachers. However, the degree of systematic differences are considered comfortingly small. This essay also makes a contribution to the empirical literature on teacher evaluation systems by incorporating objective measures of teachers' effectiveness in the examination of systematic differences in principals' evaluation ratings of teachers.

In the next section, I review the literature on classroom and school characteristics of students and instructional quality in differing school settings. I then discuss the scarce literature on principals, evaluations, and bias, followed by an overview of North Carolina's teacher evaluation system. Next, I describe the methodology including the data and analytic plan. Finally, the results and discussion are presented.

Classroom and School-Level Characteristics of Students

Much of the research on school context focuses primarily on the effect of classroom and school composition on student achievement as measured by standardized test performance. A few studies examined the effects of composition on other measures of student achievement such as student behavior (Pahike, Cooper, & Fabes, 2013; Kelly, 2010) and course taking (Southworth & Mickelson, 2007). Socio-demographic characteristics across this body of research include ethnicity and race (van Ewijk & Sleegers, 2010; Southworth & Mickelson, 2007; Kelly, 2010; Mickelson, Bottia, & Lambert, 2013), socioeconomic or poverty status (van Ewijk & Sleegers, 2010b; Palardy, 2008), ability (Vigdor & Nechyba, 2004; Hanushek, 2003), and gender (Southworth & Mickelson, 2007; Pahike, Cooper, & Fabes, 2013).

Race and Ethnic Composition and Student Achievement

Several studies show that on average predominately White schools benefit students of color by providing access to academic resources (Crain & Mahard, 1983; Wells & Crain, 1994). However, the effect of the racial and ethnic composition of schools on student achievement remains inconclusive. For example, two large meta-regression analyses examining the effect of racial and ethnic school composition on student achievement across a 20-year period found inconsistent results. The first meta-regression analysis employed study-fixed effects and found that large proportions of students of color in a school had a negative effect on the achievement of students from the same racial or ethnic group, but no effect for students' belonging to other racial or ethnic groups (van Ewikj & Sleegers, 2010). In the second meta-regression analysis, exclusively focused on math achievement, attending a racially segregated school had a negative effect on all students' math achievement (Mickelson, Bottia, & Lambert, 2013). Unlike van Ewikj and Sleegers' (2010), this study employed a two-level hierarchical linear model, which accounts for within study variations, but fails to account for unobserved characteristics of the studies.

To date, much of the literature on racial and ethnic composition and student achievement has focused on the Black/White binary, with minimal attention paid to other marginalized ethnic groups, such as Hispanics. However, research on Hispanic students found a positive relationship for Hispanic students attending predominately Hispanic schools and student achievement in reading, math, science and history (Goldsmith, 2003). Conversely, the authors found that, among Black students, there was either no relationship (reading, math and history achievement) or a negative relationship (science achievement) between achievement and attending a predominately Black school, which provides some support for van Ewikj and Sleegers' (2010) meta-regression analysis results.

Additional work extends the school composition research by studying course taking as a measure of student achievement. The research shows a link between the racial composition of students' elementary schools and later enrollment in advanced English courses. More specifically, attending elementary schools that served a large majority of Black students reduced the likelihood of enrolling in college-prep tracks for all students – Black and White (Southworth & Mickelson, 2007). Furthermore, the racial makeup of students' high school also influenced enrollment patterns. With the exception of Black female students, the likelihood of enrolling in college-prep English courses was reduced for Black and White students attending high schools that served a large majority of Black students.

It is possible that enrollment in advanced courses is affected by tracking practices within schools, which create unequal opportunities to rigorous content and learning during earlier grades (Rubin, 2003; Yonezawa, Wells, & Serna, 2002; Oakes, 2005). Researchers reported that ability-grouping practices were used more often in racially balanced schools and schools serving a high proportion of students of color than predominately White schools, at the elementary school level (Buttaro, Catsambis, Mulkey, & Steelman, 2010).

Race, Ethnic, Social Class Composition and Student Achievement
Some scholars argue that social class has the greatest influence on student achievement compared with other factors such as race or ethnicity (Rumberger & Palardy, 2005; Palardy, 2008). Although there is no convincing evidence to support this claim, few would argue that social class or socio-economic status (SES) has no relationship on student achievement. The lack of consensus on the influence of SES composition on achievement is at least a function of the varied measurement and estimating approaches used by researchers (van Ewikj & Sleegers, 2010b).

The research presented to this point limits composition to a single socio-demographic characteristic; however, the intersectionality of race, ethnicity and social class allows for the recognition that students are members of multiple social groups. For example, Black students [or any racialized student] *belong* to the "Black race" as well as a social class. Notwithstanding theoretical, political, and social shifts surrounding race, has led to greater attention on the intersection of race, ethnicity, and social class. Examining the intersection of race and socioeconomic status (as measured by eligibility for free or reduced priced lunch), Southworth (2010) found that when compared to racially and economically balanced schools[4], racially balanced–economically advantaged schools have higher test score performance in reading and math. In contrast, as the percent of students of color increased in economically balanced schools, test performance decreased. Examining science achievement, Hogrebe & Tate (2010) found that schools with large proportions of students of color and economically disadvantaged students had lower test scores and higher dropout rates than comparison schools. However, when taught by highly qualified teachers, students in these schools outperformed their peers.

Although there is general agreement that race, ethnicity, and class composition of classrooms and schools matter for student performance; there is no consensus on which students are influenced most by compositional factors. While these are important factors in understanding student achievement, studies must begin to address a broader set of classroom and school factors to understand mechanisms affecting student achievement. Furthermore, much of the literature on composition place students at the center of the analysis and do not address how compositional structures affect teachers' instructional practices and behaviors.

---

[4] Racially and economically balanced schools were defined as schools with white student populations between 25% and 75% and between 26% and 75% of students eligible for free or reduced priced lunch.

Classroom and School-Level Characteristics and Instructional Quality

Despite the plethora of research on context and student outcomes, there is a paucity of

research on classroom and school context and teachers' instructional practices and quality. For

many teachers, self-efficacy influences the quality of their instruction. However, classroom

composition among other things affects self-efficacy (Stipek, 2012). There is a preponderance of

evidence that self-efficacy impacts teachers' effort (Emmer, 1994), persistence (Bandura, 1997;

Ware & Kitsantas, 2007), instructional practices (Wahlstrom & Louis, 2008; Cousins & Walker,

2000), and student-teacher relationships (Ashton & Webb, 1986). The findings from a few

studies on the composition of students and instructional practices show that instructional

practices and quality differ based on the students assigned to teachers. Much of these differences

are based on students' race, ethnicity, and ability.

Kelly (2010) found that teachers assigned to classrooms with high proportions of Black

students assigned more seatwork, conducted more read-aloud activities, and were less likely to

engage in question and answer with students than their peers. The author also found that teachers

staffed in predominately Black schools struggled with classroom management, reporting higher

levels of student behavioral problems including tardiness, absenteeism, disrespect, and

threatening behaviors than comparison schools.

Teachers' also vary their instructional practices based on the ability composition of the

classes they teach. An ethnographic case study found that students in advanced courses were

exposed to less explicit test preparation instruction, taught more rigorous content, given

challenging writing assignments, and received more immediate verbal and written feedback on

essays than students placed in "regular" courses (Watanabe, 2008).

Evidence reveals that teachers do, in fact, alter their instructional practices and behaviors

based on their perceptions and expectations they hold about students, which affects student

learning (Stipek, 2012; Campbell, 2012; Ferguson, 2003; Tenenbaum & Ruck, 2007; Oakes, 2005). Most of the research in this area focuses on White teachers perceptions of Black students and find that teachers hold more negative perceptions of Black students than White students (Tenenbaum & Ruck, 2007; Oakes, 2005; Ogbu, 2003). Teacher perceptions are a function of the racial or ethnic congruence between teachers and students (Dee, 2005; Farkas, Grobe, Sheehan, & Shuan, 1990; Crosnoe, Johnson, & Elder, 2004). More pointedly, teachers hold more positive attitudes about students with whom they share the same racial background.

A study showed that teachers held more positive expectations, used more positive speech, and made fewer negative referrals for White students than Black students (Tenenbaum & Ruck, 2007). These researchers also report that teachers hold more deficit-oriented beliefs about low-income and Black students than White, Chinese, or middle-class students (Diamond, Randolph, & Spillane, 2004). Additionally, teachers' sense of responsibility for student learning was higher in contexts where teachers perceived students as being exposed to more resources for learning. In other words, teachers felt more responsible for the learning of students from middle-class backgrounds. For low-performing students, teachers felt they were unable to effectively teach these students due to students' lack of motivation, family background, and limited academic skills (Diamond, Randolph, & Spillane, 2004).

In summary, there is strong evidence regarding classroom and school context and instructional quality to suggest that teachers "perform teaching" based on individual, classroom, and school factors. The performance of teaching should contribute to how principals evaluate teachers during classroom observation. Formal evaluations are possibly a stress-inducing event, which may change how teachers perform teaching. Little is known about how school context influences evaluations. To my knowledge, there is only one study, which shows that teachers in

70

high-performing schools receive higher ratings than those in low-performing schools (Jacob &
Walsh, 2011).

Principals, Evaluations, and Bias

Figuratively, principals wear multiple hats within schools, from providing a safe and
orderly environment to efficiently allocating resources. Principals do not manage schools based
on their individual characteristics alone; contextual factors influence their behaviors, as well. For
example, principals with similar knowledge and dispositions related leadership may have
differing outcomes due to differing contextual factors between schools (Robinson, Lloyd, &
Rowe, 2008).

Within schools, principals' role as instructional leaders has understandably received
increased attention. Although the concept of an instructional leader – "a strong, directive
leadership focused on curriculum and instruction from the principal (p. 329)," is over three
decades old, accountability policies have created a resurgence of focus on instructional
leadership in efforts to improve student achievement. As instructional leaders in the evaluation
process, one of the most important outcomes of classroom observations is the detailed feedback
principals provide teachers during the post-observation conference. However, the quality and
perhaps quantity of the feedback is dependent on principals' knowledge on the subject matter. In
examining principals' knowledge of math, Nelson and Sassi (2005) highlight that after
conducting classroom observations, principals who were less knowledgeable about the math
were more likely to provide feedback related to "process oriented classroom practices (p. 28)"
such as student behavior and their motivations to learn. As such, principals' degree of content
knowledge and content specific pedagogy influences the ways in which they interpret of
classroom dynamics.

Research in the private sector provides considerable evidence that evaluators' are motivated by various goals, which consciously or unconsciously encourages behavioral biases in evaluation ratings (Wang, Wong & Kwong, 2010; Golman & Bhatia, 2012; Lewis, 1997). Lewis (1997) highlights four types of evaluator bias that intentionally distorts rating: halo effect, recency effect, central tendency effect, and leniency effect. Each of these forms of bias reduces the usefulness of evaluation ratings and can impact teacher performance. The halo effect occurs when principals' overall impression of a teacher influences his/her evaluation of the teachers' specific behaviors. Recency effect occurs when evaluation ratings are based on the most recent event without considering teacher's performance throughout the year. The central tendency effect occurs when evaluation ratings are compressed across all teachers resulting in less variance in scores. The leniency effect occurs when principals give all teachers high ratings regardless of performance.

Principals may provide biased ratings for various reasons such as conflict avoidance, ensuring fairness, favoritism, nepotism, and discrimination (Wang, Wong & Kwong, 2010; Jacob & Walsh, 2011). The research on the presence, magnitude, and effect of biases in teacher evaluations is scant. Research on the relationships between principal-teacher relationships suggests that principals provide more intangible benefits to racially or ethnically congruent teachers (e.g., principals and teachers from the same racial or ethnic group) (Grissom and Keiser, 2011). In addition, gender congruence influenced working conditions and teacher retention. Specifically, male teachers supervised by a female principal were less satisfied with their jobs and more likely to leave their current school than female teachers (Grissom, Nicholson-Crotty & Keiser, 2012).

Overall, I posit that differences in the implementation of evaluation systems are likely to stifle state and federal goals, which are to provide accurate and fair assessments of teachers' performance that can be used to improve teacher performance. While teacher biases towards students are an important issue and their existence is relevant to the current study, it is outside of the scope of this study. I focus more exclusively on evidence of systematic differences in principal ratings during the evaluation process. The next section briefly describes North Carolina's standards-based teacher evaluation system.

North Carolina Educator Evaluation System

North Carolina's initial teacher evaluation system grew out of education reform efforts of the late 1970s to improve the quality of education (Stacey, Holdzkom, & Kuligowski, 1989; Ellett & Garland, 1987). In 1978, the General Assembly of North Carolina developed a statewide evidence-based evaluation system, known as the Teacher Performance Assessment Instrument (TPAI). All teachers were required to participate in the evaluation process; however, requirements varied based on teachers' license status (i.e. career status (tenured) vs. probationary teachers). The evaluation system was later revised (name changed to TPAI-R) in response to the implementation of the School-Based Management and Accountability Program of 1995 and the Excellent Schools Act of 1997.

Recognizing current challenges to ensure all students have the knowledge and skills to succeed in the 21st century; North Carolina's Department of Public Instruction, in partnership with the Mid-continent Research for Education and Learning, developed a standards-based evaluation system – North Carolina Educator Evaluation System (NCEES) to replace the TPAI-R. The NCEES simultaneously serves dual purposes: originally designed to identify professional

development needs, in 2010 because of the state's Race to the Top application, the system was adapted for making high stakes personnel decisions (e.g., tenure) of teachers.

NCEES was implemented in three phases. The first phase of NCEES occurred during the 2008-2009 school year and included 13 voluntary districts across North Carolina. During the 2009-2010 school year, an additional 39 districts voluntarily participated in phase two of the evaluation process. By the 2010-2011 school year, the final online evaluation system launched and included the remaining 63 districts.

During the NCEES process, teachers receive four classroom observations throughout the school year by a school administrator and peer evaluator. In order to serve as a peer evaluator, a teacher must complete training on the North Carolina Teacher Evaluation Process. All new teachers are required to participate in NCEES until they are granted career status (commonly referred to as "tenure"). Career status teachers fully participate in NCEES (i.e. receive a summative evaluation) during their license renewal year.

The NCEES rating system is based on the North Carolina Professional Teaching Standards created by the North Carolina Professional Teaching Standards Commission.[5] The rubric includes five standards and twenty-five elements that describe the knowledge, skills and dispositions of an effective teacher. The five standards are: Standard 1: Demonstrates leadership; Standard 2: Establishes a respectful environment for a diverse population of students; Standard 3: Knows the content they teach; Standard 4: Facilitates students learning; and Standard 5: Reflects on their practice (SBE, n.d.). Within each standard, teachers are evaluated on a set of elements, which are subdivided into descriptors. Each descriptor contains language that describes the performance responsibilities at each level of the rubric: Distinguished, Accomplished,

---

[5] As of the 2011-12 academic year, North Carolina including a sixth standard based on student achievement growth using univariate response models and multivariate response models to estimate teachers' value-added.

Proficient, Developing, and Not Demonstrated. Evaluators assign ratings for individual

descriptors within an element. The rating for each element is based on the lowest rating received

across all descriptors. As an example, Figure 1 contains an element and descriptors for Standard

1: Teachers demonstrate leadership. As noted, this fictional teacher would receive a rating of

"Developing" on the "Teachers lead in the classroom" element, despite receiving higher ratings

on other descriptors within this element, because "Developing" is the lowest rating in which all

descriptors are observed/marked.

The evaluation process includes seven components: training, orientation, teacher self-

assessment, pre-observation conference, classroom observations, post-observation conferences,

and a summary evaluation conference. At the end of the school year, the principal conducts

summary evaluations to determine teachers' formal rating on each standard. The rating is not a

simple average of observation scores. Principals use multiple evidences including classroom

observations and artifacts (lesson plans, student work, service on committees, etc.) to determine

the final ratings for each standard.

Data and Sample

The data for this study includes administrative data on students, teachers, classrooms, and

schools from the North Carolina Department of Public Instruction. The dataset includes North

Carolina public middle school teachers in their first five years of teaching, in tested subjects,

during the 2010-2011 school year. During the 2010-2011 school year, tested subjects include

English language arts (ELA), math, Algebra I, and science. Teachers are linked to students using

classroom rosters. NCEES evaluation rating data also come from the North Carolina Department

of Public Instruction and are combined with the administrative data using unique teacher

identifiers.

Complete evaluation ratings were available for 45,900 teachers across grade levels and subjects for the 2010-2011 school year. Because I am interested in beginning middle school teachers in tested subjects, the number of teachers rated and included in the analysis sample varied from 139 to 1,004. The sample varies significantly due to testing requirements. For instance, the end-of-grade science test is only taken in eighth grade; therefore, the sample of teachers will be considerably smaller than those in ELA or math models. Across North Carolina, students typically enroll in Algebra I during the ninth grade; however, an increasing number of students take the course during middle grades. Similar to the science teachers, the sample of Algebra I teachers will be considerably smaller than ELA and math teachers.

I focus on beginning teachers for two primary reasons. First, research is clear that teachers improve in their effectiveness over their first five years of teaching (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Clotfelter, Ladd, & Vigdor, 2007, 2010; Harris & Sass, 2011; Henry, Bastian, & Fortner, 2011; Henry, Fortner, & Bastian, 2012). Therefore, more variation in evaluation ratings is likely to exist among beginning teachers. Second, unlike career status teachers, beginning teachers are required to undergo the full NCEES process; therefore, makeup a larger portion of those evaluated on all five standards.

I focus on middle school because it is often an under-researched, despite being a crucial point in students' academic success. Teachers staffed to this grade level must ensure students are prepared for rigorous high school material. Failing to do this increases the risk that students will drop out of school (Bridgeland, Dilulio, & Morison, 2006; Rumberger, 2001). Examining middle schools also allows for a more nuanced examination of evaluation ratings across multiple subject areas that are not possible in self-contained elementary classrooms. [6] Finally, beginning teachers

---

[6] This assumes that ELA, math, Algebra I and science teachers are mutually exclusive, which is not the case in all schools across North Carolina. For example, if a teacher teaches math and Algebra I, of the

in middle grades teach multiple classes and are accountable for the test performance of more students per year than elementary grade teachers. The increased number of classes and students taught provides more statistical power in the measurement models discussed later.

Measures

Dependent Variable: The outcome of interest is teachers' rating on the NCEES. As previously discussed, teachers receive summative ratings on each of the five standards during the end of the school year. The summative ratings are converted into a 5-point scale, where a rating of 1 indicates "Not Demonstrated" and a rating of 5 indicates "Distinguished." For this study, I create a unit-weighted composite rating comprised of the five standards as the dependent variable. I also use the ratings on each standard as separate continuous dependent variables.

Focal Variables. The goal of this study is to examine differences in evaluation ratings based classroom and school characteristics. Therefore, several classroom and school characteristics represent focal variables of interest.

Classroom Characteristics. Teachers assigned to teach, for example, a large proportion of students of color and/or economically disadvantaged students need to be fairly evaluated with their peers. I am not suggesting that teachers assigned to more challenging classroom environments are unaccountable for student learning or classroom management, but principals should not systematically rate these teachers differently than teachers in less challenging classroom environments.

Several key socio-demographic variables of students are included in the analysis. The proportion of Black, White, Asian, Hispanic, Native American, Pacific Islander and Multiracial/ethnic students a given teacher teaches measures classroom racial and ethnic

---

three required observations the principal evaluate the teacher solely on Algebra I performance, vice versa or a combination of the two courses.

composition. Other socio-demographic classroom composition variables include the proportion of male students, the proportion of students eligible for free and reduced priced lunch, the proportion of students that receive or previously received English language learner (ELL) services, and the proportion of students classified as exceptional (i.e., having a disability or classified as academically and intellectually gifted (gifted)).

Classroom contextual variables also affect classroom dynamics and instructional practices. It is clear from the literature that student absenteeism not only impacts student achievement, but also classroom learning environments (Lamdin, 1996). Students with poor attendance have more behavior problems, increased risk of suspension, and higher dropout rates. At the classroom-level, absentee students may create an increased need for classroom management, scaffolding, and differentiated learning all of which can influence teacher evaluations. The average number of students absent during the school year for a given teacher is included in the analysis to account for absenteeism.

School structures and programs may also influence evaluation ratings. School structures such as class size may have an inverse relationship on evaluation ratings. That is, teachers with large classes may receive low evaluation ratings due to the need to differentiate lessons and manage more students. For the analysis, I include the average number of students assigned to each teacher as a focal variable.

Finally, student sorting based on perceived academic ability creates classroom groupings, which may also benefit some teachers over others. The average reading and math peer dispersion are included in the analysis as measures of the heterogeneity of ability within classrooms. Arguably, a teacher with a majority of high-performing students may receive high ratings due to the relative ease in meeting the standards compared to a teacher with majority of low-performing

students or a heterogeneous group of students. Similarly, the proportion of remedial and advanced classes taught by teachers is included as covariates.

School Characteristics. Several key school-level characteristics are included in the analysis to examine whether these factors influence teacher ratings. The racial, ethnic, and economic proportions of schools' student population are included in the analysis. As a measure of school size, average daily membership is included. As a proxy for school resources, I include the total per pupil expenditures across the sample schools. The presence of a safe and orderly learning environment is measured using two variables: 1) the number of acts of crime or violence reported per 1,000 students and 2) the average number of short-term (10 days or less) suspensions per 100 students. Geographic location is also included in the analyses and includes the four major urban-centric locales determined by the National Center for Education Statistics – city, suburb, town and rural. Dummy variables are created for each of the locales, with rural, the modal category, as the reference group. Principals are partitioned into six race-by-gender groups: Black females, Black males, White females, White males, nonBlack females of color, and nonBlack males of color. All groups were dummy coded 1 or 0. White females, the modal group, are the reference category. Finally, a continuous variable of principals' years of experience as a principal is included as a covariate.

Other Covariates. In addition to the focal variables included in the analysis, several teacher-level controls are used in the analysis, to account for factors that may affect the evaluation ratings of teachers. These characteristics include teacher effectiveness, socio-demographic characteristics, experience, and teacher preparation.

Teacher Effectiveness. In an effort to control for an objective measure of teachers' actual effectiveness in the classroom, teacher value-added (TVA) models are also included as a

covariate in the analysis. Teacher value-added modeling is a statistical technique that estimates teachers' impact on student learning after controlling for factors outside of the teacher's control, such as student socio-demographics and school composition. I derive individual teacher value-added estimates for a given subject using a three-level hierarchical linear model (see Henry, Bastian & Fortner, 2011 and Henry, Fortner & Bastian, 2012). Test scores are standardized across grade and subject with a mean of zero and standard deviation of one to account for differences in the tests. In addition to student prior test scores, a rich set of student, classroom, teacher and school-level covariates, shown in Table 2.11, are used to derive teacher value-added estimates. The purpose of this covariate is to test whether systematic differences based on classroom and school characteristics exist in teachers' evaluation ratings after controlling for an objective measure of their performance. For example, if teachers' with higher proportions of students of color are rated lower after controlling for the teachers' value-added scores, this may support the hypothesis that the evaluation ratings are systematically biased. Furthermore, it would suggest that being assigned more students of color might result in lower evaluation ratings, which calls into question the integrity of the evaluation process.

Socio-demographics. Similar to the principal socio-demographic controls, I include six race/ethnicity-by-gender groupings of teachers: Black females, Black males, White females, White males, nonBlack female teachers of color, and nonBlack male teachers of color. NonBlack teachers of color are combined due to sample size limitations and include Hispanic Americans, Asian Americans, Native Americans and unspecified racial and ethnic groups. All groups are dummy coded 1 or 0. White females, the modal cohort, are the reference category.

Experience and Preparation. Teachers' experience is based on the total years of actual teaching and does not include credit from related work experience gained from non-education

industries. Because the sample is limited to beginning teachers, which is defined as teachers in their first five years of teaching, five dichotomous experience variables are created for this analysis. The reference category is teachers with zero years of experience, which means that they are in their first year of teaching.

In recent years, the type of preparation teachers enter the classroom with has become increasingly diverse (Henry et al., 2014a). As a result, beginning teachers may exhibit differences in their content knowledge, skills, dispositions and other important aspects of teaching. These differences may have important implications on teachers' evaluation ratings. For example, in-state public undergraduate prepared teachers might receive high ratings due to their preservice coursework and student teaching directly related to the North Carolina Professional Teaching Standards than out-of-state undergraduate prepared teachers.

Five preparation portals, which prior research indicates affects student achievement (Henry et al., 2014a; Henry et al., 2014b), are included as controls: in-state public undergraduate prepared (reference group), Teach for America (TFA), out-of-state undergraduate prepared, lateral entry, and all other portals. All other portals include in-state and out of state graduate prepared, in-state private undergraduate and graduate programs, visiting international faculty, licensure only, and unclassifiable. These preparation portals are grouped together because they individually represent a small proportion of beginning teachers and the differences in their value-added estimates of effectiveness were relatively small. While teachers who entered the classroom via TFA preparation also represents a small proportion of the beginning teacher workforce, they are examined separately because studies show that TFA teachers are more effective in secondary grades and STEM subjects (Henry et al., 2014a; Xu, Hannaway, & Taylor, 2011).

Analytic Approach

The goal of this study is to examine whether principals rate teachers systematically

different based on classroom and school characteristics, net of teacher performance. To address

the study's goal, a two-step approach that includes a measurement and analysis model are

employed. The measurement model is used to derive individual teacher value-added estimates

for a given subject using a three-level hierarchical linear model (Henry, Bastian & Fortner, 2011

and Henry, Fortner & Bastian, 2012). The hierarchical linear model accounts for the nesting of

students within classrooms, which are nested within schools. The TVA estimates, include a

robust set of covariates such as students' prior test scores and other student, classroom, and

school characteristics to adjust for factors, which are arguably outside of teachers' control, but

affect student achievement gains (see Table 1.18 for a list of the covariates).

Teacher characteristics are omitted from the value-added estimates because of the

possible correlation with student performance and evaluation ratings. For each teacher, TVA

estimates are generated across classrooms and subjects. The reduced form equation for the

estimation of the TVA is:

$$Y_{ijk} = \varphi_{000} + \beta_1 Y_{ijk(t-1)} + \beta_2 X_{ijk} + \gamma_{01} C_{jk} + \varphi_{001} S + \varepsilon_{ijk} + \rho_{0jk} + \phi_k \quad (1)$$

where

Y_{ijk} is the current test performance for student i taught by teacher j in school k;

$Y_{ijk(t-1)}$ is the prior test performance for student i, taught by teacher j in school k;

X_{ijk} is a vector of individual characteristics for student i taught by teacher j in school k;

C_{jk} is a vector of the teacher (classroom)-level characteristics for teacher j in school k;

S_k is a vector of school level characteristics common to all students and teachers in

school k;

ε_{ijk} is the individual error term of student i taught by teacher j in school k;

ρ_{0jk} is the error between teachers within schools for teacher j in school k and yields shrunken empirical Bayes residuals that are used as the measure of teacher effectiveness for the analysis models; and

ϕ_k is the error between schools.

In the analysis models, I employ ordinary least squares (OLS) regression to estimate the relationship between the focal classroom and school variables and evaluation ratings. Cluster-adjusted standard errors are used at the school-level to account for the clustering of teachers within schools. Separate analysis models are conducted for each of the tested subjects – ELA, math, Algebra I, and science.

To address the first research question concerning whether there are systematic differences in teachers' evaluation ratings based on classroom composition after controlling for teacher characteristics, I first estimate a naïve model that includes only the focal classroom-level covariates. The equation used to estimate this model is as follows:

$$\text{ER}_{tk} = \beta_0 + \beta_1 C_{tk} + \varepsilon_{tk} \quad (2)$$

where

ER_{tk} is the composite evaluation rating of teacher t in school k;

C_{tk} is a vector of classroom characteristics of teacher t in school k; and

ε_{tk} is the individual error term of teacher t in school k.

To examine whether the naïve association, presented in Equation 2, weakens after adjusting for teachers' effectiveness, I include the TVA estimates from Equation 1, and adjust for other teacher characteristics. The equation used to estimate the model is as follows:

$$\text{ER}_{tk} = \beta_0 + \beta_1 C_{tk} + \beta_2 \rho_{0jk} + \beta_3 T_{tk} + \epsilon_{tk} \quad (3)$$

where

ER_{tk} is the composite evaluation rating of teacher t in school k;

C_{tk} is a vector of classroom characteristics of teacher t in school k;

ρ_{0jk} is the individual teacher value-added estimates of teacher t in school k;

T_{tk} is a vector of teacher characteristics of teacher t in school k; and

ϵ_{tk} is the individual error term of teacher t in school k.

The variance inflation factor (VIF) was used in each of the models to check for possible multicollinearity. The presence of multicollinearity may make the test of significance overly conservative and create unstable coefficient estimates. The results from the VIF do not indicate the presence of multicollinearity among the covariates in the fully specified model and the mean VIF ranged from 3.32 to 5.08 across the subjects; therefore, all covariates were are included in the fully specified model.

While the use of covariate adjustments attempt to isolate the effect of classroom characteristics on evaluation ratings and reduce omitted variable bias, other sources of bias potentially remain. It is unclear whether differences in teachers' ratings are a result of their sorting into schools. For example, it is plausible to assume based on the literature that teachers staffed in more challenging environments may be lower performing than their peers. Similarly, it is also plausible that teachers assigned more challenging students within schools may be lower performing than their peers. Due to the nonrandom assignment of teachers to schools and classrooms, endogeneity issues are present. Data limitations do not allow me to effectively control for observed and unobserved factors that influence selection into schools and classrooms; therefore, I use school-level fixed effects models to partially address the problem of endogeneity. The school-fixed effects specification accounts for observed and unobserved school by

controlling for all differences between schools, that is only examining differences that occur within schools, thereby also controlling for differences in the way principals rate teachers. It is important to note that the use of school-fixed effects does not eliminate the endogeneity problem; however, it does reduce the bias in the estimates.

The second research question examines systematic differences within schools using school-fixed effects. As previously stated, school-fixed effects account for the presence of nonrandom assignment of teachers to classrooms and unmeasured school factors such as principals' leadership ability. The equation used to estimate the fixed effect model is as follows:

$$\text{ER}_{tk} = \beta_0 + \beta_1 C_{tk} + \beta_2 \rho_{0jk} + \beta_3 T_{tk} + \mu_k + \upsilon_{jk} \quad (4)$$

where

ER_{tk} is the composite evaluation rating of teacher t in school k;

C_{tk} is a vector of classroom characteristics of teacher t in school k;

ρ_{0jk} is the individual teacher value-added estimates of teacher t in school k;

T_{tk} is a vector of teacher characteristics of teacher t in school k;

μ_k is the school-specific error component that is constant for every teacher in schools;

and υ_{jk} is idiosyncratic error that varies across teachers within schools.

The third research question examines the systematic differences in evaluation ratings based on school characteristics, net of teacher and classroom characteristics. The equation is as follows:

$$\text{ER}_{tk} = \beta_0 + \beta_1 C_{tk} + \beta_2 \rho_{0jk} + \beta_3 T_{tk} + \beta_4 S_{tk} + \tau_{tk} \quad (5)$$

where

ER_{tk} is the composite evaluation rating of teacher t in school k;

C_{tk} is a vector of classroom characteristics of teacher t in school k;

ρ_{0jk} is the individual teacher value-added estimates of teacher t in school k;

T_{tk} is a vector of teacher characteristics of teacher t in school k;

S_{tk} is a vector of school characteristics of teacher t in school k; and

τ_{tk} is the individual error term of teacher t in school k.

Results

Descriptive statistics

Table 2.1 shows descriptive statistics separately for the sample of teachers in ELA, math, Algebra I and science. On average, teachers across subjects are rated at least 'Proficient' on the NCEES. In terms of race/ethnicity-gender, the majority of teachers in the models are White women, which is representative of teachers statewide. However, White women are underrepresented in science in comparison to the other subjects. In general, men are heavily concentrated in STEM related subjects. Black women are uniformly represented across subject areas. With the exception of science teachers, teachers' value-added estimates are below average, which is consistent with the evidence about the effectiveness of beginning teachers. The distribution of years of teaching experience was spread out across the five experience variables with the lowest proportion of teachers being those with either one or four years of teaching experience. With the exception of science teachers, in-state public undergraduate programs prepare the majority of teachers in models. Thirty-eight percent of science teachers are prepared via a lateral entry program. In addition, a larger percentage of TFA teachers teach science.

In terms of classroom characteristics, White students make up the largest proportion of students taught by teachers. Between 42% and 50% of the students are eligible for free or reduced priced lunch, with Algebra I teachers with the lowest percentage of eligible students. Across all subjects, students are absent an average of 8 days during the school year. The average class size ranged from 22 to 24 students, which is consistent with state averages for middle

schools. ELA teachers, on average, have a larger percentage of students with disabilities than all other teachers. Not surprising, Algebra I teachers teach large percentage of gifted students. Teachers teach relatively homogenous classrooms, based on math and reading peer ability.

Across the schools in the sample, over half of the student population is eligible for free or reduced priced lunch. The majority of the student population is White, (~ 49%), with Black students making up the second largest racial or ethnic student population at nearly 31 percent. Sample schools are staffed with mostly White principals of which, women make up a slightly larger percentage than men. On average, the principals in the sample have 5 years of experience as a principal in North Carolina public schools.

Findings on Classroom Characteristics

The goal of the first two research questions is to examine whether principals rate teachers systematically different based on classroom characteristics. Results varied across subjects and are presented in Tables 2.2-2.5.

ELA. The first column of Table 2.2 shows that an increase in the average number of males in a classroom statistically decreases ELA teachers' evaluation ratings by 0.41 points. In addition, an increase in the number of gifted students has a positive relationship on ratings by 0.52 points, on average. Conditional on teacher effectiveness and other covariates, the results remained significant although the magnitude of the effects slightly decreased. In both models, the magnitude of the effects are small to moderate. Once school-fixed effects (third column in Table 2) are introduced in the analysis, the results on the number of males and gifted students are no longer significant, which suggests that systematic differences based on these characteristics occur between not within schools.

Math. Turning to math teachers, the results in the first column of Table 2.3 show no significant relationship between evaluation ratings and classroom characteristics. Interestingly,

after controlling for teacher effectiveness and other teacher covariates, the class size has a statistically negative effect on ratings, although the effect is small – 0.012 points. Once school-fixed effects (third column in Table 2.3) are introduced in the analysis, the result on class size disappears. To this point, it appears that classroom characteristics do not influence how principals rate teachers. Surprisingly, after controlling for school characteristics, increasing the proportion of Native American students assigned to math teachers improves their evaluation ratings. Class size has a statistically negative effect on ratings, which is consistent with OLS specification that includes classroom and teacher characteristics. However, similar to the OLS specification the effects are considerably small.

Algebra I. Table 2.4 shows the results for Algebra I teachers. In the analysis without teacher characteristics, the results show that as the number of economically disadvantaged students in classrooms increase, teachers receive higher ratings, which is an unexpected finding. However, adjusting for teacher characteristics the finding is no longer significant. Interestingly, teaching a large proportion of advanced courses has a negative relationship on ratings, net of teacher characteristics. Consistent with math teachers, Algebra I teachers also receive lower ratings as class size increases. There were not enough variation in classroom characteristics among Algebra I teachers to estimate the school-fixed effect model.

Science. In the naïve model, student attendance, class size and proportion of remedial classes a teacher teaches were all negatively related to evaluation ratings (see Table 2.5). After controlling for teacher effectiveness, the effect of student attendance is no longer significant. Increases in the average class size and teaching a large proportion of remedial classes continue to have a negative relationship on ratings. However, these focal variables are not significant in the school-fixed effects model.

88

Findings on School Characteristics

The third research question examines the relationship between school characteristics and teacher evaluation ratings.

ELA. As shown in fourth column on Table 2.2, per pupil expenditures has a positive relationship on ELA teachers' ratings. Additionally, White male principals gave higher ratings than White female principals to ELA teachers, by about 0.10 points. No other school-level variables have a significant relationship on evaluation ratings.

Math. Column 3 on Table 2.3 shows that math teachers' ratings are negatively related to the rate of short-term suspensions in schools. Additionally, there is a positive relationship between total per pupil expenditures and evaluation ratings. Compared to White female principals, White males give higher ratings and nonBlack female principals of color give math teaches lower ratings. Principals' years of leadership experience has a positive relationship on evaluation ratings.

Algebra I. As shown in the last column on Table 2.4, the only significant school-level variable associated with evaluation ratings is total per pupil expenditures. The lack of statistically significant findings may be a function of the lack of power to detect effects due to the limited sample size.

Science. Among science teachers, there is no relationship between the school-level variables and evaluation ratings. The results offer some confidence that principals are not systematically rating teachers based on the characteristics of the schools in which they are staffed.

Additional Specifications

I use two alternative specifications of the OLS regression model shown in Equation 2 to assess the robustness of evaluation rating estimates. The full set of covariates previously

mentioned are used in each of the alternative specification models. First, I used an alternative

measure of the evaluation ratings. The current composite evaluation rating, includes two

standards which are arguably more closely related to teacher value-added estimates, because of

the "direct" effect of these knowledge, skills and dispositions on student learning – Standard 3:

Knows the content and Standard 4: *Facilitates student learning*. The alternative measure of

evaluation ratings is constrained to include only standards two and three.

The results for each subject are shown in Table 2.6. Overall, the results are consistent

with the five-standard composite ratings; however, there are some notable differences. In the

ELA model, the relationship between the proportion of male students and ratings are no longer

significant. There are several inconsistent results in the math model. As shown in the fourth

column on Table 2.3, class size and rate of short-term suspensions in schools are both negatively

related to evaluation ratings. The proportion of Native American students at the classroom level

and total per pupil expenditures are positively related to evaluation ratings. After constraining the

composite ratings to include standards 3 and 4 only, these associates are no longer significant.

The results on principal characteristics remain statistically significant. In the Algebra I model,

total per pupil expenditures is no longer significant. Among science teachers, class size is

negatively associated with evaluation ratings.

Second, the composite rating measure may mask variation in the direction and magnitude

of the evaluation ratings. To account for the variation, each standard serves as a separate

dependent variable. The results are included in Tables 2.7-2.10. The results across subjects

varied substantially. Additionally, none of the significant classroom or school characteristics are

significant across all five individual standards. The most salient finding at the classroom-level is

the negative relationship between evaluation ratings and class size in math and science. In both

90

of these models, there is a negative relationship between class size and Standard 3: *Knows the content.* There is also a negative relationship among math teachers between class size and ratings on standard two – establishes a respectful environment for a diverse population of students. Although the interpretation is speculative, the result may be due to classroom management issues such as dealing with disruptions. The research on class size suggests that smaller class have fewer disruptions and discipline problems and provides more time for individual instruction (Leithwood, Seashore Louis, Anderson & Wahlstrom, 2004; Odden, 1990). Teachers with fewer students are able to facilitate learning for their students (Standard 4), which may be evident during principal observations.

At the school-level, the most salient result is the total per pupil expenditures is positively associated with evaluation ratings. Examining the evaluation ratings by standard and subject, the results show that Standard 1: *demonstrates leadership*, Standard 2: e*stablishes a respectful environment for a diverse population of students*, and Standard 5: *reflects on teaching practices* drives the positive relationship. Interestingly, these standards do not measure instructional practices directly tied to student learning. This may suggest that more resourced schools are able to focus more on teachers' professional growth and development and building relationship building.

Discussion

In response to the increased attention on measuring teacher effectiveness, states are using standards-based evaluation systems because they provide a more comprehensive picture of teaching effectiveness and teaching practices. The few studies on standards-based evaluations, have focused primarily on the predictive validity, with no examinations of the fairness of standards-based evaluations. Given that North Carolina's expects its evaluation system is to be

fair to teachers, it is important that we began critically examining these evaluation systems. In this study, I begin a preliminary exploration into the fairness of North Carolina's evaluation system. More specifically, I examine whether systematic differences exist in teacher evaluation ratings based on classroom and school characteristics.

The current literature on the influence of classroom and school characteristics on teacher evaluation ratings is limited. Prior research does provide us with some evidence that teachers are rated differently based on the academic performance of the schools where they are staffed. At a more micro level, the findings from my study does not indicate a statically significant relationship between student ability at the classroom level and teachers' evaluation ratings. Moreover, this study suggest little or no evidence of systematic differences in the evaluation of teachers based on various classroom and school characteristics, which is promising for the current and future use of the evaluation system.

The most salient school characteristic finding is the positive relationship between total per pupil expenditures and ratings, which is significant in all subject areas except science. Caution must be made when interpreting these findings, given the magnitude of the effects being all less than one point differences. Examining the patterns from the individual standard specification, one possible explanation for the result is that teachers in resourced schools are able to focus more on professional growth and development and building relationships with students. It is unclear how expenditures are allocated within schools. Although studies would suggest that funding geared toward improving student achievement has concentrated on regular classroom instruction, professional development, and instructional support, especially in schools with a large underserved student population (Henry, Fortner, & Thompson, 2010). Future studies

should examine funding in a more nuanced fashion to help understand if and how resources improve instructional practices.

There are two major limitations of this study that should be considered in future studies of teacher evaluation systems. First, only one academic year is used in the analysis; therefore, the conclusions drawn from this study may not appropriately explain the relationship between evaluation ratings and classroom and school characteristics. Future studies should incorporate a longitudinal design to whether the results persist over time or are an artifact of a single time point. Second, this study only examines middle grades. From a policy perspective, expanding this analysis to the elementary and high schools would help in examining the generalizability the findings in this study. For example, in elementary grades, which have, self-contained classrooms, the influence of classroom characteristics may have a greater impact on teachers' instructional practices, because teachers are with the same group of students all day.

Overall, this study begins to unlock the black box regarding how principals evaluate teachers, but more research is warranted to understand whether principals rate teachers systematically differently. Principals may not evaluate teachers based on classroom or school characteristics; however, the controls for teachers' socio-demographic characteristics indicate that systematic bias may occur at the individual teacher-level. Therefore, more research is warranted to understand the non-school factors that may contribute to evaluation ratings.

Figure 2.1 Example of North Carolina Educator Evaluation System Rubric

Standard I: Teachers demonstrate leadership

a. Teachers lead in their classrooms. Teachers demonstrate leadership by taking responsibility for the progress of all students to ensure that they graduate from high school, are globally competitive for work and postsecondary education, and are prepared for life in the 21st century. Teachers communicate this vision to their students. Using a variety of data sources, they organize, plan, and set goals that meet the needs of the individual student and the class. Teachers use various types of assessment data during the school year to evaluate student progress and to make adjustments to the teaching and learning process. They establish a safe, orderly environment, and create a culture that empowers students to collaborate and become lifelong learners.

(Observation)

Developing	Proficient	Accomplished	Distinguished	Not Demonstrated (Comment Required)
	. . . and	. . . and	. . . and	
☑ Understands how they contribute to students graduating from high school.	❑ Takes responsibility for the progress of students to ensure that they graduate from high school.	❑ Communicates to students the vision of being prepared for life in the 21st century.	❑ Encourages students to take responsibility for their own learning.	
☑ Uses data to understand the skills and abilities of students.	☑ Provides evidence of data driven instruction throughout all classroom activities.	☑ Evaluates student progress using a variety of assessment data.	❑ Uses classroom assessment data to inform program planning.	
✓	☑ Establishes a safe and orderly classroom.	❑ Creates a classroom culture that empowers students to collaborate.	☑ Empowers and encourages students to create and maintain a safe and supportive school and community environment.	

Summary Rating Form for Teachers	Developing	Proficient	Accomplished	Distinguished	Not Demonstrated
Standard I: Teachers demonstrate leadership					
A. Leads in the classroom.	X				
B. Leads in the school.		X			
C. Leads the teaching profession.		X			
D. Advocates for the school and students.		X			
E. Demonstrates high ethical standards.		X			
Overall Rating for Standard I		X			

Table 2.1 Summary statistics for middle school teachers, 2010-2011

Variables	ELA Model	Math Model	Algebra I Model	Science Model
Effectiveness				
Teacher Value Added	-0.089 (0.940)	-0.102 (1.027)	-0.083 (0.917)	0.062 (0.906)
NCEES ratings				
Composite ratings	3.337 (0.551)	3.310 (0.555)	3.426 (0.584)	3.348 (0.498)
Std1: Demonstrate leadership	3.306 (0.649)	3.305 (0.652)	3.391 (0.688)	3.359 (0.604)
Std2: Establish positive environment	3.305 (0.614)	3.293 (0.628)	3.391 (0.667)	3.328 (0.591)
Std3: Content knowledge	3.321 (0.614)	3.268 (0.624)	3.464 (0.696)	3.350 (0.597)
Std4: Facilitate learning	3.415 (0.650)	3.349 (0.640)	3.435 (0.639)	3.359 (0.584)
Std5: Reflect on practices	3.336 (0.642)	3.334 (0.659)	3.449 (0.684)	3.343 (0.600)
Teacher socio-demographics				
Male	14.06%	25.41%	25.36%	33.84%
Black teacher	15.48%	15.22%	10.22%	17.43%
White teacher	80.80%	79.92%	83.21%	77.98%
Nonblack teacher of color	3.68%	4.83%	6.52%	4.56%
Black female	12.92%	11.50%	9.42%	12.46%
Black male	2.39%	3.59%	0.72%	4.86%
White female	68.99%	58.93%	60.14%	49.85%
White male	10.83%	20.23%	22.46%	27.66%
Nonblack female teacher of color	2.88%	3.39%	4.35%	3.65%
Nonblack male teacher of color	0.70%	1.44%	2.17%	0.91%
Experience and Preparation				
First year of teaching	22.07%	24.64%	21.01%	20.67%
One year of teaching experience	16.00%	18.40%	15.22%	14.89%
Two years of teaching experience	25.94%	23.52%	18.84%	23.10%
Three years of teaching experience	22.37%	22.23%	31.88%	26.75%
Four years of teaching experience	13.62%	11.14%	13.04%	14.59%
UNC undergraduate prepared	31.21%	35.52%	39.13%	26.14%
Out of state undergraduate	20.78%	20.43%	18.84%	19.76%
Teach For America	3.58%	4.11%	3.62%	6.08%
Lateral entry	22.47%	22.90%	26.09%	37.69%
Other preparation	21.97%	17.04%	12.32%	10.33%
Classroom Characteristics				
Male students	52.28%	52.35%	50.10%	52.41%
Black students	29.08%	29.21%	30.12%	30.15%
White students	51.19%	51.07%	50.09%	50.91%
Hispanic students	12.75%	12.83%	12.31%	11.54%
Asian students	2.29%	2.18%	3.29%	2.25%
Native American students	1.36%	1.31%	0.67%	1.41%
Pacific Islander students	0.10%	0.08%	0.13%	0.08%

Multiracial students	3.23%	3.31%	3.38%	3.67%
Average Days Absent	7.728	7.854	7.539	8.214
Eligible for FRL program	48.54%	49.77%	42.34%	47.41%
Currently receives ELL services	7.21%	6.95%	5.44%	6.05%
Previously received ELL services	4.24%	4.41%	4.42%	3.63%
Students classified with a disability	18.10%	16.67%	8.76%	12.64%
Gifted students	12.51%	11.86%	24.62%	14.11%
Remedial curriculum	8.65%	6.68%	0.42%	2.05%
Advanced curriculum	6.09%	16.37%	50.60%	1.46%
Average Class Size	21.673	22.358	23.868	23.420
Average Math Peer Dispersion	0.686	0.633	0.576	0.743
Average Reading Peer Dispersion	0.713	0.741	0.702	0.801
School Characteristics				
Eligible for FRL program	57.76	58.62%	54.09%	58.44%
Black student population	30.89%	30.38%	32.04%	30.97%
White student population	48.85%	49.17%	47.61%	48.74%
Hispanic student population	12.93%	13.19%	12.69%	13.03%
Asian student population	2.35%	2.32%	3.38%	2.27%
Native American student population	1.33%	1.32%	0.56%	1.42%
Pacific Islander student population	0.09%	0.09%	0.12%	0.10%
Multiethnic or racial student population	3.55%	3.52%	3.59%	3.47%
Average Daily Membership per 100	7.244	7.252	7.916	7.075
Suspension rate per 100 students	33.120	35.238	31.030	39.450
Violent acts per 1,000 students	12.365	12.621	13.258	13.696
Total Per Pupil expenditures	83.394	83.125	82.019	85.787
Urbanicity - City	27.03%	25.15%	34.06%	23.70%
Urbanicity - Suburb	12.03%	10.84%	9.42%	10.03%
Urbanicity - Town	12.92%	14.72%	8.70%	12.16%
Urbanicity - Rural	44.53%	46.42%	47.10%	50.46%
Principal Characteristics				
Male principal	42.47%	46.00%	47.83%	50.61%
Black principal	27.66%	29.36%	29.71%	29.23%
White principal	71.64%	69.81%	70.29%	69.54%
Nonblack principal of color	0.70%	0.82%	0.00%	1.22%
Black female principal	17.15%	16.32%	18.12%	13.11%
Black male principal	10.37%	12.73%	11.59%	15.85%
White female principal	39.88%	36.45%	34.06%	35.67%
White male principal	31.41%	32.65%	36.23%	33.23%
Nonblack female principal of color	0.30%	0.21%	0.00%	0.00%
Nonblack male principal of color	0.40%	0.62%	0.00%	1.22%
Years of experience	5.319	5.137	5.008	5.049
Number of Teachers	1,006	974	138	329

Note: Standard deviations are in parentheses.

Table 2.2 Relationship between teacher evaluation ratings, classroom and school characteristics in ELA

Composite Evaluation Rating	Classroom Characteristics	Classroom & Teacher Characteristics	Classroom & Teacher Characteristics (Fixed Effects)	Classroom, Teacher & School Characteristics
Classroom Characteristics				
Proportion of male students	-0.410	-0.372	-0.260	-0.365
	(0.206)*	(0.175)*	(0.254)	(0.180)*
Proportion of Asian students	-0.706	-0.792	-0.383	-0.833
	(0.555)	(0.511)	(0.856)	(0.635)
Proportion of Black students	-0.172	-0.153	0.255	-0.118
	(0.284)	(0.283)	(0.548)	(0.317)
Proportion of Hispanic students	-0.102	-0.018	0.366	0.162
	(0.433)	(0.415)	(0.726)	(0.495)
Proportion of Multiethnic students	0.143	0.004	-0.412	0.647
	(0.583)	(0.592)	(0.991)	(0.696)
Proportion of Native American students	-0.270	-0.164	-0.257	-0.580
	(0.400)	(0.416)	(1.386)	(0.924)
Proportion of Pacific Islander students	2.347	-0.718	0.891	1.830
	(3.314)	(2.869)	(8.705)	(3.898)
Proportion of White students	0.198	0.239	0.198	0.294
	(0.244)	(0.245)	(0.413)	(0.278)
Average days absent	0.006	0.008	-0.000	0.011
	(0.010)	(0.009)	(0.016)	(0.009)
Average number of students eligible for FRL	0.067	0.081	-0.408	-0.099
	(0.182)	(0.169)	(0.415)	(0.201)
Receives ELL services	0.292	0.167	0.137	0.078
	(0.366)	(0.330)	(0.553)	(0.350)
Previously received ELL services	0.224	0.160	-0.416	0.168
	(0.697)	(0.610)	(0.989)	(0.636)
Proportion of students classified with disabilities	-0.079	-0.068	-0.107	-0.012
	(0.167)	(0.161)	(0.274)	(0.174)
Proportion of gifted students	0.515	0.457	0.261	0.416
	(0.177)*	(0.156)*	(0.242)	(0.174)*
Proportion of remedial courses	0.150	0.087	0.209	0.094
	(0.123)	(0.127)	(0.213)	(0.139)
Proportion of advanced courses	-0.073	-0.056	0.092	-0.063
	(0.129)	(0.122)	(0.190)	(0.131)

Average class size	-0.006	-0.008	-0.000	-0.006
	(0.005)	(0.004)	(0.007)	(0.005)
Average math peer dispersion	0.115	0.156	0.040	0.190
	(0.137)	(0.133)	(0.215)	(0.138)
Average reading peer dispersion	0.125	0.117	-0.024	0.023
	(0.121)	(0.118)	(0.203)	(0.119)
Teacher Characteristics				
		0.071	0.061	0.071
		(0.017)*	(0.026)*	(0.018)*
Black women		-0.005	0.043	0.001
		(0.053)	(0.092)	(0.056)
Black men		-0.051	-0.107	-0.085
		(0.107)	(0.201)	(0.107)
White men		-0.151	-0.119	-0.163
		(0.053)*	(0.065)	(0.054)*
NonBlack women of color		-0.037	0.003	-0.012
		(0.097)	(0.144)	(0.103)
NonBlack men of color		0.473	0.309	0.559
		(0.155)*	(0.516)	(0.203)*
One year of teaching experience		0.250	0.222	0.241
		(0.055)*	(0.089)*	(0.057)*
Two years of teaching experience		0.399	0.358	0.386
Three years of teaching experience		0.493	0.494	0.460
		(0.050)*	(0.075)*	(0.051)*
Four years of teaching experience		0.587	0.514	0.560
		(0.061)*	(0.097)*	(0.064)*
Out of state undergrad prepared		-0.004	0.053	-0.002
		(0.050)	(0.078)	(0.053)
Teach for America prepared		0.103	0.103	0.081
		(0.077)	(0.182)	(0.083)
Lateral entry prepared		-0.005	0.036	-0.026
		(0.048)	(0.076)	(0.049)
Other preparation		-0.040	-0.048	-0.049
		(0.044)	(0.075)	(0.047)
School Characteristics				
Average number of students eligible for FRL				0.000
				(0.002)

Suspension rate per 100 students				-0.000
				(0.001)
Violent acts per 1,000 students				-0.003
				(0.002)
Proportion of Asian students				0.029
				(0.017)
Proportion of Black students				0.024
				(0.013)
Proportion of Hispanic students				0.023
				(0.013)
Proportion of Native American students				0.026
				(0.016)
Proportion of White students				0.022
				(0.013)
Proportion of Pacific Islander students				-0.068
Total per pupil expenditures				0.003
				(0.001)*
City				-0.049
				(0.057)
Suburb				0.090
				(0.073)
Town				0.044
				(0.066)
Average daily membership				0.008
				(0.010)
Black women principals				0.015
				(0.062)
Black men principals				0.020
				(0.073)
White men principals				0.106
				(0.050)*
NonBlack women principals				0.150
				(0.267)
NonBlack men principals				0.120
				(0.464)
Principal's years of experience				0.007
				(0.005)

Constant	3.336	3.010	3.157	0.367
	(0.238)*	(0.242)*	(0.404)*	(1.292)
R^2	0.05	0.22	0.71	0.25
N	1,004	1,004	923	923

Standard errors in parentheses are clustered at the school level. ELA=English language arts. * indicates significance at the $p<0.05$ level.

Table 2.3 Relationship between teacher evaluation ratings, classroom and school characteristics in Math

Composite Evaluation Rating	Classroom Characteristics	Classroom & Teacher Characteristics	Classroom & Teacher Characteristics (Fixed Effects)	Classroom, Teacher & School Characteristics
Classroom Characteristics				
Proportion of male students	-0.174	-0.222	-0.196	-0.171
	(0.230)	(0.190)	(0.330)	(0.200)
Proportion of Asian students	-0.433	-0.226	-0.213	0.448
	(0.683)	(0.636)	(0.983)	(0.838)
Proportion of Black students	-0.078	0.005	-0.273	0.153
	(0.279)	(0.285)	(0.448)	(0.330)
Proportion of Hispanic students	-0.660	-0.343	0.591	0.507
	(0.561)	(0.557)	(0.754)	(0.584)
Proportion of Multiethnic students	-0.611	-0.252	-0.313	0.090
	(0.654)	(0.624)	(0.944)	(0.690)
Proportion of Native American students	-0.023	0.091	0.567	2.167
	(0.447)	(0.411)	(1.853)	(0.959)*
Proportion of Pacific Islander students	-3.000	-3.364	-4.012	-5.176
	(4.411)	(3.996)	(4.514)	(3.636)
	(0.246)	(0.251)	(0.314)	(0.281)
Average days absent	0.006	0.010	-0.009	0.013
	(0.009)	(0.009)	(0.013)	(0.010)
Average number of students eligible for FRL	0.089	0.020	-0.269	-0.348
	(0.186)	(0.179)	(0.362)	(0.240)
Receives ELL services	0.350	0.074	-0.263	-0.005
	(0.489)	(0.475)	(0.502)	(0.460)
Previously received ELL services	1.074	0.802	-0.193	0.943
	(0.655)	(0.650)	(0.847)	(0.693)
Proportion of students classified with disabilities	0.035	-0.008	0.119	-0.030
	(0.192)	(0.174)	(0.221)	(0.166)
Proportion of gifted students	0.325	0.254	0.186	0.169
	(0.170)	(0.158)	(0.233)	(0.169)

Proportion of remedial courses	-0.032	-0.054	0.006	0.029
	(0.146)	(0.139)	(0.170)	(0.143)
Proportion of advanced courses	0.033	-0.012	0.089	-0.010
	(0.085)	(0.079)	(0.123)	(0.084)
Average class size	-0.008	-0.012	0.000	-0.011
	(0.005)	(0.005)*	(0.007)	(0.006)*
Average math peer dispersion	-0.236	-0.080	-0.049	-0.087
	(0.167)	(0.159)	(0.222)	(0.161)
Average reading peer dispersion	0.050	0.107	0.136	0.119
	(0.125)	(0.120)	(0.166)	(0.122)
Teacher Characteristics				
Teacher value-added		0.123	0.095	0.119
		(0.015)*	(0.022)*	(0.016)*
Black women		-0.135	-0.173	-0.164
		(0.055)*	(0.075)*	(0.060)*
Black men		-0.088	-0.083	-0.021
		(0.091)	(0.132)	(0.095)
White men		-0.064	-0.069	-0.086
		(0.043)	(0.062)	(0.045)
NonBlack women of color		0.108	0.018	0.087
		(0.086)	(0.102)	(0.090)
NonBlack men of color		-0.213	-0.492	-0.260
		(0.151)	(0.266)	(0.128)*
One year of teaching experience		0.245	0.304	0.226
		(0.052)*	(0.077)*	(0.054)*
Two years of teaching experience		0.405	0.361	0.391
		(0.049)*	(0.066)*	(0.048)*
Three years of teaching experience		0.462	0.455	0.442
		(0.049)*	(0.070)*	(0.050)*
Four years of teaching experience		0.510	0.428	0.484
		(0.065)*	(0.101)*	(0.070)*
Out of state undergrad prepared		-0.070	-0.015	-0.059
		(0.046)	(0.062)	(0.050)
Teach for America prepared		0.171	0.291	0.182
		(0.088)	(0.194)	(0.118)
Lateral entry prepared		-0.076	-0.051	-0.071
		(0.046)	(0.061)	(0.048)
Other preparation		-0.086	-0.038	-0.095

	(0.046)	(0.066)	(0.048)
School Characteristics			
Average number of students eligible for FRL			0.003
			(0.002)
Suspension rate per 100 students			-0.002
			(0.001)*
Violent acts per 1,000 students			-0.000
			(0.001)
Proportion of Asian students			-0.006
			(0.017)
Proportion of Black students			0.005
			(0.013)
Proportion of Hispanic students			-0.005
			(0.013)
Proportion of Native American students			-0.017
Proportion of White students			0.003
			(0.013)
Proportion of Pacific Islander students			0.035
			(0.106)
Total per pupil expenditures			0.003
			(0.001)*
City			-0.075
			(0.077)
Suburb			0.067
			(0.074)
Town			-0.022
			(0.075)
Average daily membership			0.010
			(0.011)
Black women principals			-0.004
			(0.076)
Black men principals			0.046
			(0.088)
White men principals			0.143
			(0.053)*

NonBlack women principals				-0.534
				(0.193)*
NonBlack men principals				0.114
				(0.193)
Principal's years of experience				0.011
				(0.005)*
Constant	3.538	3.278	3.405	2.440
	(0.265)*	(0.260)*	(0.309)*	(1.299)
R^2	0.04	0.23	0.79	0.27
N	978	978	904	904

Standard errors in parentheses are clustered at the school level.* indicates significance at the p<0.05 level.

Table 2.4 Relationship between evaluation ratings, classroom and school characteristics in Algebra I

Composite Evaluation Rating	Classroom Characteristics	Classroom & Teacher Characteristics	Classroom, Teacher & School Characteristics
Classroom Characteristics			
Proportion of male students	-0.282	-0.494	-0.402
	(0.803)	(0.766)	(1.015)
Proportion of Asian students	-0.851	-1.765	-0.083
	(2.423)	(2.031)	(2.573)
Proportion of Black students	1.087	-0.603	0.933
	(2.341)	(2.065)	(2.578)
Proportion of Hispanic students	-0.626	-1.552	-1.984
	(2.406)	(2.071)	(2.616)
Proportion of Multiethnic students	0.030	-2.308	-1.728
	(2.815)	(2.610)	(3.645)
Proportion of Native American students	-6.384	-6.239	-1.670
	(4.131)	(4.439)	(7.469)
Proportion of Pacific Islander students	-9.245	-3.852	-5.327
	(5.177)	(4.445)	(7.427)
Proportion of White students	1.616	-0.260	-0.224
	(2.351)	(2.080)	(2.235)
Average days absent	-0.028	-0.029	-0.091
	(0.021)	(0.022)	(0.044)*
Average number of students eligible for FRL	1.041	0.658	0.637
	(0.503)*	(0.511)	(0.951)
Receives ELL services	-0.146	-1.446	-1.452
	(1.695)	(1.605)	(2.121)
Previously received ELL services	2.319	2.386	1.860
	(1.698)	(1.624)	(2.069)
Proportion of students classified with disabilities	1.672	1.072	1.360
	(0.925)	(0.993)	(1.076)
Proportion of gifted students	0.765	0.599	0.487
	(0.353)*	(0.332)	(0.536)
Proportion of remedial courses	-1.939	-2.586	-2.728
	(1.336)	(1.575)	(1.885)
Proportion of advanced courses	-0.430	-0.536	-0.528
	(0.197)*	(0.214)*	(0.307)
Average class size	-0.017	-0.027	-0.011
	(0.012)	(0.014)*	(0.019)
Average math peer dispersion	0.430	0.688	0.508
	(0.575)	(0.584)	(0.696)
Average reading peer dispersion	0.109	-0.221	-0.032
	(0.630)	(0.577)	(0.783)
Teacher Characteristics			
Teacher value-added		0.080	0.098
		(0.045)	(0.059)
Black women		0.140	0.086
		(0.180)	(0.201)
Black men		0.896	1.150
		(0.238)*	(0.358)*
White men		-0.050	-0.063
		(0.118)	(0.182)

NonBlack women of color		-0.083	-0.364
		(0.147)	(0.214)
NonBlack men of color		-0.234	-0.282
		(0.353)	(0.351)
One year of teaching experience		0.398	0.315
		(0.147)*	(0.199)
Two years of teaching experience		0.604	0.537
		(0.186)*	(0.241)*
Three years of teaching experience		0.546	0.461
		(0.130)*	(0.150)*
Four years of teaching experience		0.459	0.441
		(0.163)*	(0.258)
Out of state undergrad prepared		0.121	0.144
		(0.163)	(0.185)
Teach for America prepared		-0.089	-0.142
		(0.196)	(0.232)
Lateral entry prepared		-0.033	-0.107
		(0.148)	(0.182)
Other preparation		-0.145	-0.118
		(0.147)	(0.216)
School Characteristics			
Average number of students eligible for FRL			-0.002
			(0.009)
Suspension rate per 100 students			0.002
			(0.003)
Violent acts per 1,000 students			-0.001
			(0.005)
Proportion of Asian students			0.042
			(0.107)
Proportion of Black students			0.042
			(0.105)
Proportion of Hispanic students			0.064
			(0.103)
Proportion of Multiethnic students			0.059
			(0.119)
Proportion of White students			0.059
			(0.106)
Proportion of Pacific Islander students			0.214
			(0.336)
Total per pupil expenditures			0.013
			(0.005)*
City			-0.064
			(0.160)
Suburb			0.108
			(0.238)
Town			-0.028
			(0.245)
Average daily membership			0.005
			(0.037)
Black women principals			-0.163
			(0.250)
Black men principals			0.110
			(0.229)

White men principals			0.183
			(0.146)
Principal's years of experience			-0.021
			(0.018)
Constant	2.265	4.333	-2.367
	(2.622)	(2.305)	(10.518)
R^2	0.24	0.40	0.51
N	138	138	129

Standard errors in parentheses are clustered at the school level.* indicates significance at the $p<0.05$ level.

Table 2.5 Relationship between teacher evaluation ratings, classroom and school characteristics in Science

Composite Evaluation Rating	Classroom Characteristics	Classroom & Teacher Characteristics	Classroom & Teacher Characteristics (Fixed Effects)	Classroom, Teacher & School Characteristics
Classroom Characteristics				
Proportion of male students	0.159	0.157	0.527	-0.089
	(0.287)	(0.301)	(0.488)	(0.359)
Proportion of Asian students	-0.697	-0.911	0.568	-0.543
	(0.903)	(0.853)	(2.721)	(1.340)
Proportion of Black students	0.395	0.275	0.919	0.602
	(0.395)	(0.402)	(1.803)	(0.586)
Proportion of Hispanic students	-0.004	-0.145	5.914	0.196
	(0.850)	(0.871)	(3.647)	(1.046)
Proportion of Multiethnic students	-0.857	-0.893	0.026	-3.081
	(0.949)	(0.921)	(4.402)	(1.445)*
Proportion of Native American students	0.241	0.142	-0.147	-0.677
	(0.817)	(0.810)	(1.573)	(1.535)
Proportion of Pacific Islander students	1.654	2.301	20.550	3.772
	(4.164)	(4.632)	(19.758)	(7.540)
Proportion of White students	0.721	0.561	0.248	0.292
	(0.394)	(0.404)	(1.418)	(0.652)
Average days absent	-0.026	-0.022	-0.016	-0.025
	(0.012)*	(0.012)	(0.033)	(0.017)
Average number of students eligible for FRL	0.133	0.054	-1.405	-0.217
Receives ELL services	-0.491	-0.489	-2.551	-0.330
	(0.959)	(0.910)	(2.695)	(1.096)
Previously received ELL services	0.757	1.175	-2.587	1.843
	(1.200)	(1.220)	(3.528)	(1.330)
Proportion of students classified with disabilities	0.465	0.454	-0.142	0.738
	(0.277)	(0.273)	(1.029)	(0.298)*
Proportion of gifted students	0.403	0.246	-0.273	0.279
	(0.309)	(0.280)	(0.604)	(0.312)
Proportion of remedial courses	-0.621	-0.686	0.221	-0.744
	(0.260)*	(0.256)*	(0.680)	(0.293)*
Proportion of advanced courses	0.491	0.407	0.457	0.587
	(0.336)	(0.328)	(0.952)	(0.360)
Average class size	-0.016	-0.019	-0.018	-0.014

	(0.008)*	(0.007)*	(0.043)	(0.009)
Average math peer dispersion	0.180	0.264	0.248	0.472
	(0.276)	(0.273)	(0.545)	(0.316)
Average reading peer dispersion	0.030	0.120	0.787	-0.028
	(0.263)	(0.256)	(0.802)	(0.305)
Teacher Characteristics				
Teacher value-added		0.066	0.079	0.059
		(0.028)*	(0.069)	(0.032)
Black women		-0.066	0.066	-0.078
		(0.094)	(0.243)	(0.103)
Black men		0.037	0.003	-0.041
		(0.107)	(0.261)	(0.149)
White men		-0.114	-0.150	-0.107
		(0.061)	(0.161)	(0.068)
NonBlack women of color		-0.074	0.074	-0.104
		(0.169)	(0.311)	(0.198)
NonBlack men of color		-0.275	0.342	-0.383
		(0.413)	(0.279)	(0.402)
One year of teaching experience		0.258	-0.232	0.273
		(0.088)*	(0.209)	(0.094)*
		(0.079)*	(0.199)	(0.090)*
Three years of teaching experience		0.420	0.162	0.410
		(0.083)*	(0.240)	(0.091)*
Four years of teaching experience		0.471	0.144	0.498
		(0.104)*	(0.242)	(0.106)*
Out of state undergrad prepared		-0.084	-0.071	-0.109
		(0.086)	(0.205)	(0.097)
Teach for America prepared		-0.082	0.165	-0.084
		(0.105)	(0.320)	(0.131)
Lateral entry prepared		-0.154	-0.114	-0.164
		(0.078)*	(0.159)	(0.091)
Other preparation		-0.130	-0.125	-0.145
		(0.088)	(0.197)	(0.098)
School Characteristics				
Average number of students eligible for FRL				0.002
				(0.003)
Suspension rate per 100 students				0.000

				(0.001)
Violent acts per 1,000 students				-0.003
				(0.002)
Proportion of Asian students				-0.040
				(0.027)
Proportion of Black students				-0.033
				(0.023)
Proportion of Hispanic students				-0.036
				(0.023)
Proportion of Native American students				-0.018
				(0.025)
Proportion of White students				-0.027
				(0.023)
Proportion of Pacific Islander students				-0.016
				(0.162)
Total per pupil expenditures				0.003
				(0.002)
City				0.001
				(0.076)
Suburb				0.125
				(0.108)
Town				-0.118
				(0.124)
Average daily membership				-0.007
				(0.016)
Black women principals				0.040
				(0.101)
Black men principals				0.037
				(0.120)
White men principals				-0.057
				(0.080)
NonBlack men principals				-0.146
				(0.376)
Principal's years of experience				0.011
				(0.007)
Constant	3.111	3.058	2.760	5.832
	(0.441)*	(0.456)*	(1.398)*	(2.162)*
R^2	0.11	0.26	0.98	0.31

| | *N* | 329 | 329 | 307 | 307 |

Standard errors in parentheses are clustered at the school level, * indicates significance at the p<0.05 level.

Table 2.6. Relationship between teacher evaluation ratings, classroom and school characteristics

Composite Rating - Standards 3 & 4	ELA OLS	Math OLS	Algebra I OLS	Science OLS
Classroom Characteristics				
Proportion of male students	-0.167	-0.302	-0.287	-0.343
	(0.215)	(0.211)	(1.013)	(0.392)
Proportion of Asian students	-0.902	0.089	0.544	-0.838
	(0.691)	(0.998)	(2.849)	(1.635)
Proportion of Black students	-0.134	0.140	3.001	0.620
	(0.338)	(0.355)	(2.609)	(0.636)
Proportion of Hispanic students	0.197	0.485	-1.095	0.558
	(0.525)	(0.647)	(3.011)	(1.206)
Proportion of Multiethnic students	0.929	-0.127	0.691	-2.316
	(0.736)	(0.727)	(4.123)	(1.514)
Proportion of Native American students	-0.612	1.832	-4.437	-0.823
	(0.918)	(1.013)	(7.959)	(1.488)
Proportion of Pacific Islander students	3.104	-0.357	2.425	2.888
	(3.831)	(3.993)	(8.442)	(8.382)
Proportion of White students	0.239	0.199	1.283	0.030
	(0.304)	(0.305)	(2.351)	(0.686)
Average days absent	0.010	0.010	-0.107	-0.029
	(0.009)	(0.011)	(0.048)*	(0.018)
Average number of students eligible for FRL	-0.126	-0.403	0.764	-0.230
	(0.207)	(0.251)	(0.963)	(0.460)
Receives ELL services	0.075	-0.030	-0.550	-0.594
	(0.383)	(0.517)	(2.092)	(1.231)
Previously received ELL services	0.223	1.008	2.617	1.492
	(0.646)	(0.732)	(2.240)	(1.499)
Proportion of students classified with disabilities	-0.074	0.016	2.092	0.749
	(0.174)	(0.184)	(1.241)	(0.339)*
Proportion of gifted students	0.437	0.069	0.678	0.401
	(0.174)*	(0.177)	(0.546)	(0.326)
Proportion of remedial courses	0.127	0.024	-3.467	-0.805
	(0.138)	(0.154)	(2.232)	(0.300)*
Proportion of advanced courses	-0.092	0.050	-0.597	0.698
	(0.139)	(0.089)	(0.341)	(0.352)*
Average class size	-0.006	-0.014	-0.010	-0.020
	(0.005)	(0.006)*	(0.019)	(0.010)*
Average math peer dispersion	0.182	0.009	0.815	0.509
	(0.142)	(0.166)	(0.714)	(0.322)
Average reading peer dispersion	-0.044	0.056	-0.499	-0.080
	(0.126)	(0.131)	(0.804)	(0.321)
Teacher Characteristics				
Teacher value-added	0.075	0.115	0.106	0.060
	(0.019)*	(0.016)*	(0.064)	(0.039)
Black women	0.015	-0.148	0.106	-0.080
	(0.060)	(0.066)*	(0.192)	(0.114)
Black men	-0.127	-0.059	1.260	0.032
	(0.098)	(0.103)	(0.363)*	(0.175)
White men	-0.171	-0.083	-0.108	-0.080
	(0.056)*	(0.048)	(0.189)	(0.077)
NonBlack women of color	0.001	0.136	-0.357	-0.080
	(0.119)	(0.099)	(0.372)	(0.216)

NonBlack men of color	0.449	-0.273	-0.269	-0.190
	(0.180)*	(0.122)*	(0.405)	(0.449)
One year of teaching experience	0.261	0.202	0.423	0.253
	(0.059)*	(0.057)*	(0.199)*	(0.108)*
Two years of teaching experience	0.392	0.369	0.549	0.164
	(0.051)*	(0.051)*	(0.232)*	(0.089)
Three years of teaching experience	0.483	0.423	0.324	0.408
	(0.054)*	(0.053)*	(0.161)*	(0.096)*
Four years of teaching experience	0.585	0.458	0.471	0.404
	(0.068)*	(0.071)*	(0.237)*	(0.113)*
Out of state undergrad prepared	-0.025	-0.022	0.249	-0.099
	(0.056)	(0.056)	(0.205)	(0.105)
Teach for America prepared	0.028	0.123	-0.207	-0.170
	(0.089)	(0.117)	(0.320)	(0.150)
Lateral entry prepared	-0.023	-0.046	-0.112	-0.111
	(0.053)	(0.051)	(0.200)	(0.094)
Other preparation	-0.026	-0.060	-0.141	-0.136
	(0.051)	(0.051)	(0.231)	(0.105)
School Characteristics				
Average number of students eligible for FRL	0.000	0.003	-0.001	0.000
	(0.002)	(0.002)	(0.010)	(0.003)
Suspension rate per 100 students	0.000	-0.001	0.002	0.001
	(0.001)	(0.001)	(0.003)	(0.001)
Violent acts per 1,000 students	-0.004	-0.001	0.003	-0.003
	(0.002)	(0.001)	(0.006)	(0.002)
Proportion of Asian students	0.027	-0.004	0.028	-0.026
	(0.019)	(0.019)	(0.115)	(0.025)
Proportion of Black students	0.023	0.002	--	--
	(0.014)	(0.014)	--	--
Proportion of Hispanic students	0.021	-0.006	--	--
	(0.014)	(0.014)	--	--
Proportion of Native American students	0.024	-0.017	--	-0.007
	(0.017)	(0.017)	--	(0.027)
Proportion of White students	0.021	0.001	0.022	-0.016
	(0.014)	(0.014)	(0.116)	(0.025)
Proportion of Pacific Islander students	-0.132	-0.007	0.090	0.032
	(0.118)	(0.116)	(0.333)	(0.162)
Total per pupil expenditures	0.004	0.002	0.009	0.001
	(0.001)*	(0.001)	(0.005)	(0.002)
City	-0.071	-0.032	-0.066	0.069
	(0.057)	(0.083)	(0.184)	(0.090)
Suburb	0.075	0.047	0.026	0.121
	(0.076)	(0.083)	(0.261)	(0.122)
Town	0.050	-0.008	0.137	-0.111
	(0.063)	(0.077)	(0.241)	(0.130)
Average daily membership	0.010	0.010	-0.007	-0.009
	(0.011)	(0.012)	(0.039)	(0.017)
Black women principals	0.049	0.023	-0.159	0.025
	(0.064)	(0.081)	(0.289)	(0.116)
Black men principals	0.031	-0.021	0.179	0.031
	(0.076)	(0.093)	(0.218)	(0.121)
White men principals	0.104	0.131	0.156	-0.037
	(0.051)*	(0.057)*	(0.155)	(0.081)

NonBlack women principals	0.136	-0.548	--	--
	(0.262)	(0.203)*	--	--
NonBlack men principals	0.280	0.159	--	-0.158
	(0.514)	(0.177)	--	(0.421)
Principal's years of experience	0.008	0.014	-0.019	0.009
	(0.006)	(0.006)*	(0.017)	(0.008)
Constant	0.481	2.843	0.293	5.468
	(1.406)	(1.352)*	(11.890)	(2.410)*
R^2	0.24	0.23	0.50	0.26
N	923	904	129	307

Standard errors in parentheses are clustered at the school level.* indicates significance at the p<0.05 level.

Table 2.7 Relationship between teacher evaluation ratings, classroom, school characteristics – ELA by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Classroom Characteristics					
Proportion of male students	-0.494	-0.395	-0.175	-0.158	-0.603
	(0.224)*	(0.233)	(0.254)	(0.239)	(0.220)*
Proportion of Asian students	-0.392	-1.297	-0.863	-0.941	-0.673
	(0.823)	(0.715)	(0.762)	(0.784)	(0.808)
Proportion of Black students	0.012	-0.121	0.066	-0.334	-0.212
	(0.366)	(0.359)	(0.369)	(0.384)	(0.400)
Proportion of Hispanic students	0.381	-0.279	0.172	0.222	0.312
	(0.567)	(0.571)	(0.582)	(0.560)	(0.592)
Proportion of Multiethnic students	0.678	0.606	0.497	1.361	0.095
	(0.889)	(0.855)	(0.777)	(0.865)	(0.826)
Proportion of Native American students	-0.425	-1.087	-1.854	0.630	-0.165
	(1.175)	(1.058)	(1.084)	(1.041)	(1.034)
Proportion of Pacific Islander students	4.394	0.600	1.223	4.985	-2.051
	(5.435)	(4.248)	(3.757)	(4.750)	(4.212)
	(0.309)	(0.326)	(0.343)	(0.327)	(0.321)
Average days absent	0.017	0.004	0.002	0.018	0.012
	(0.010)	(0.010)	(0.009)	(0.011)	(0.010)
Average number of students eligible for FRL	-0.085	0.028	-0.166	-0.086	-0.186
	(0.252)	(0.234)	(0.220)	(0.239)	(0.266)
Receives ELL services	0.024	0.206	0.002	0.148	0.010
	(0.399)	(0.444)	(0.430)	(0.421)	(0.420)
Previously received ELL services	0.038	0.335	-0.018	0.464	0.022
	(0.750)	(0.811)	(0.730)	(0.743)	(0.745)
Proportion of students classified with disabilities	0.084	0.074	-0.157	0.010	-0.068
	(0.204)	(0.202)	(0.189)	(0.193)	(0.202)
Proportion of gifted students	0.489	0.513	0.474	0.401	0.204
	(0.210)*	(0.196)*	(0.182)*	(0.201)*	(0.205)
Proportion of remedial courses	0.071	-0.076	0.144	0.109	0.219
	(0.174)	(0.161)	(0.147)	(0.152)	(0.144)
Proportion of advanced courses	-0.029	-0.056	-0.142	-0.042	-0.048
	(0.150)	(0.146)	(0.146)	(0.154)	(0.164)
Average class size	-0.005	-0.005	-0.003	-0.008	-0.007

	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Average math peer dispersion	0.225	0.239	0.083	0.281	0.120
	(0.171)	(0.162)	(0.154)	(0.169)	(0.167)
Average reading peer dispersion	0.141	-0.051	-0.053	-0.034	0.111
	(0.151)	(0.149)	(0.131)	(0.148)	(0.137)
Teacher Characteristics					
Teacher value-added	0.057	0.069	0.077	0.073	0.079
	(0.021)*	(0.020)*	(0.021)*	(0.021)*	(0.023)*
Black women	-0.048	0.011	0.005	0.026	0.015
	(0.067)	(0.064)	(0.064)	(0.070)	(0.070)
Black men	-0.047	-0.052	-0.129	-0.125	-0.074
	(0.132)	(0.123)	(0.112)	(0.109)	(0.117)
White men	-0.165	-0.179	-0.131	-0.210	-0.131
	(0.064)*	(0.068)*	(0.063)*	(0.063)*	(0.065)*
NonBlack women teachers of color	-0.094	-0.035	0.015	-0.014	0.066
	(0.111)	(0.125)	(0.132)	(0.134)	(0.126)
NonBlack men teachers of color	0.747	0.348	0.478	0.420	0.803
	(0.264)*	(0.162)*	(0.249)	(0.152)*	(0.323)*
	(0.069)*	(0.064)*	(0.059)*	(0.071)*	(0.070)*
Two years of teaching experience	0.374	0.357	0.369	0.416	0.416
	(0.062)*	(0.055)*	(0.058)*	(0.059)*	(0.057)*
Three years of teaching experience	0.450	0.410	0.490	0.476	0.473
	(0.062)*	(0.060)*	(0.060)*	(0.060)*	(0.061)*
Four years of teaching experience	0.496	0.552	0.554	0.617	0.582
	(0.076)*	(0.071)*	(0.074)*	(0.076)*	(0.077)*
Out of state undergrad prepared	-0.004	0.009	-0.007	-0.043	0.036
	(0.063)	(0.059)	(0.061)	(0.061)	(0.065)
Teach for America prepared	0.156	0.042	0.017	0.039	0.149
	(0.108)	(0.096)	(0.098)	(0.119)	(0.119)
Lateral entry prepared	-0.032	-0.006	-0.018	-0.028	-0.047
	(0.060)	(0.056)	(0.057)	(0.058)	(0.056)
Other preparation	-0.067	-0.081	-0.008	-0.043	-0.047
	(0.054)	(0.056)	(0.056)	(0.058)	(0.053)
School Characteristics					
Average number of students eligible for FRL	-0.001	0.001	-0.000	0.000	0.002
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Suspension rate per 100 students	-0.000	0.000	0.000	-0.000	-0.000

	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Violent acts per 1,000 students	-0.001	-0.002	-0.004	-0.004	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Proportion of Asian students	0.035	0.030	0.031	0.023	0.024
	(0.018)	(0.019)	(0.019)	(0.021)	(0.018)
Proportion of Black students	0.032	0.028	0.027	0.019	0.017
	(0.015)*	(0.016)	(0.015)	(0.016)	(0.014)
Proportion of Hispanic students	0.030	0.030	0.028	0.015	0.013
	(0.015)	(0.016)	(0.016)	(0.016)	(0.014)
Proportion of Multiethnic students	0.032	0.035	0.042	0.006	0.014
	(0.019)	(0.019)	(0.019)*	(0.018)	(0.017)
Proportion of White students	0.029	0.028	0.025	0.018	0.012
	(0.014)*	(0.016)	(0.015)	(0.016)	(0.014)
Proportion of Pacific Islander students	-0.031	-0.050	-0.028	-0.236	0.005
	(0.171)	(0.135)	(0.122)	(0.131)	(0.137)
Total per pupil expenditures	0.002	0.003	0.004	0.003	0.003
City	-0.054	-0.032	-0.105	-0.037	-0.018
	(0.069)	(0.067)	(0.061)	(0.065)	(0.069)
Suburb	0.111	0.099	0.088	0.063	0.091
	(0.089)	(0.080)	(0.081)	(0.082)	(0.079)
Town	0.050	0.043	0.025	0.074	0.027
	(0.083)	(0.069)	(0.070)	(0.068)	(0.078)
Average daily membership	0.000	0.011	0.013	0.007	0.011
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Black women principals	-0.085	0.033	0.040	0.058	0.030
	(0.073)	(0.068)	(0.071)	(0.071)	(0.074)
Black men principals	-0.065	0.036	-0.027	0.089	0.067
	(0.091)	(0.078)	(0.080)	(0.085)	(0.075)
White men principals	0.053	0.142	0.076	0.131	0.129
	(0.060)	(0.056)*	(0.053)	(0.056)*	(0.054)*
NonBlack women principals	0.199	0.165	0.215	0.057	0.116
	(0.341)	(0.307)	(0.308)	(0.254)	(0.247)
NonBlack men principals	0.016	-0.255	-0.006	0.567	0.278
	(0.678)	(0.494)	(0.653)	(0.390)	(0.267)
Principal's years of experience	0.001	0.011	0.007	0.008	0.006
	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
Constant	-0.225	-0.097	0.039	0.922	1.195

	(1.431)	(1.551)	(1.478)	(1.562)	(1.382)
R^2	0.17	0.20	0.21	0.22	0.20
N	923	923	923	923	923

Table 2.8 Relationship between teacher evaluation ratings, classroom, school characteristics – Math by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Classroom Characteristics					
Proportion of male students	-0.237	0.017	-0.156	-0.449	-0.030
	(0.244)	(0.241)	(0.217)	(0.291)	(0.248)
Proportion of Asian students	0.878	0.717	0.540	-0.362	0.465
	(0.784)	(1.004)	(1.014)	(1.100)	(0.912)
Proportion of Black students	0.047	0.408	0.165	0.115	0.031
	(0.372)	(0.341)	(0.376)	(0.406)	(0.426)
Proportion of Hispanic students	0.808	0.455	0.418	0.552	0.302
	(0.613)	(0.653)	(0.666)	(0.727)	(0.707)
Proportion of Multiethnic students	0.332	-0.305	-0.196	-0.059	0.679
	(0.931)	(0.796)	(0.803)	(0.851)	(0.784)
Proportion of Native American students	4.081	2.366	2.832	0.832	0.725
	(1.347)*	(1.145)*	(1.249)*	(1.022)	(1.040)
Proportion of Pacific Islander students	-7.107	-6.383	-1.269	0.554	-11.678
	(5.170)	(4.545)	(4.365)	(4.136)	(3.826)*
Proportion of White students	0.170	0.204	0.198	0.200	0.268
	(0.303)	(0.308)	(0.302)	(0.356)	(0.323)
Average days absent	0.016	0.007	-0.005	0.024	0.021
	(0.011)	(0.010)	(0.011)	(0.012)	(0.013)
Average number of students eligible for FRL	-0.082	-0.375	-0.369	-0.438	-0.477
	(0.271)	(0.263)	(0.278)	(0.273)	(0.313)
Receives ELL services	-0.166	0.009	-0.161	0.101	0.191
	(0.460)	(0.491)	(0.531)	(0.571)	(0.555)
Previously received ELL services	0.621	0.600	1.025	0.991	1.477
	(0.757)	(0.786)	(0.748)	(0.851)	(0.806)
Proportion of students classified with disabilities	-0.095	-0.156	-0.154	0.185	0.070
	(0.185)	(0.174)	(0.188)	(0.209)	(0.192)
Proportion of gifted students	0.360	0.184	0.046	0.092	0.162
	(0.193)	(0.191)	(0.194)	(0.203)	(0.201)
Proportion of remedial courses	0.119	-0.055	0.022	0.026	0.035
	(0.158)	(0.154)	(0.152)	(0.187)	(0.175)
Proportion of advanced courses	-0.076	-0.041	0.086	0.014	-0.034
	(0.103)	(0.096)	(0.101)	(0.100)	(0.096)
Average class size	-0.010	-0.016	-0.017	-0.012	-0.003

	(0.006)	(0.006)*	(0.006)*	(0.007)	(0.006)
Average math peer dispersion	-0.017	-0.249	0.080	-0.061	-0.187
	(0.198)	(0.178)	(0.172)	(0.184)	(0.189)
Average reading peer dispersion	0.103	0.242	0.002	0.110	0.138
	(0.148)	(0.149)	(0.138)	(0.153)	(0.163)
Teacher Characteristics					
Teacher value-added	0.131	0.120	0.121	0.109	0.115
	(0.019)*	(0.019)*	(0.018)*	(0.019)*	(0.020)*
Black women	-0.181	-0.182	-0.134	-0.161	-0.160
	(0.072)*	(0.068)*	(0.075)	(0.073)*	(0.074)*
Black men	-0.022	-0.121	-0.165	0.047	0.154
	(0.126)	(0.114)	(0.120)	(0.117)	(0.111)
White men	-0.097	-0.112	-0.122	-0.044	-0.056
	(0.053)	(0.050)*	(0.051)*	(0.056)	(0.053)
NonBlack women of color	-0.041	0.153	0.160	0.112	0.052
	(0.109)	(0.110)	(0.094)	(0.109)	(0.114)
NonBlack men of color	-0.209	-0.265	-0.279	-0.267	-0.281
	(0.123)	(0.150)	(0.110)*	(0.163)	(0.159)
One year of teaching experience	0.219	0.261	0.198	0.207	0.243
	(0.064)*	(0.066)*	(0.064)*	(0.063)*	(0.060)*
Two years of teaching experience	0.373	0.419	0.361	0.377	0.424
	(0.063)*	(0.058)*	(0.056)*	(0.056)*	(0.053)*
Three years of teaching experience	0.410	0.481	0.469	0.376	0.475
	(0.062)*	(0.062)*	(0.061)*	(0.056)*	(0.059)*
Four years of teaching experience	0.511	0.444	0.469	0.446	0.549
	(0.087)*	(0.078)*	(0.075)*	(0.085)*	(0.087)*
Out of state undergrad prepared	-0.073	-0.088	-0.017	-0.026	-0.091
	(0.058)	(0.057)	(0.061)	(0.060)	(0.057)
Teach for America prepared	0.288	0.152	0.041	0.205	0.226
	(0.146)*	(0.118)	(0.116)	(0.147)	(0.165)
Lateral entry prepared	-0.059	-0.072	-0.025	-0.068	-0.129
	(0.062)	(0.053)	(0.056)	(0.057)	(0.059)*
Other preparation	-0.116	-0.076	-0.020	-0.100	-0.161
	(0.055)*	(0.060)	(0.058)	(0.060)	(0.058)*

School Characteristics					
Average number of students eligible for FRL	0.001	0.005	0.003	0.003	0.003
	(0.002)	(0.002)*	(0.002)	(0.002)	(0.003)
Suspension rate per 100 students	-0.002	-0.002	-0.001	-0.001	-0.002
	(0.001)*	(0.001)*	(0.001)	(0.001)	(0.001)*
Violent acts per 1,000 students	-0.000	-0.000	-0.001	-0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Proportion of Asian students	-0.015	-0.001	-0.018	0.011	-0.006
	(0.019)	(0.021)	(0.020)	(0.022)	(0.018)
Proportion of Black students	0.002	0.009	-0.004	0.009	0.009
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
Proportion of Hispanic students	-0.010	0.003	-0.011	-0.001	-0.004
	(0.016)	(0.015)	(0.015)	(0.016)	(0.014)
Proportion of Multiethnic students	-0.040	-0.014	-0.038	0.004	0.001
	(0.020)	(0.019)	(0.020)	(0.018)	(0.018)
Proportion of White students	-0.001	0.011	-0.004	0.007	0.004
	(0.016)	(0.015)	(0.015)	(0.016)	(0.014)
Proportion of Pacific Islander students	-0.009	0.122	-0.029	0.015	0.076
	(0.120)	(0.122)	(0.139)	(0.121)	(0.105)
Total per pupil expenditures	0.004	0.004	0.003	0.002	0.003
	(0.002)*	(0.002)*	(0.002)	(0.002)	(0.002)*
City	-0.072	-0.093	-0.031	-0.033	-0.146
	(0.083)	(0.081)	(0.089)	(0.086)	(0.084)
Suburb	0.135	0.021	0.014	0.081	0.085
	(0.074)	(0.076)	(0.087)	(0.087)	(0.091)
Town	-0.009	-0.034	-0.003	-0.013	-0.052
	(0.087)	(0.078)	(0.079)	(0.084)	(0.081)
Average daily membership	0.011	0.019	0.016	0.003	0.002
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Black women principals	-0.074	-0.012	0.020	0.026	0.020
	(0.082)	(0.085)	(0.082)	(0.089)	(0.082)
Black men principals	0.088	0.059	-0.024	-0.018	0.126
	(0.096)	(0.097)	(0.092)	(0.104)	(0.100)
White men principals	0.186	0.140	0.112	0.150	0.126

	(0.059)*	(0.057)*	(0.061)	(0.059)*	(0.061)*
NonBlack women principals	-0.717	-0.493	-0.497	-0.599	-0.365
	(0.283)*	(0.206)*	(0.278)	(0.177)*	(0.195)
NonBlack men principals	0.128	-0.012	0.413	-0.095	0.139
	(0.332)	(0.238)	(0.236)	(0.191)	(0.168)
Principal's years of experience	0.010	0.012	0.017	0.010	0.007
	(0.006)	(0.005)*	(0.006)*	(0.006)	(0.005)
Constant	2.746	1.629	3.349	2.338	2.136
	(1.606)	(1.509)	(1.454)*	(1.526)	(1.378)
R^2	0.22	0.24	0.22	0.19	0.23

Standard errors in parentheses are clustered at the school level.* indicates significance at the p<0.05 level.

Table 2.9 Relationship between teacher evaluation ratings, classroom, school characteristics – Algebra I by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Classroom Characteristics					
Proportion of male students	-0.903	0.296	-0.517	-0.058	-0.828
	(1.203)	(1.267)	(1.244)	(0.928)	(1.214)
Proportion of Asian students	0.523	-1.685	0.256	0.832	-0.342
	(3.466)	(3.338)	(2.854)	(3.635)	(3.580)
Proportion of Black students	0.204	-0.866	2.073	3.930	-0.678
	(3.381)	(3.226)	(2.784)	(3.151)	(3.200)
Proportion of Hispanic students	0.185	-3.176	-0.251	-1.939	-4.738
	(3.776)	(3.411)	(3.268)	(3.795)	(3.140)
Proportion of Multiethnic students	-1.182	-3.469	2.206	-0.825	-5.369
	(4.286)	(4.681)	(4.550)	(4.740)	(4.841)
Proportion of Native American students	0.615	-1.012	-5.052	-3.821	0.921
	(8.691)	(9.982)	(8.531)	(8.769)	(10.311)
Proportion of Pacific Islander students	-13.309	-3.655	5.324	-0.474	-14.520
	(10.474)	(9.281)	(11.327)	(8.219)	(8.295)
Proportion of White students	-0.061	-1.652	0.339	2.227	-1.972
	(3.134)	(2.781)	(2.431)	(3.122)	(3.019)
Average days absent	-0.083	-0.081	-0.151	-0.062	-0.076
	(0.036)*	(0.050)	(0.045)*	(0.058)	(0.066)
Average number of students eligible for FRL	1.242	0.639	1.125	0.403	-0.225
	(1.088)	(1.100)	(1.164)	(0.984)	(1.142)
Receives ELL services	-3.181	-1.897	-1.995	0.895	-1.084
	(2.359)	(2.411)	(2.439)	(2.213)	(2.858)
Previously received ELL services	0.307	0.079	1.348	3.886	3.681
	(2.450)	(2.534)	(2.525)	(2.364)	(2.757)
Proportion of students classified with disabilities	1.125	0.409	2.631	1.553	1.082
	(1.410)	(1.342)	(1.335)	(1.498)	(1.410)
Proportion of gifted students	0.583	0.402	0.910	0.447	0.091
	(0.672)	(0.603)	(0.633)	(0.621)	(0.653)
Proportion of remedial courses	-2.404	-1.939	-5.538	-1.395	-2.361

	(2.611)	(2.133)	(2.224)*	(3.261)	(2.407)
Proportion of advanced courses	-0.337	-0.648	-0.634	-0.559	-0.462
	(0.348)	(0.371)	(0.367)	(0.364)	(0.359)
Average class size	-0.006	-0.018	-0.008	-0.011	-0.010
	(0.022)	(0.023)	(0.023)	(0.021)	(0.026)
Average math peer dispersion	0.130	0.546	0.371	1.259	0.235
	(0.827)	(0.813)	(0.827)	(0.787)	(0.924)
Average reading peer dispersion	0.685	-0.082	0.071	-1.070	0.238
	(0.954)	(0.811)	(0.881)	(0.854)	(1.034)
Teacher Characteristics					
Teacher value-added	0.047	0.135	0.082	0.130	0.097
	(0.083)	(0.073)	(0.071)	(0.069)	(0.074)
Black women	-0.003	0.090	0.084	0.128	0.129
	(0.245)	(0.246)	(0.199)	(0.206)	(0.277)
Black men	1.118	1.135	1.222	1.298	0.975
	(0.409)*	(0.395)*	(0.384)*	(0.393)*	(0.484)*
White men	-0.103	-0.010	-0.149	-0.066	0.014
	(0.197)	(0.205)	(0.200)	(0.204)	(0.232)
NonBlack women of color	-0.620	-0.022	-0.455	-0.258	-0.463
	(0.305)*	(0.314)	(0.413)	(0.351)	(0.231)*
NonBlack men of color	-0.351	-0.275	-0.370	-0.168	-0.246
	(0.400)	(0.401)	(0.471)	(0.435)	(0.459)
One year of teaching experience	0.285	0.192	0.490	0.355	0.254
	(0.231)	(0.250)	(0.217)*	(0.237)	(0.275)
Two years of teaching experience	0.426	0.605	0.644	0.455	0.555
	(0.270)	(0.270)*	(0.271)*	(0.240)	(0.308)
Three years of teaching experience	0.513	0.505	0.497	0.152	0.637
	(0.191)*	(0.182)*	(0.197)*	(0.165)	(0.178)*
Four years of teaching experience	0.306	0.509	0.514	0.429	0.448
	(0.296)	(0.303)	(0.283)	(0.256)	(0.343)
Out of state undergrad prepared	0.056	0.011	0.233	0.264	0.156
	(0.218)	(0.207)	(0.228)	(0.206)	(0.205)
Teach for America prepared	-0.239	-0.200	-0.382	-0.033	0.143

	(0.399)	(0.259)	(0.390)	(0.316)	(0.304)
Lateral entry prepared	0.050	-0.274	-0.042	-0.182	-0.088
	(0.200)	(0.215)	(0.214)	(0.218)	(0.214)
Other preparation	-0.162	-0.160	-0.109	-0.173	0.015
	(0.242)	(0.257)	(0.272)	(0.248)	(0.258)
School Characteristics					
Average number of students eligible for FRL	-0.006	-0.004	-0.009	0.006	0.004
	(0.011)	(0.011)	(0.012)	(0.010)	(0.011)
Suspension rate per 100 students	0.004	-0.002	0.006	-0.002	0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Violent acts per 1,000 students	-0.005	-0.003	0.001	0.005	-0.001
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
Proportion of Asian students	0.021	0.073	-0.013	0.037	0.091
	(0.127)	(0.132)	(0.131)	(0.142)	(0.155)
Proportion of Black students	0.026	0.084	-0.019	0.015	0.103
	(0.126)	(0.127)	(0.128)	(0.136)	(0.157)
Proportion of Hispanic students	0.031	0.102	0.011	0.046	0.129
	(0.125)	(0.127)	(0.128)	(0.132)	(0.153)
Proportion of Multiethnic students	0.063	0.063	-0.019	0.054	0.136
	(0.141)	(0.139)	(0.148)	(0.156)	(0.174)
Proportion of White students	0.032	0.097	0.006	0.038	0.120
	(0.126)	(0.129)	(0.127)	(0.135)	(0.156)
Proportion of Pacific Islander students	0.291	0.525	0.162	0.018	0.072
	(0.408)	(0.361)	(0.411)	(0.342)	(0.515)
Total per pupil expenditures	0.017	0.019	0.012	0.005	0.010
	(0.006)*	(0.005)*	(0.006)*	(0.006)	(0.007)
City	-0.150	0.109	-0.109	-0.023	-0.148
	(0.180)	(0.191)	(0.220)	(0.188)	(0.200)
Suburb	0.221	0.199	-0.169	0.220	0.068
	(0.247)	(0.254)	(0.298)	(0.280)	(0.310)
Town	-0.307	-0.056	-0.132	0.406	-0.054
	(0.278)	(0.266)	(0.254)	(0.289)	(0.322)
Average daily membership	0.013	0.006	-0.029	0.016	0.019
	(0.046)	(0.044)	(0.046)	(0.044)	(0.046)

Black women principals	-0.101	-0.277	-0.165	-0.154	-0.120
	(0.296)	(0.270)	(0.304)	(0.305)	(0.221)
Black men principals	0.205	-0.101	0.144	0.215	0.089
	(0.267)	(0.258)	(0.254)	(0.232)	(0.335)
White men principals	0.380	0.074	0.170	0.142	0.149
	(0.167)*	(0.169)	(0.179)	(0.168)	(0.188)
Principal's years of experience	-0.035	-0.011	-0.020	-0.019	-0.018
	(0.017)*	(0.020)	(0.022)	(0.017)	(0.027)
Constant	-0.841	-4.995	2.974	-2.388	-6.584
	(13.049)	(13.199)	(12.912)	(14.052)	(15.575)
R^2	0.49	0.48	0.50	0.44	0.42
N	129	129	129	129	129

Standard errors in parentheses are clustered at the school level.* indicates significance at the $p < 0.05$ level.

Table 2.10 Relationship between teacher evaluation ratings, classroom, school characteristics – Science by Standard

	Standard 1 Demonstrate Leadership	Standard 2 Respectful Classroom	Standard 3 Content Knowledge	Standard 4 Facilitate Learning	Standard 5 Reflect on Practice
Classroom Characteristics					
Proportion of male students	-0.075	0.203	-0.299	-0.387	0.113
	(0.382)	(0.438)	(0.450)	(0.441)	(0.438)
Proportion of Asian students	-0.888	-0.215	-0.609	-1.068	0.067
	(1.728)	(1.830)	(2.041)	(1.799)	(1.752)
Proportion of Black students	0.681	0.632	0.332	0.908	0.457
Proportion of Hispanic students	-0.282	-0.019	0.618	0.499	0.165
	(1.161)	(1.226)	(1.284)	(1.390)	(1.281)
Proportion of Multiethnic students	-3.425	-4.534	-2.660	-1.972	-2.812
	(1.980)	(1.661)*	(1.672)	(1.754)	(1.523)
Proportion of Native American students	-1.560	-0.574	-0.884	-0.762	0.395
	(1.662)	(1.977)	(1.736)	(1.582)	(1.843)
Proportion of Pacific Islander students	2.262	4.941	6.416	-0.641	5.880
	(8.102)	(8.923)	(10.219)	(8.501)	(8.638)
Proportion of White students	0.102	0.606	-0.342	0.403	0.690
	(0.758)	(0.768)	(0.772)	(0.829)	(0.861)
Average days absent	-0.022	-0.029	-0.021	-0.037	-0.016
	(0.019)	(0.018)	(0.018)	(0.020)	(0.018)
Average number of students eligible for FRL	-0.035	-0.379	-0.158	-0.301	-0.212
	(0.497)	(0.460)	(0.468)	(0.560)	(0.459)
Receives ELL services	0.437	-0.777	-1.061	-0.127	-0.123
	(1.230)	(1.204)	(1.278)	(1.364)	(1.286)
Previously received ELL services	1.446	3.191	2.812	0.173	1.596
	(1.529)	(1.521)*	(1.685)	(1.619)	(1.503)
Proportion of students classified with disabilities	0.919	0.695	0.569	0.930	0.578
	(0.344)*	(0.360)	(0.381)	(0.383)*	(0.371)
Proportion of gifted students	0.372	0.096	0.428	0.373	0.127
	(0.366)	(0.373)	(0.398)	(0.348)	(0.374)
Proportion of remedial courses	-0.785	-0.683	-0.839	-0.772	-0.639

	(0.363)*	(0.316)*	(0.348)*	(0.415)	(0.443)
Proportion of advanced courses	0.251	0.503	0.897	0.499	0.786
	(0.485)	(0.450)	(0.426)*	(0.336)	(0.553)
Average class size	-0.012	-0.004	-0.022	-0.018	-0.012
	(0.011)	(0.010)	(0.011)*	(0.011)	(0.011)
Average math peer dispersion	1.054	0.218	0.678	0.340	0.072
	(0.358)*	(0.402)	(0.363)	(0.367)	(0.368)
Average reading peer dispersion	-0.325	0.290	-0.172	0.011	0.053
	(0.383)	(0.370)	(0.339)	(0.372)	(0.331)
Teacher Characteristics					
Teacher value-added	0.082	0.054	0.061	0.060	0.037
	(0.042)	(0.036)	(0.045)	(0.041)	(0.037)
Black women	-0.101	-0.101	-0.120	-0.041	-0.026
	(0.135)	(0.117)	(0.132)	(0.123)	(0.132)
Black men	-0.341	-0.143	-0.038	0.102	0.212
	(0.176)	(0.160)	(0.199)	(0.177)	(0.149)
White men	-0.155	-0.148	-0.100	-0.061	-0.072
	(0.083)	(0.076)	(0.088)	(0.086)	(0.083)
NonBlack women of color	-0.324	0.032	0.027	-0.186	-0.068
	(0.167)	(0.254)	(0.218)	(0.286)	(0.243)
NonBlack men of color	-0.455	-0.608	-0.183	-0.198	-0.471
	(0.321)	(0.608)	(0.357)	(0.694)	(0.262)
One year of teaching experience	0.241	0.323	0.360	0.145	0.295
	(0.114)*	(0.111)*	(0.118)*	(0.125)	(0.110)*
Two years of teaching experience	0.242	0.329	0.227	0.101	0.154
	(0.110)*	(0.115)*	(0.104)*	(0.100)	(0.110)
Three years of teaching experience	0.438	0.423	0.449	0.367	0.373
	(0.113)*	(0.104)*	(0.105)*	(0.113)*	(0.109)*
Four years of teaching experience	0.547	0.562	0.431	0.377	0.573
	(0.129)*	(0.126)*	(0.118)*	(0.133)*	(0.134)*
Out of state undergrad prepared	-0.130	-0.155	-0.058	-0.140	-0.061
	(0.118)	(0.115)	(0.115)	(0.114)	(0.107)
Teach for America prepared	-0.021	0.024	-0.046	-0.294	-0.082
	(0.188)	(0.169)	(0.163)	(0.175)	(0.155)
Lateral entry prepared	-0.212	-0.219	0.022	-0.243	-0.170
	(0.112)	(0.111)*	(0.110)	(0.102)*	(0.105)
Other preparation	-0.143	-0.184	-0.117	-0.154	-0.127

	(0.121)	(0.131)	(0.123)	(0.129)	(0.116)
School Characteristics					
Average number of students eligible for FRL	0.002	0.005	-0.000	0.001	0.000
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Suspension rate per 100 students	-0.000	-0.000	0.001	0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Violent acts per 1,000 students	-0.003	-0.001	-0.005	-0.002	-0.005
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)*
Proportion of Asian students	-0.015	-0.075	-0.027	-0.033	-0.048
	(0.034)	(0.034)*	(0.035)	(0.033)	(0.030)
Proportion of Black students	-0.018	-0.055	-0.020	-0.031	-0.043
	(0.029)	(0.027)*	(0.029)	(0.029)	(0.025)
Proportion of Hispanic students	-0.019	-0.059	-0.028	-0.024	-0.052
	(0.029)	(0.029)*	(0.029)	(0.030)	(0.025)*
Proportion of Multiethnic students	0.005	-0.045	-0.000	-0.013	-0.037
	(0.030)	(0.031)	(0.030)	(0.032)	(0.028)
Proportion of White students	-0.010	-0.050	-0.010	-0.023	-0.043
	(0.029)	(0.028)	(0.029)	(0.029)	(0.025)
Proportion of Pacific Islander students	0.022	0.013	-0.037	0.101	-0.176
	(0.215)	(0.222)	(0.185)	(0.200)	(0.196)
Total per pupil expenditures	0.005	0.003	0.001	0.001	0.004
	(0.002)*	(0.002)	(0.002)	(0.002)	(0.002)*
City	-0.003	0.045	0.151	-0.014	-0.176
	(0.094)	(0.093)	(0.105)	(0.098)	(0.092)
Suburb	0.177	0.140	0.201	0.040	0.067
	(0.125)	(0.144)	(0.135)	(0.137)	(0.126)
Town	-0.130	-0.077	-0.137	-0.084	-0.160
	(0.143)	(0.135)	(0.152)	(0.131)	(0.147)
Average daily membership	0.006	0.000	-0.017	-0.002	-0.022
	(0.019)	(0.020)	(0.019)	(0.020)	(0.017)
Black women principals	-0.013	0.092	0.013	0.037	0.072
	(0.123)	(0.121)	(0.129)	(0.124)	(0.116)
Black men principals	-0.047	0.094	0.032	0.030	0.075
	(0.138)	(0.124)	(0.138)	(0.131)	(0.153)
White men principals	-0.112	-0.011	-0.049	-0.025	-0.087

	(0.096)	(0.095)	(0.091)	(0.092)	(0.097)
NonBlack men principals	-0.400	0.089	-0.550	0.234	-0.104
	(0.308)	(0.565)	(0.354)	(0.546)	(0.328)
Principal's years of experience	0.013	0.016	0.008	0.010	0.007
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Constant	3.754	7.271	5.113	5.822	7.201
	(2.816)	(2.579)*	(2.715)	(2.828)*	(2.333)*
R^2	0.29	0.28	0.26	0.21	0.31
N	307	307	307	307	307

Standard errors in parentheses are clustered at the school level.* indicates significance at the $p<0.05$ level.

References: Chapter II

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education*, *1*(2), 176-216.

Bridgeland, J. M., DiIulio Jr, J. J., & Morison, K. B. (2006). The silent epidemic: Perspectives of high school dropouts. *Civic Enterprises*.

Buttaro, A., Catsamis, S., Mulkey, L.M., & Steelman, L.C. (2010). An organizational perspective on the origins of instructional segregation: school composition and use of within class ability grouping in American kindergartens. *Teachers College Record*, 112(5), 1300-1337.

Campbell, S. L. (2012). For Colored Girls? Factors that Influence Teacher Recommendations into Advanced Courses for Black Girls. *The Review of Black Political Economy*, 39(4), 389-402.

Clotfelter, C. T., Ladd, H.F., & Vigdor, J.L. (2007). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. Working Paper #11. Washington, DC: CALDER.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Cousins, J., & Walker, C. (2000). Predictors of educators' valuing of systematic inquiry in schools. *Canadian Journal of Program Evaluation*, 15 (Special Issue), 25–53

Crain, R. L., & Mahard, R. E. (1983). The effect of research methodology on desegregation-achievement studies: A meta-analysis. *American Journal of Sociology*, 839-854.

Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004). Intergenerational bonding in school: The behavioral and contextual correlates of student-teacher relationships. *Sociology of Education*, *77*(1), 60-81.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195-210.

Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology & Education Quarterly*, *35*(1), 75-98.

Doherty, K. M., & Jacobs, S. (2013). State of the states 2013: Connect the dots: Using evaluations of teacher effectiveness to inform policy and practice.

Ellett, C. D., & Garland, J. S. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to 'state-of-the-art'systems?. *Journal of Personnel Evaluation in Education*, *1*(1), 69-92

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review*, 127-142.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the black-white test score gap. *Urban Education*, 38(4), 460-507.

Goldsmith, P. A. (2003). All segregation is not equal: The impact of Latino and Black school composition. *Sociological Perspectives*, *46*(1), 83-105.

Golman, R., & Bhatia, S. (2012). Performance evaluation inflation and compression. *Accounting, Organizations and Society*, 37(8), 534-543.

Grissom, J. A., & Keiser, L. R. (2011). A supervisor like me: Race, representation, and the satisfaction and turnover decisions of public sector employees. *Journal of Policy Analysis and Management*, 30(3), 557-580.

Grissom, J. A., Nicholson-Crotty, J., & Keiser, L. (2012). Does My Boss's Gender Matter? Explaining Job Satisfaction and Employee Turnover in the Public Sector. *Journal of Public Administration Research and Theory*, 22(4), 649-673.

Hanushek, E. A. (Ed.). (2003). The Economics of schooling and school quality (Vol. 159). Edward Elgar Pub.

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

Henry, G. T., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Purtell, K. M., Thompson, C. L., & Zulli, R. A. (2014a). Teacher preparation policies and their effects on student achievement. *Education Finance and Policy*.

Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014b). The Effects of Teacher Entry Portals on Student Achievement. *Journal of Teacher Education*, *65*(1), 7-23.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and Leavers Early-Career Teacher Effectiveness and Attrition. Educational Researcher, 40(6), 271-280.

Henry, G. T., Fortner, C. K., & Bastian, K. C. (2012). The Effects of Experience and Attrition for Novice High-School Science and Mathematics Teachers. *Science,* 335(6072), 1118-1121.

Hogrebe, M. C., & Tate, W. F. (2010). School Composition and Context Factors That Moderate and Predict 10th-Grade. *The Teachers College Record*, 112(4), 1096-1136.

Jacob, B. A., & Walsh, E. (2011). What's in a rating? *Economics of Education Review*, 30(3), 434-448.

Kelly, S. (2010). A Crisis of Authority in Predominantly Black Schools? *The Teachers College Record*, 112(5), 2-3.

Lamdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *The Journal of Educational Research*, 89, 155–162.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How leadership influences student learning.

Lewis, G. B. (1997). Race, sex, and performance ratings in the federal service. *Public Administration Review*, 479-489.

Mickelson, R. A., Bottia, M. C., & Lambert, R. (2013). Effects of School Racial Composition on K–12 Mathematics Outcomes A Metaregression Analysis. *Review of Educational Research*, 83(1), 121-158.

Milanowski, A.T, Kimball, S.M., & White, B. (2004). The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites. (CPRE-UW Working Paper Series, TC-04-01). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research, Consortium for Policy Research in Education.

Nelson, B. S., & Sassi, A. (2005). *The effective principal: Instructional leadership for high-quality learning*. Teachers College Press.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis,* 26, 237-257.

Oakes, J. (2005). Keeping track: How schools structure inequality (2nd ed.). New Haven, CT; London: Yale University Press.

Odden, A. (1990). Class size and student achievement: Research-based policy alternatives. Educational evaluation and policy analysis, 12(2), 213-227.

Ogbu, J. U. (2003). Black American students in an affluent suburb: A study of academic disengagement. Mahwah, N.J.: L. Erlbaum Associates.

Pahlke, E., Cooper, C. E., & Fabes, R. A. (2013). Classroom Sex Composition and First-Grade School Outcomes: The Role of Classroom Behavior. *Social Science Research*, 42(6), 1650-1658

Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19(1), 21-49.

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types.*Educational administration quarterly*, *44*(5), 635-674.

Rubin, B. C. (2003). Unpacking detracking: When progressive pedagogy meets student's social worlds. *American Educational Research Journal*, 40(2), 539-573.

Rumberger, R. W. (2001). Why students drop out of school and what can be done.

Rumberger, R., & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *The Teachers College Record*, 107(9), 1999-2045.

SBE Highlights - North Carolina Public Schools. (n.d.). Retrieved from http://www.ncpublicschools.org/sbe-archives/highlights/2008/06highlights

Southworth, S., & Mickelson, R. A. (2007). The interactive effects of race, gender and school composition on college track placement. *Social Forces*, 86(2), 497-523.

Southworth, S. (2010). Examining the Effects of School Composition on North Carolina Student Achievement Over Time. *Education Policy Analysis Archives*, 18 (29).

Stacey, D. C., Holdzkom, D., & Kuligowski, B. (1989). Effectiveness of the North Carolina teacher performance appraisal system. *Journal of personnel evaluation in education*, *3*(1), 79-106.

Stipek, D. (2012). Context Matters. *The Elementary School Journal*, *112*(4), 590-606.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of educational psychology*, 99(2), 253.

Thrupp, M., & Lupton, R. (2006). Taking school contexts more seriously: The social justice challenge. *British Journal of Educational Studies*, 54(3), 308-328.

van Ewijk, R., & Sleegers, P. (2010). Peer ethnicity and achievement: a meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3), 237-265.

van Ewijk, R., & Sleegers, P. (2010b). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2), 134-150.

Vigdor, J., & Nechyba, T. (2004). Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment. Unpublished manuscript.

Wahlstrom, K., & Louis, K. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Educational Administration Quarterly*, 44, 458–495.

Wang, X. M., Wong, K. F. E., & Kwong, J. Y. (2010). The roles of rater goals and ratee performance levels in the distortion of performance ratings. *Journal of Applied Psychology*, *95*(3), 546.

Ware, H., & Kitsantas, A. (2007). Teacher and collective efficacy beliefs as predictors of professional commitment. *The Journal of Educational Research*, *100*(5), 303-310.

Watanabe, M. (2008). Tracking in the era of high stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record*. 110(3), 489-534.

Wells, A. S., & Crain, R. L. (1994). Perpetuation theory and the long-term effects of school desegregation. *Review of Educational Research*, 64(4), 531-555.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach for America in high school. *Journal of Policy Analysis and Management*, *30*(3), 447-469.

Yonezawa, S., Wells, A. S., & Serna, I. (2002). Choosing tracks: Freedom of choice in detracking schools. *American Educational Research Journal*, 39(1), 37-67.

Chapter III

"WE SPREAD THE WEALTH"
UNDERSTANDING TEACHER STAFFING AND ASSIGNMENT OF BEGINNING
TEACHERS

Despites reform efforts to improve educational outcomes for all students, educational leaders,

policymakers, and researchers continue to highlight disparities among students. Challenged with

federal and state pressures to ensure all students receive equitable educational opportunities,

school leaders are tasked with strategically allocating the most important resource – teachers.

This qualitative study explores how middle school teachers and school leaders make staffing

decisions. I also examine three mechanisms – federal and state policies, institutional politics, and

institutional structures that may influence how school leaders assign teachers to classrooms.

The findings show little evidence that beginning teachers hold specific preferences when

applying for and accepting teaching positions. Teachers seem most interested in simply gaining

employment. Conversely, school leaders do not intentionally seek out beginning teachers in

staffing their schools. In regards to teacher assignment, the study's most salient finding is school

leaders use of a "spread the wealth" philosophy among ELA and math teachers. That is, school

leaders ensure that, within a subject, a given teacher does not exclusively teach high-performing

classrooms nor are they assigned to only low-performing classrooms.

As states and districts continue to create school-wide policies and interventions aimed at

closing test score and opportunity gaps – especially by race, ethnicity, and economic status –

who teaches our children is necessary question that we must continue to examine in hopes that

all students have the opportunity to receive a 'so und basic education.'

Introduction

Today's global economy depends heavily on the availability of knowledgeable and skilled workers. Since the 1980s, education reform efforts in the United States have used accountability driven policies and practices to improve student outcomes. However, research continues to highlight disparities in the educational opportunities among students. For example, students of color and economically disadvantaged students are disproportionately taught by less qualified teachers (Boyd, Lankford, & Wyckoff, 2008; Ingersoll, 2002); more often taught by beginning teachers (Clotfelter, Ladd, & Vigdor, 2005); and more likely tracked in less rigorous courses (Oakes, 1992; Attewell & Domina, 2008; Alvarez & Mehan, 2006) than their White and economically advantaged peers.

Challenged with mounting federal pressures to ensure all students receive equitable educational opportunities, school leaders are tasked with strategically allocating school resources to improve student achievement. There is consensus among scholars, educators and policymakers that teachers are the most important school resource. Evidence suggests that achievement gains for students consistently taught by an ineffective teacher, measured by value-added estimates, are less than those taught by effective teachers (Sanders & Rivers, 1996; Rockoff, 2003). These effects are often strongest among schools with high concentrations of students of color, economically disadvantaged, and low-performing students, which are also disproportionately staffed with ineffective teachers. Nonrandom sorting is likely to produce the inequitable distributions of teachers, which exacerbates achievement and opportunity gaps among students. Within schools, accountability policies place greater weight on student achievement within particular grades and subjects. These policies establish incentives for school

leaders to strategically assign or reassign teachers to these grades and subjects to improve student test score achievement.

Despite decades of teacher quality research, little is known about beginning teachers' career decisions and their subsequent assignments within schools. The extant literature on teachers' career decisions often examines how to retain teachers in the workforce with little empirical research on what factors attract individuals to the teaching profession. Literature related to teacher assignment focuses primarily on elementary and middle schools within Florida, which is a collective bargaining state (Cohen-Vogel & Osborne-Lampkin, 2007; Cohen-Vogel, 2011; Chingos & West, 2011).

This study extends the research by examining beginning teachers staffing and assignment among core subjects in North Carolina public middle schools. As a non-collective bargaining state, studying North Carolina enables us to understand whether the lack of constraints from a bargaining agreement influence how school leaders allocate teachers. Furthermore, the study addresses the staffing and assignment of beginning teachers – teachers in their first two years of teaching, an important time in teachers' careers due to high rates of turnover (Henry, Bastian, & Smith, 2012) and the greening of the teacher workforce (Ingersoll, Merrill, & Stuckey, 2014).

Staffing is defined as the process of hiring teachers into schools. The term teacher assignment refers to school leaders' process of placing teachers in specific grades, subjects and ability-level courses within schools. This differs from student assignment, which refers to the process in which students are assigned to classrooms and teachers (Cohen-Vogel, Osborne-Lampkin, & Houck, 2013). To provide a deeper understanding of staffing and teacher assignment decisions this study examines four research questions:

1. What factors influence where beginning teachers apply and accept teaching positions?

2. What factors influence how school leaders recruit and hire beginning teachers?

3. How does the quantity and quality of the teacher applicant pool affect hiring and assignment decisions of beginning teachers?

4. What roles do federal and state policies; institutional politics; and institutional structures play in teacher assignment practices, especially for beginning teachers?

Overall, these preliminary findings regarding teacher staffing, show little evidence that beginning teachers hold specific preferences when applying for and accepting teaching positions. Teachers seem most interested in simply gaining employment. Regardless of teachers' experience level, school leaders hire teachers based on being a "good fit." In regards to teacher assignment, the study's most salient finding is school leaders use of a "spread the wealth" philosophy in the teacher assignment process among ELA and math teachers. That is, school leaders ensure that, within a subject, a given teacher does not teach high-performing classrooms only nor are they assigned to only low-performing classrooms. Accountability policies are used to reassign and nudge ineffective teachers to transfer to another school. Consistent with prior research, parents and more experienced teachers do not impact assignment decisions; however, school leaders do elicit information regarding all teachers' preferences.

In the next section, I review the literature on teacher staffing across schools and teacher assignment within schools. I then describe the conceptual framework that guides the research questions, followed by a discussion of the data and methods used to address the research questions. Next, I discuss the study's findings and conclude with a discussion of the contributions of this study for the field of education policy.

Literature Review

The assignment of teachers into classrooms is a complex process that varies among districts and schools. To address the complexities of hiring practices and teacher assignment this literature review is divided into two sections: (1) an examination of the literature on hiring and staffing processes between schools and (2) an examination of the literature on teacher assignment within schools.

Hiring and Staffing Processes Between Schools

Staffing of schools is a function of several factors including the supply and demand of teachers. Literature suggests that ineffective teachers, as measured by individual attributes, are staffed in schools with high concentrations of students of color, economically disadvantaged, and low-performing students than effective teachers (Ballou, 1996; Lankford, Loeb, & Wyckoff, 2002; DeAngelis, Presley, & White, 2005). In this section, I discuss the demand side of the teacher staffing process – schools and districts hiring practices, followed by a discussion of teacher preferences and working conditions, which influence the supply of teachers in the workforce.

Hiring Practices. According to Mason & Schroeder (2010), hiring occurs in three stages. The first stage involves pre-screening to reduce the candidate pool to a manageable size and includes an evaluation of a candidate's credentials and letters of recommendation. In the second stage, formal interviews are conducted with eligible candidates. The final stage involves the selection and hiring of teacher candidates. Studies find that when considering which teacher candidates to hire, districts and schools consider professional attributes such as an interview, references, prior academic performance, and personal attributes such as enthusiasm, appearance and communication skills (Mason & Schroeder, 2010). However, the research is inconclusive regarding the importance of any one of these attributes.

Some scholars suggest that institutional prestige, academic major, and academic performance do not increase a teacher candidate's probability of receiving a job offer (Ballou, 1996). Others contend that subject matter certification, academic major, and references are the most important criteria to consider when hiring teachers (Balter & Duncombe, 2005). Using administrative data from the New York City schools, Boyd and colleagues found that schools seek to hire teachers based on institutional prestige, exam scores, experience, teacher value-added scores, race and ethnicity (i.e., Black and Hispanic teachers) (Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff, 2011).

The attributes districts and schools seek in teacher candidates are also a function of external policies such as No Child Left Behind (NCLB) and test-based accountability requirements. In a mixed methods study of the influence of external policies on hiring decisions, Rutledge, Harris, and Ingle (2010) found that in the presence of No Child Left Behind's provision to ensure all students receive a "highly qualified teacher," principals align their hiring preferences based on direct mandates. Principals in the study placed substantial weight on candidates' certifications and credentials. Test-based accountability pressures also influenced hiring decisions made by principals in that they were more likely to give greater weight to teaching skills and content knowledge of candidates. However, this was more prominent among low-performing schools. Rutledge, Harris, and Ingle (2010) suggest that although external policies shape most hiring practices, principals also place emphasis on racial diversity and personality characteristics when hiring candidates.

Other research shows that personal attributes serve as important characteristics, which influence the hiring practices. In a study of first and second year teachers across three states, Johnson and colleagues found that beginning teachers in economically disadvantaged schools

were more often hired without being interviewed than beginning teachers in economically advantaged schools. When interviews were conducted, beginning teachers in economically disadvantaged schools were less likely to interact with future colleagues and other school personnel during the interview process (Johnson, Kardos, Kauffman, Liu, & Donaldson, 2004). Consistent with Johnson et al.'s study, Lui and Johnson-Moore (2006) found that teacher candidates in districts with decentralized hiring practices also experienced limited interactions with school personnel during the interview process.

In addition to the professional and personal attributes, organizational practices have a significant impact on hiring practices. In a study of four underserved urban districts[7], Levin and Quinn (2003) found that inefficiencies in the hiring process at the district-level caused highly qualified teachers, as measured by individual attributes, to accept positions in suburban districts, leaving under-qualified teachers to fill vacancies in underserved urban districts. Late hiring at the district-level occurs for three primary reasons (Levin & Quinn, 2003; Johnson et al., 2004). First, districts are hampered by late or nonexistent notification requirements to fill vacancies for retiring and resigning teachers. Second, seniority-based transfers and the transfer process often delay the hiring of teacher candidates. Finally, student enrollment uncertainties and delayed state budgets prevent districts from accurately determining and budgeting for teacher demand in advance.

Teacher Preferences and Working Conditions. The hiring process does not occur mechanically, whereby educational leaders interview candidates, select desirable candidates, and immediately place them into schools and classrooms. This process unfolds in a negotiated space.

---

[7] Levin and Quinn (2003) describe the districts as "hard-to-staff" urban districts; however, the author's classification was based on the demographics of the student population, which does not predict difficulty in hiring (Opfer, 2011). Therefore, I reclassify these districts as "underserved" because they serve student groups traditionally underserved in U.S. public schools.

Teachers also interview and select schools based on several factors including preferences and working conditions. Much of the research on beginning teachers' workforce decisions is situated within the teacher turnover and retention literature, which bypasses initial decisions. In general, beginning teachers who remain in their initial school of hire, transfer to another school, or leave the teaching profession are influenced by pay, student characteristics, and working conditions (Hanushek, Kain, O'Brien, & Rivkin, 2005; Boyd, Lankford, Loeb, & Wyckoff, 2006; Johnson & Birkeland, 2003). More specifically, lower salaries when compared to other districts and professions, influence who enters and remains in teaching (Dolton & Makepeace, 1993; Clotfelter, Glennie, Ladd, & Vigdor, 2008). Teachers also indicate a preference for assignments in schools that are simultaneously majority White, middle-class and high-performing (Hanushek, Kain, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebrickner, 2007). After considering the level of poverty within schools, Scafidi et al. (2007) found that the racial composition of a school strongly predicted teacher turnover. That is, schools with a high proportion of Black students had higher teacher turnover rates than all other school types.

Teacher working conditions are defined and measured in various ways, with no clear consensus on which aspects matter most in maintaining a positive and supportive environment for teachers. However, there is consensus that these conditions are important in teacher retention, teacher performance, and teacher satisfaction. Poor working conditions such as large course loads, lack of administrative support, and lack of supports for beginning teachers negatively affect retention, performance, and satisfaction (Parker, Goe, Hicks, & McCreadie, 2007). In a study of beginning Teach For America teachers staffed in underserved schools, Donaldson and Johnson (2010) found these teachers were more likely to leave their current schools when given more challenging assignments, such as teaching multiple subjects and out-of-field, than those

with relatively easier assignments. In addition to school structural factors, leadership is shown to predict teacher mobility. In a study of North Carolina's Teacher Working Conditions Survey, Ladd (2011) found that, across all grade levels, higher perceived quality of leadership predicted planned and actual teacher mobility. However, after allowing cross-level variance, Schweig (2013) found that teacher mobility was related to only 'Distributed Leadership' and not all factors of school leadership, as was found in Ladd's study.

To summarize, the literature suggests that school leaders use professional and personal attributes when hiring beginning teachers. Although there is no consensus on which attributes are most salient during the hiring process, school leaders are influenced by accountability policies. Moreover, timing-related challenges constrain the quality and quantity of teachers available to staff schools. Beginning teachers search for jobs based on their own personal preferences and workforce condition preferences. The literature is clear that key preferences are salary, student characteristics and extent to which leadership is distributed across the school.

Teacher Assignment Within Schools

The research on teacher staffing between schools and across districts often overlooks what is arguably considered an important level of sorting – the sorting that goes on within schools (i.e., teacher assignment). Teacher assignment is influenced by several mechanisms. In this section, I examine the literature on federal and state policies, institutional politics, and institutional structures as mechanisms that impact how school leaders make teacher assignment decisions.

Federal and State Policies

Federal and state accountability policies affect hiring decisions and teacher assignments by placing pressures on schools to develop ways to recruit, retain and dismiss teachers. Examining the impact of NCLB on teacher reassignment, Hanushek and Rivkin (2010) found

144

that accountability pressures did not cause principals to move ineffective teachers to non-tested grades or subjects. Conversely, in a qualitative study of high- and low-performing schools in Florida, principals assigned teachers based on student test score data (Cohen-Vogel, 2011). Principals in the study considered a combination of teacher preferences and teacher effectiveness − measured by student test score data − when making assignment decisions by grade and subject area. Chingos and West (2011) also found differences in teacher reassignment based on accountability reforms. Schools with low accountability ratings were more likely to staff teachers in the top effectiveness quartile to tested subjects than schools those with high accountability ratings. Additionally, teachers in the top effectiveness quartile were less likely to be reassigned to non-tested grades.

Institutional Politics

Parental Preferences. Within schools, inequities in teacher assignments may result from demand side constraints school leaders experience from parents (Clotfelter, Ladd, & Vigdor, 2005). Parents who effectively navigate the educational system bargain with school administrators in an effort to ensure their children receive the greatest educational opportunities. The primary request is classroom assignments. Parents, who are characteristically White and from middle- to upper-social class backgrounds, leverage their bargaining power and cultural capital with implicit threats to remove their students from the school if requests are not met (McGrath & Kuriloff, 1999a; Horvat, Weininger, & Lareau, 2003; Wells & Serna, 1996).

In a district-wide study of parental involvement and math track placement, McGrath & Kuriloff (1999b) found that parents were able to successfully intervene to create homogenous classrooms and reassign teachers. Although evidence suggests that some parents are able to influence school structures, a nationally representative study of tracking among secondary public schools found that principals had the greatest influence on teacher assignment. Seventy percent

145

of principals reported that parents had no influence on teacher assignment (Carey, Farris, & Carpenter, 1994).

Teacher Preferences. On the supply side, teachers demonstrate their preferences through negotiations with administrators for specific subjects, grades, and academically rigorous assignments, though this is somewhat constrained by teachers' license qualifications at the secondary level. On average, teachers are less amenable to teaching low-performing students and use their existing power to broker for average or above average-performing classrooms. This teacher leveraging creates positive matching, which occurs when effective teachers are assigned to high-performing students (Clotfelter, Ladd, & Vigdor, 2006). More generally, positive matching is referred to as teacher tracking, which is the systematic "mapping of teachers onto the hierarchically structured course sequences of students" (Kelly, 2004b: 55). Although, teacher tracking does not affect attrition (Kelly, 2004), this practice appears to not only exacerbate student-learning gaps, but also negatively affect teachers' instructional efficacy (Talbert & Ennis, 1990). Positive matching between teachers and students may be more common among experienced teachers given their seniority status (Carey, Farris, & Carpenter, 1994); however, recent studies suggest teacher assignments are influenced predominately by student test score data not seniority (Cohen-Vogel, 2011).

School Leadership Preferences. School leaders also hold preferences in assignment decisions. A number of factors including principals' beliefs about students and teachers, school-wide goals, and political concerns may influence these preferences (Heck, Marcoulides, & Glasman, 1989; Glasman & Heck, 1987). For instance, principals must manage and motivate teachers for retention purposes. Without monetary incentives to do so, principals often use nonpecuniary incentives such as more desirable classrooms or reduced course loads to retain

146

quality teachers. The use of nonpecuniary may be especially present across underserved schools. Consequently, providing nonpecuniary incentives may increase nonrandom assignment of teachers to classrooms and exacerbate teacher tracking practices within schools.

<u>Institutional Structures</u>

The organization of schools is a complex process. School leaders must allocate teachers around academic teams, course offerings and student ability by grade level. Nonempirical literature on academic teams suggests that school leaders must strategically consider the skill set of each teacher assigned to a team (Neubig, 2006). These skill sets include the ability to integrate curriculum across various disciplines and willingness to work well in a team. Limited empirical research has shown that academic teams or professional learning communities improve student achievement (Reed & Groth, 2009; Stearns, 1999) and instructional practices (Martin & Williams, 2012).

Beginning in elementary school, the use of instructional structures such as within-classroom ability groups and advanced courses for gifted students carry over to secondary grades. Across many middle schools, students are frequently assigned to a classroom through "neo tracking," which allows for curricular differentiation within-subject (Mickelson & Everett, 2008). While there is more flexibility in this structure, tracking involves development of courses based on students' perceived and actual academic ability. In addition, the rigor of the content and the nature of instruction differ based on the course (Watanabe, 2008; Oakes, 2005; Mickelson, 2001).   Teachers are assigned to these courses through positive matching (Kelly, 2004b) and experience-level. Faced with parental pressures, school leaders disproportionally assign beginning teachers to lower track courses, which are comprised of low-performing students, based on standardized test performance (Clotfelter, Ladd, & Vigdor, 2005).

Taken together, the literature on teacher assignment practices suggests that little is known about school leaders' decision-making process as it relates to beginning teachers. However, there is some evidence that school leaders use evidence-based staffing and course structures when assigning teachers to classrooms. Finally, there is also evidence to suggest that parental and teacher preferences do not directly influence school leaders' assignment decisions.

Conceptual Framework

Figure 3.1 illustrates the conceptual framework developed based on the literature, which guides the study's research questions. As shown, prior to the staffing of teachers into schools and classrooms, two parallel events occur – teachers' willingness to participate in the teaching workforce and schools' demand for teachers. The supply of teachers is based on many factors; however, for the purpose of this study I focus on teachers' preferences and workforce conditions. Other factors may include salary and competing occupations (Boe & Gilford, 1992). According to the extant literature, geographic location and student composition are two salient personal preferences that contribute to teacher workforce participation (Boyd, Lankford, Loeb, & Wyckoff, 2013). Working conditions such as quality of leadership, camaraderie among teachers, and student related issues (e.g., student behavior) also contribute to workforce participation. Concurrently, schools demand teachers due to vacancies caused primarily by retirements, leaving the profession, transferring to another school/district, and fluctuations in student enrollment.

The concurrent job search by teachers and recruitment efforts by school leaders leads to the hiring process. For teacher candidates the hiring process involves submitting applications, interviewing schools, and selecting a school to be staffed. For school leaders, the hiring process involves pre-screening applications, interviewing candidates, and selecting candidates to fill a teaching position.

Finally, beginning teachers are assigned to classrooms. The assignment is based in part on teachers' license qualifications and the needs of the school. Three major mechanisms school leaders utilize when assigning teachers to classrooms include the federal and/or state policies, institutional politics, and institutional structures. Federal and state policies refer to accountability policies that impact, what Cohen-Vogel (2011) terms, *evidence-based staffing,* in which assignment practices are driven on improving student outcomes. Institutional politics refers to the influence of parents and teachers in teacher assignment. Institutional structures include the ways schools organize academic teams, courses, and students and make teacher assignment decisions based on these factors.

Data and Methods

To understand beginning teachers' initial career decisions and schools' teacher assignment practices, I designed a qualitative case study. A case study design was most appropriate approach in understanding the process of a complex phenomenon – teacher staffing and assignment – within a bounded system (Merriam, 2009; Yin, 2009). The bounded system or case in this study is a single large district in North Carolina. Examining a district provides what Yin (2009) refers to as a representative or typical case of other large districts. It also allows for emergent themes within a district that may be unobserved by large nationally representative samples (Southworth & Mickelson, 2007). However, this case study method has limitations. For instance, a single district may not be representative of other large districts in North Carolina and/or other states. Furthermore, focusing on a single district raises the issue of whether the findings are generalizable to other school settings.

Sample and Site Selection

As Figure 2 illustrates, the sample and site selection was determined using several district and school-level criteria during the 2010-11 school year. The sample district was selected based

on two inclusion criteria. First, the district had a substantial proportion of beginning teachers. Substantial was defined as a district where at least 25 percent of the teachers are in their first two years of teaching. This criterion helped ensure a sizable sample of teachers to recruit into the study. Second, to capture the variation in performance levels of schools, the district had a variation of high-, average- and low-performing schools. This criterion attempts to remove potential selection bias based on student performance at the district level. That is, high-performing districts may exhibit different hiring and placement practices than average - or low-performing districts.

Performance was determined based on schools' performance composite, which is the percentage of student test scores at or above grade level. The criteria used for this study were derived from the five North Carolina ABCs of Public Education categories – Schools of Excellence, Schools of Distinction, Schools of Progress, Priority Schools, and Low-Performing Schools[8]. Low-performing schools were identified as those with a performance composite below 60 percent. This cutoff was based on either 'Priority Schools' or 'Low-Performing Schools' designations. Average-performing districts were identified as those with a performance composite between 60%-79%, which is based on the 'Schools of Progress' designation. High-performing districts were identified as those with a performance composite at or above 80 percent. This cutoff was based on either 'Schools of Excellence' or 'Schools of Distinction' designations. Based on these criteria, a single large district was selected for this study. A proposal requesting to conduct research was sent to the qualifying district for approval. The request was approved, however, principals had to agree to participate and/or allow teachers in their schools to participate in the study.

---

[8] Because there is no district-level designation, I use school-level designations to determine the cutoff for identifying districts based on performance.

After determining the eligible district, I further limit the sample to public middle schools staffed with at least one beginning teacher. Middle schools are defined in this study as schools containing grades six through eight, which is the most common structure of middle grades in North Carolina. Schools with wider middle grade ranges may exhibit systematically different characteristics and assignment practices; therefore, to limit the potential between school confounders, only schools with grades 6-8 were included.

I focus on public middle schools, because demand and supply side constraints at this grade level create an opportunity to examine the hiring process given constraints. For example, nationwide middle schools are currently experiencing teacher shortages, especially in math and science (U.S. DOE, 2011). Additionally, on the supply side, middle schools received fewer candidate applications and more transfer requests compared to elementary schools (Boyd et al., 2011). It remains unclear why teachers favor elementary and high schools; however, some suggest that students' age and curriculum content were top reasons teachers preferred elementary and high school grades (Radcliffe & Mandeville, 2007). Teachers and administrators from high- and average-performing schools were included in the sample. Low-performing schools were excluded from the study, because I was unable to recruit any of these middle schools to participate[9]. School performance levels are defined using similar cutoff levels from the district inclusion requirements previously discussed.

For the purpose of this study, beginning teachers are defined as teachers in their first year of teaching in North Carolina public schools. Schools with at least one beginning teacher were eligible to participate in the study. Furthermore, only beginning teachers teaching core subjects

---

[9] Several efforts were made to recruit low-performing schools including, multiple email and telephone correspondences as well as school visits; however, principals were unavailable and/or unresponsive. One principal from a low-performing school declined participation. Two other school leaders agreed to participate, but were not in the study, because upon scheduling an interview time, one decided not to participate, the other did not respond to email correspondences or voice messages.

(i.e., math, English language arts, social studies, and science) are included in the analysis, to capture the influence of high-stakes accountability on hiring practices and teacher assignment practices. Social studies and sixth and seventh grade science are non-tested subjects; therefore, teachers in these grades/subjects serve as a comparison group with teachers staffed to tested subjects. Of the 82 middle school beginning teachers in the district, 24 were eligible to participate in the study. Finally, principals and assistant principals (APs) in the eligible schools were recruited to participate in the study.

Based on these exclusions, interviews were conducted within four schools through a convenience sampling of beginning teachers, principals and/or assistant principals. In all, 10 participants were interviewed, including one principal, two assistant principals and seven teachers. Two of the schools are classified as average-performing and two are classified as high-performing. Unintentionally, the high-performing schools are both located in rural areas, whereas, the average-performing schools are located in the city. Among the teachers, four were prepared via a lateral entry program. The remaining three teachers were prepared via an in-state public undergraduate education program; however, one teacher received her master's degree in education prior to entering the classroom from an out-of-state private institution. The majority of the teacher participants taught English language arts (ELA), although the sample does include teachers from all four core subject areas. In terms of the representativeness of the sample, I examine gender, race, ethnicity, and college GPA of the full population and study sample of teachers. The percentage of female teachers in the sample is slightly higher than the full population of beginning middle school teachers in the district. The racial composition and college GPAs of beginning teachers in the sample was equivalent to the full population. See

Tables 3.1 and 3.2 for characteristics of the participating schools and individual participants, respectively.

Data Collection

Semi-structured interviews ranging from 45 minutes to 75 minutes were conducted once with each participant. Interviews were conducted from May 2013 to January 2014. The semi-structured interviews focused primarily on the job search process, hiring, recruiting and assignment within schools. In-depth interviews provide information-rich data for analyzing the assignment processes that could not be obtained by other methods of data collection such as surveys, observations, or content analyses. The semi-structured nature of the interviews allowed the flexibility to probe in greater detail, where necessary. Interviews with teacher participants focused on their job search experience and their teacher assignments. For example, teachers were asked, "What things were you looking for in a school?" and "As a beginning teacher, what was the process used to assign you to a specific grade, subject, and level in your school?" School leader participant interviews focused on the recruiting, hiring and assignment of teachers. Key questions included, for example, "How do you recruit teachers?" and "Tell me about how you make assignment decisions. Who's involved in the process?" See Appendices A and B for teacher and school leader interview protocols. Fieldnotes included memos about the participants, the interview setting, and information about the district was also collected.

Data Analysis

Following Yin (2009), this study employed an explanation-building technique. The primary purpose of this technique is to build on "how" or "why" teachers are staffed between and within schools. This strategy is useful for developing ideas for future studies (Yin. 2009). The use of semi-structured questions or "case study protocols" (Yin, 2009; George, 1979) was

created to provide evidence for the study's research questions and aligned with the conceptual framework.

Teacher and school leader interviews were collected, coded and analyzed using ATLAS.ti, a qualitative data analysis software tool. Prior to inputting the data into ATLAS.ti, a codebook was developed from the interview protocols for the teachers and school leaders; however, additional codes were added. Data was then inputted into the software and was coded into categories developed from the research questions. Major, a priori, categories derived from the conceptual framework included teachers' job market search and application process, school leader recruitment, the hiring process, and teacher assignment. Multiple readings of interviews ensured immersion and familiarity with the data and to establish additional codes and themes. Codes used for the analysis were selected based on the research questions and conceptual framework.

Role of the researcher. Although this study is based on systematic inquiry, the interpretations reflect my understanding of how teachers and school leaders describe the staffing and teacher assignment process. My interest in understanding how educational inequities develop and persist can potentially obscure my ability to assess the issue more broadly. For example, I may be hypersensitive and thus probe more on topics of tracking or teachers' decisions to apply to schools based on socio-demographic compositional characteristics.

As a graduate student who has never taught at the K-12 level, I may be considered an outsider to the participants, which has benefits and consequences for the study. My outsider status as a graduate student may provide more candid responses because participants perceived my role as a student less powerful or threatening than a professor, educational leader or policymaker. Moreover, respondents may openly provide detailed responses regarding staffing

processes due to my lack of lived experience. Conversely, I may potentially overlook contextual factors that affect teacher assignment choices due to my limited experience. However, as an outsider I am able to maintain criticality in the analysis that could be overlooked if I were familiar with being a teacher (Tinker & Armstrong, 2008).

<u>Reliability and Validity</u>. Creswell and Miller (2000) provide eight validation strategies used by qualitative researchers in design, data collection, analysis, and reporting. In this study, I focus on three of the eight strategies. First, I triangulate the data by conducting interviews with teachers and their school leaders to provide corroborating evidence of research findings. Second, I critically reflect on my positionality and recognize biases from the outset of this study. Finally, I use a peer review or debriefing strategy with peers external to the study that challenged the assumptions and interpretations I made in the study.

Findings
Where Beginning Teachers Apply and Accept Positions

Teachers had three major reasons for selecting middle grades. First, compared to elementary school students, middle school students are perceived as more mature and independent. Elementary school students were referred to several times as "babies." Second, teachers were not confident in their ability to teacher high school-level content. Third, choosing middle grades hinged on teachers' implicit *savior mentality*. That is, teachers believed they could directly impact students' life trajectories during students' adolescent years. For instance, one teacher said, "there is so much potential in middle school, and that's the age where they are starting to turn. Where they can take a right turn or left and I really wanted to be a guiding force working in middle grades." According to the teachers, high school students are considered "already too set in the ways to affect them very much."

In terms of the factors that influence the schools where beginning teachers apply, there are no consistent patterns. Many applied to several schools with varying socio-demographic characteristics, performance levels and geographic locations. Regardless of school characteristics, beginning teachers were simply interested in obtaining employment. The number of applications submitted by beginning teachers ranged from one to twenty. Location was a primary factor for two lateral entry teachers, who were hired in the school they previously worked as substitutes. Only one teacher mentioned setting criteria when determining where to apply. Specifically, she examined student test scores and teacher turnover:

> I did research the schools…I was looking at high turnover rates because if they have a high turnover rate that means something is going wrong that your teachers are leaving. I looked there. I also looked at EOG scores, but I didn't want that to also deter me away from the school. But it said something about the school if their EOG scores over a three or four-year span continued to decline and they're not making the uphill. Then something's going on. That was pretty much it. That was all I was really looking for.

Interestingly, despite these criteria, she accepted a position at Bell Middle School[10], which has a high teacher turnover rate and is on the cusp of being classified as a low-performing school.

In terms of where teachers accept positions, many discussed the importance of a supportive administrative team in their decision making. For instance, the teacher who accepted the position at Bell Middle School despite her desire for a school with a low turnover rate did so because she really liked the principal when they met at a job fair. Another teacher states that during her interview with the principal from Bell Middle School, "I told her that…I wanted to be at a place where I thought my principal would advocate for me."

Other factors that are important in accepting positions are the camaraderie among staff and current teachers who enjoy their jobs. According to a teacher, "it's very crucial to have that, a friendly environment. I wanted to be able to walk in the school and be like 'oh I'm home.' Not

---

[10] For anonymity purposes, pseudonyms are used for school, teachers, and school leaders' name.

like, 'uh, I hate coming to work everyday.'" However, many of the teachers in the sample were not interviewed by their potential colleagues nor did candidates meet with potential colleagues prior to accepting their teaching position. Therefore, it is important to note that teachers' responses may reflect their current working condition preferences.

How and Who Schools Recruit

Although most schools hire heavily during the summer months, the recruitment process begins as soon as there is a known vacancy. Across the schools in the study, there are various reasons teachers leave the schools, including retirements, promotions, moving for family reasons, and changes in student enrollment. When vacancies occur, schools use a combination of recruiting tactics to fill positions. For example, the AP at Bethune Middle School reports:

> When there's a vacancy we work within our school, but we do have to notify human resources that the vacancy is up an then they will pull - say if we had one next week, then we'd notify them and they'd look through their candidates. And sometimes we even share across schools, put an email out, "looking for a language arts teachers." If they've interviewed one and didn't hire them they may share the name with us and we contact that way."

Schools also recruit using resumes from teacher candidates that had previously applied to the school. As the principal at DuBois Middle School states, "people send me resumes year round. There's somebody looking for a job in education year round and I always keep those resumes in a file...and that's usually the first place I look before I go to the direct website." University and district-level job fairs are another avenue schools recruit teacher candidates. Finally, schools recruit individuals already working in their schools as substitutes or preservice teachers.

It is common knowledge that the number of applications a school receives impacts their ability to recruit and hire teachers. School leaders, without exception, report that the number of applications varied given the time of the year and the subject matter of the vacant position.

157

Consistent with national teacher shortage statistics, the principal at DuBois Middle School reports, "You're going to have more applicants in social studies, possibly language arts and you're going to have fewer in math and science, but you still have quite a few."

Among the schools, teacher education preparation plays a significant factor in recruitment. For example, school leaders are very reluctant to recruit lateral entry teachers, because of their perceived lack of quality and experience. In describing her reason for not hiring lateral entry candidates, the principal at DuBois Middle School said,

> In general, I don't look at lateral entry candidates first. I don't. I have hired several lateral entry people, but they're not my first people to look at because in general most of the time you're at an advantage if you can get someone who already has their school behind them and maybe has some experience, but it depends upon what you're looking for and how hard that is to staff.

In addition, schools had mixed hiring practices among preservice teachers from traditional preparation programs. For instance, school leaders at Bell Middle School, an average-performing city school, do not actively recruit preservice teachers assigned to their schools. However, school leaders at DuBois Middle School, a high-performing rural school, actively recruits and hires preservice teachers assigned to their school.

Finally, while Bell Middle School did not actively recruit preservice teachers from traditional preparation programs, school leaders were interested in recruiting out-of-state prepared teacher candidates. The AP at Bell Middle School stated, "we felt like in certain positions out-of-state people had a more qualified certification. Their qualifications have lined up a little bit better with what the position asked for."

School leaders all discussed how student enrollment and budget reductions affect the quality of teacher candidates they are able to recruit and hire. If student enrollment declines, schools are required to transfer teachers. According to the district's policy, teachers are

transferred based on their willingness to be reassigned and seniority, which could potentially mean losing high quality teachers. On the other hand, when student enrollment increases schools are required to hire first from the district's transfer list without an opportunity to recruit and hire more qualified candidates. Often referred to as the "dance of the lemons," principals expressed how this reassignment process impacts the quality and quantity of candidates. The principal at DuBois Middle School discussed how this reassignment process affects school culture as well as students:

> The problem lies in the fact that sometimes you don't get people that are a really good fit for the school...You try to welcome the new teacher in as much as possible. You hope that they're going to be a good fit and sometimes they are but sometimes they're not. And if they're not, you have to ask people to be patient with them. Really because it's not like you can change your schedule back. You can't do it. I think you try to assign somebody from the school that can help work with them to, I guess, acclimate themselves better to the school, but sometimes there's not a whole lot that can be done...And that's tough because if you have a situation like that where kids have gotten to know their teachers for 20 days of school now and some of their schedule's going to be changed.

The Hiring Process of Beginning Teachers

As a moderately decentralized district, all candidates must be screened at the district-level prior to being hired; although there are cases where teachers were hired prior to being screened. The district screening is a formal, structured interview where 2-3 principals conduct a face-to-face interview with teacher candidates. The goal of the screening process as one principal described it is to "give you a little bit more of an idea of their [teachers] philosophy." The screening tool was developed to efficiently increase the number of qualified teachers in urban districts. Teachers are scored by each of the principals and receive a final score, which goes into their application file; however, school leaders considered the score a facet of the hiring process, but not the ultimate determining factor.

Principals serve as the "gatekeepers" for hiring decisions within their schools and are the primary person responsible for screening candidates. However, if principals are overwhelmed,

159

assistant principals help with the screening process. The principal at DuBois Middle School further supported her role as the gatekeeper by allowing APs to call references and screen for non-teaching positions.

School leaders typically invite five to seven teacher candidates in for interviews. To understand schools' level of rigor in hiring the best candidates, several elements of the interview process are considered: materials teacher candidates were asked to bring into the interview, the number of interviewees, and the type of interview questions. There was no evidence that school leaders request performance related materials (i.e., test score data, Praxis scores, student transcripts, etc.) of beginning teachers during the interview, although, one school leader mentioned asking for Educator Value-Added Assessment System (EVAAS) data. Despite the lack of required materials, candidates usually brought their portfolio or sample lesson plans and reported using these as evidence to support their responses during the interview.

The literature suggests that requiring teacher candidates to perform sample lessons is an indicator of a rigorous selection process. In general, school leaders did not require teacher candidates to perform a sample lesson. The principal at DuBois Middle School hoped to incorporate sample lessons, but found it to be difficult; especially given interviews typically take place during the summer. However, a lateral entry teacher from Bethune Middle School, an average-performing city school, reported that she was required to submit a written lesson plan and a video when she applied to her current school. Interestingly, the AP at the school did not discuss this during our interview.

The individual conducting the interviews varies across schools. At the most basic level, the principal was the only person that conducts interviews. A more complex panel included the principal, all APs, the curriculum specialist and the curriculum coordinator. As the principal at

160

DuBois Middle School reports "I really do like having other people with that because they many many times will pick up on something I missed. They might be able to red flag or something that was very positive that maybe I didn't really hear..."

Teachers are also included on the panel of interviewers for some schools, although, school leaders had different opinions about the use of teachers. The AP at Bell Middle School, an average-performing city school, shared that teachers are not always included during the interview process because school leaders lack trust in their opinion. In his words, "…but on occasions like that, we'll have teachers come in. We really have to trust the teacher's opinion though. They have to be pretty dedicated to the school." Conversely, when asked about including teachers in the interview process, the principal at DuBois Middle School, a high-performing rural school reports:

> The ultimate way and the way of the future I think is to have more input from your teachers – maybe some people from the team they'd be working on or the subject area department they'd be working on but I don't always do that. That's a goal and they are sometimes involved, but that is something I still like to have some control over because I'm the one who has to deal with them if they don't function well and that's something I'm not quite willing to give away without having some input in, but we bring them in for the interview…

Across both schools, maintaining power over the interview process is important for the school leaders; however, they differed in actively using teachers during the process. Although, school leaders at Bell Middle School did not use teachers during the interview process, teacher participants from this school reported that having teachers in the interview process would have been helpful as a gauge of compatibility.

Most of the interviewers used scripted interview questions that covered questions about classroom management, skills in differentiation, teaching philosophy, use of technology, what a typical day in their classroom would look like, and self-assessments (i.e., why did you become a

teacher). Interviews with teachers confirm these are the questions most asked during interviews.

There are some differences across schools in personality-related versus professional skill-related

questions. School leaders from Bethune Middle School, an average-performing city school, asks

more personality-related questions:

> We had a list this year and we just go back…the list, giving a scenario, asking about taking us through a typical day, telling us about yourself, why do you think you're a good fit for the school…Tell us how you develop relationships, how do you communicate with students as well as parents, keep them informed of what's going on in your classroom…

Whereas, school leaders from Bell Middle School, an average-performing city

school, asks more professional skill-related questions:

> We will typically ask them to describe a lesson that they've done, especially when it comes to how have they utilized differentiated instruction. How have they dealt with a difficult student? How would they plan instruction if they saw that most of the class didn't do well on the test? If I were to walk in your classroom on one typical afternoon, what would your classroom look like? What would you be doing?

> School leaders were listening for several indicators to determine whether candidates are a

good fit for the school. In general, school leaders are interested in candidates with strong content

knowledge, able to build positive relationships, able to manage his/her classroom, and able to

continuously grow and develop professionally. The principal at Bell Middle School provides a

succinct list of qualities schools in the study look for teacher candidates to possess:

> Knowledge of the curriculum, knowing and sharing, you may not necessarily know but sharing how they develop positive relationships with students, how they discipline students, classroom management piece, just building those relationships and being able to work collaboratively with other teachers as well because that was an important piece too. And the willingness to learn, willingness to accept constructive criticism and to learn and be open to new ideas and just accept challenges that you faced each day.

> School leaders do not differentiate between beginning teachers and experienced teachers

in terms of the qualities they are looking for in a teacher candidate; however, they are aware that

beginning teachers have limited experience and are listening for potential and willingness to develop.

After the interview, school leaders report debriefing with the panel to make a final selection. There is no formal evaluation rubric that is used to evaluate candidates during or after the interview. School leaders report simply discussing the candidate with the panel of interviewers and "rate them as they go." School leaders also discussed using pre-determined cues to signal to one another that a candidate is not a good fit.

School personnel interviewed all of the teachers in the study at some point; however, two lateral entry teachers were not interviewed for their current teaching position but had been for a position that they previously applied for in the school. Both teachers served as substitute teachers in their current schools prior to being hired as a regular classroom teacher. One at Burroughs Middle School interviewed for a position in the prior year, but was not hired. She stated, "So for this – for what I'm doing right now, there was not technically – I just kind of slid in." The other teacher at Bethune Middle had also applied for another position at the school but was not hired. When her current position opened, she resubmitted her cover letter and resume to the principal, was told to go through the district-screening interview and was hired without a formal school-level interview.

Hiring Challenges

Hiring challenges expressed by school leaders are manifold and are consistent with the extant literature. First, within schools, teachers who are knowingly leaving the school delay submitting their resignation forms. The principal at DuBois Middle School reported:

> They can verbally say I'm going to retire, but until they fill out the paperwork – I can be searching for candidates and I often do but I couldn't officially hire anyone or offer them a job until the paperwork has been turned in to say I will resign on June 10th or whatever that last day of school is. The paperwork is really what speaks and they may tell you verbally.

163

Second, districts create hiring challenges such as the district-level interview requirements and screening requirements. The district requires schools to interview at least three teacher candidates for a given position; however, schools often have difficulty finding three qualified candidates, especially when hiring for specialized positions.

Bell Middle School's AP discussed how the district-level screening interview requirement slows down the hiring process:

> There are some things that people have to go through. If they haven't been through the screening interview and they happen to shoot [the principal] an email saying, "This is my resume" and they are a really good candidate then they have to be screened. We might interview them but say, "You have to be screened first." That can be a little bit of red tape I guess.

Finally, schools create hiring challenges by interviewing candidates without ensuring application materials are complete at the district-level. This occurs most often when hiring candidates, first year teachers or teachers outside of the district. Schools also experience challenges, by delaying hiring in efforts to attract underrepresented teachers. This delay in hiring directly affects current staff, especially when positions are not filled prior to the start of the school year. The principal at DuBois Middle School, a high-performing rural school, candidly describes the challenge with hiring underrepresented teacher candidates:

> It's time consuming to hire the best candidate. Sometimes there aren't as many applicants as you would like to choose from. Sometimes you're looking and it's almost time for school to start and I'm not going to move everybody around. I've already told them what they're going to teach, what room they're going to be in, I might be looking for a particular set of skills or I might have all women on a team and I'd really like to have a male or to be honest, for a school like mine, it's also hard to find diversity, to bring diversity in with any kind of just minority.

The third research question examines three mechanisms that could provide a deeper understanding about the teacher assignment process. The findings related to these mechanisms are discussed in the next section. Table 3.5 summarizes the presence of the mechanisms.

164

Federal and State Policies and Teacher Assignment

At the middle school level, teacher assignment is primarily based on teachers' license area. However, schools leaders did use student test score data (i.e., EVAAS) to reassign teachers to different subject areas. The principal at DuBois Middle School was open about moving teachers based on their EVAAS scores:

> But her math was not looking good and I said you know what, if this doesn't look better next year and I have an opening in social studies, you know I'm going to have to put you in social studies because I need to put you where you can be the most effective and for whatever reason something is going on with math and I don't know.

Conversely, the AP at Bethune Middle School has not and is reluctant to reassign teachers, but more willing to allow teachers to "either volunteer to just leave and go somewhere else" based on low EVAAS scores. A teacher from Bell Middle School also suggested that school leaders are more willing to reassign teachers across schools than reassign teachers within schools based on low test scores. In the case of within school reassignment, it was easier to reassign teachers that had dual licenses. Teachers use this information to strategically acquire dual licenses to ensure job security in the event of mandatory reassignments. A teacher from Bell Middle School who teaches ELA and social studies reports:

> One of my peer teachers, my mentor actually, she mentioned there is going to be an opening for the language arts and I would really like it if you maybe took the praxis and I think you could do it. She recommended that I do it. And so then I brought it to my principal during one of our post observations and I said, you know, is this something, and she said that would be great, yes, if you want to do it, great, go for it. I was like well, you know -- because I was worried about job security because we lost a lot of people last year and so I was like, oh, take the initiative.

Although this teacher chose to ensure job security, she also acknowledged the difficulty of teaching two subjects, ELA and social studies, and frustrations with accountability pressures she experiences while teaching ELA. Another teacher at the same school was aware of the job

security "game" but refused to play. However, she would consider teaching a different grade level if necessary:

> I love science, and I don't want to do social studies. So I like it, I love English, but I'm marketing myself as a science teacher, I want to teach science. It does put me in a box, but I like that box. Yeah. It puts me in just that science box, but I'm okay with that. I'm okay. I mean, if push comes to shove, I'm still 6th thru 9th, if [the principal] needed me for -- I'm not speaking this -- but if [the principal] needed me for 8th grade science, I've taught 8th grade before and that's what my certification is for, 6th thru 9th.

In order to build some level of "stability," first year teachers at Bell Middle School are not reassigned due to low EVAAS scores. The AP did not rule out the possibility of reassignment, but simply stated that it does not happen often.

> In my experience, we haven't changed a lot of first year teachers' assignments because when you're looking at a first year teacher, one of the best things that you can do is give them stability. It can happen, but we don't typically move somebody at least for one year I don't think.

Although, school leaders may not reassign first year teachers, they still use EVAAS scores as a tool to assign teachers based on their strengths. The principal at DuBois Middle School, a high-performing rural school, describes balancing act she experiences in assigning teachers based on EVAAS scores, while maintaining an equitable teacher assignment distribution:

> Value added data in the district tells you that you should assign teachers according to their strength by value added and with value added in math and in language arts, you're going to see who does well with the low kids, the middle kids and the high kids. And a lot of time they'll do well with two groups. Sometimes they'll grow all three groups…If [Teacher A] is fantastic with high kids, then by value added, I should give her all the high kids but then I'm going to have all the high kids on one team and then [Teacher B] is going to have all the low kids and every kid in the building and every teacher in the building and every parent in the community is going to say that's the dumb team, that's the smart team and we're not doing that. We're just not going to do that. What I would try to do is make sure that she has at least one of those high groups of kids. If her value added data showed she's really good at growing high kids, I'm going to be sure she has at least one group of those AL or upper level math or whatever kids. I'll make sure she has at least one group or maybe two, but she can't have all high kids.

The balance between data-driven decision-making and equity-based decision-making is also discussed by the AP at Bell Middle School, an average performing city school. The AP used EVAAS scores to assign students to classes, but employed a "spread the wealth" philosophy with regard to teacher assignment:

> We do use EVAAS scores. We'll also look at EOG scores across the board, how their classes did. And I think by spreading the wealth out you can't say, "Well, that team has higher students so that's why they did better." You might be able to find small differences between a team but you're not going to find glaring differences between our teams. You're not going to find that. And, as a result there's not an excuse, there's just not.

When asked about the use of the North Carolina Educator Evaluation System (NCEES), another accountability driven initiative, in teacher assignment strategies, school leaders and teachers indicate that the evaluation system is used for professional development, not teacher or student assignment strategies.

Institutional Politics and Teacher Assignment

In this study, two forms of institutional politics were examined: parental preferences, on one hand, and teacher preferences and teacher seniority, on the other. Overall, teachers and school leaders both expressed that parents have some influence on student assignment, but no influence on the teacher assignment process. Nonetheless, parents at DuBois Middle School are allowed to voice their opinion in a survey that "gives every parent a chance to provide input on the kinds of characteristics [they] want in [their] child's teachers next year." The findings are consistent with the results from Carey, Farris, and Carpenter's (1994) study.

As previously mentioned, school leaders actively assign teachers based on a "spread the wealth" philosophy where every teacher regardless of experience level receives the same course load and multiple ability level courses. However, teachers' preferences in the two high-performing schools, DuBois and Burroughs Middle Schools, were considered using a feedback

form created by school leaders. School leaders try to accommodate teachers' requests; however, this was not always possible. On the whole, teachers were able to provide their preferences, but were also flexible if reassignment was necessary.

Consistent with Cohen-Vogel's (2011) study, school leaders are adamant that seniority does not matter in teacher assignment decisions. While North Carolina is a non-collective bargaining state, seniority is important in determining reassignment under the district's mandatory reassignment policy. Under the policy, principals reassign teachers based on "least service seniority."

Institutional Structures and Teacher Assignment

Overall, licensure area is the main factor in the teacher assignment process; other factors school leaders consider include teacher's personality, skill sets, and strengths. The principal at DuBois Middle School reports how teachers are assigned based on the aforementioned characteristics. She states:

> I'll show my piece of poster board…here's my little post-it note chart where I look at the names of the teams and then this is the subject area. These are language arts. These are math. These are science and these are social studies and then with moving these around, right here, this teacher is scheduled to come back. She's been out on leave. She has her administrative certificate. She's trying very hard to get a job as an assistant principal. Right now she is here because she's on contract to come back but if she doesn't, she's going to move here from down here because she's certified in two areas. She taught math for me, but her strength is science and so she will stay here unless an opening comes up. You do these little things and I can move the post-its around and do combinations of people.

The AP at Bethune Middle School mentioned that school leaders begin with student assignment "then we find a teacher and put this group of teachers with this teacher, AA and then from there, we look at the team as well." In terms of classroom composition, across all schools in the study, math and ELA courses are homogenously grouped by student ability and science and social studies are heterogeneously grouped. The grouping strategy is based on a district-level

168

initiative to provide differentiated curriculum and instruction. Schools also organized their teachers and students into grade level academic teams. Each academic team includes four teachers – one each in ELA, math, science and social studies. In general, schools had either three or four teams per grade level. Due to a decrease in student enrollment, Bell Middle School had a six-person team, which was an anomaly for the school.

Overall, the assignment of students within each teacher is moderately equitable, in that all teachers are required to teach multiple ability-level subject specific courses. Tables 3.3 and 3.4 provide illustrative examples of the team structure and teacher assignment at an average-performing and high-performing school, respectively. As shown, Bell Middle School has three academic teams. Each teacher teaches four math courses; however, the number of high-ability, average ability and low ability classes assigned to teachers differs. Despite what appears to be an equitable distribution of teachers, Table 3.3 shows that Teacher A has the "ideal" assignment, with the greatest number of advanced classes and the fewest low ability classes. Similarly, Teacher B on Table 3.4 has the preferred teacher assignment.

Concerning teaching multiple ability-level courses, a teacher at Bethune Middle School expressed mixed feelings about the teacher assignment process. She indicates that the structure is a fair process for teachers; however, it is harmful for students because not all teachers are skilled at teaching all ability groupings of students.

> I think it's fair that this year they said every single teacher that teaches a core class like math, language arts, science, social studies has a regular, and I have my fingers doing quotation marks. Some of these kids are dumb as rocks. Sorry. I think that's fair because advanced learner kids are easy to have in class, but they're supposedly harder to grow. Okay. EC kids sometimes are difficult to have in class because of their issues developmentally, and they're harder to teach, but if you can get a hold of them, they're easy to grow because they have further to grow. Okay, and regular kids are just a crap shoot, right? So it's fair if everybody has all of it, right? Instead of me having all AL kids. Well, that's an easy day for me, but I'm not getting any money with EVAAS data.

I don't think I'm highly qualified to teach EC kids, regular kids, and AL kids. I've been doing this for a year and a half. Do you think that's fair to all those kids?...I don't care if you send [teachers] to classes to teach all three. They're not going to be good with all three.

DuBois Middle School also use diversity related politics to assign teachers across teams. The principal was deliberate about diversifying teams based on race, gender, experience and personality. In her words,

I like to have – we do have several male teachers in the building. I like to have at least one male teacher on every team. I think logistically there's a lot of good reasons why that's good. I think it's good for the kids because you do have more male teachers in middle and high school but elementary if they haven't had one before to kind of get them used to that and so I wouldn't put all the men on one team. I would try to spread them out…Whatever their differences are we spread them out whether it's that, gender, minority status, how much experience they have. I'm not going to put all the veteran people on one team and the new people on another.

Several teachers discuss their preference for teaching advanced level students. One teacher in particular expressed the delight she receives from teaching advanced students:

…they're so interesting, and they're so witty, and they've all usually been together for a long time that they've been in advanced learner classes through their elementary school. They've been in all the other same elementary school, but they've known each other for awhile, and they like each other. They're funny. They're smart. They read. They do their work. They carry on conversations. You know what I mean? They're so smart. For the most part they've got parents that are looking after them, that take that, that pay attention to what they're doing at school, that feel them, that make them take a bath.

Although science courses are heterogeneous, the science teacher in the study also indicates that she "wouldn't mind teaching an honors or an upper level science" at some point during her career.

Beginning teachers also expressed some frustrations with their assignment. For example, despite being made aware of their grade and subject assignment, teachers often do not receive classroom rosters until a week before the start of the school year. Therefore, teachers are not able

to adequately plan for their courses with such late notification. The science teacher from Bell

Middle School states:

> …before open house, I may get a roll, but I haven't seen some of them, so I don't know. I don't know their abilities or disabilities or any of that until we get an IEP at a glance or a parent wants to pull you aside on walk the schedule night and say, "My kid has got..." And I'm just like "okay", and I'm not going to remember that on Monday, so.

There are mixed views regarding teaching dual subjects. The teacher at DuBois Middle

School did not mind the dual subjects and expressed that he was able to see his students twice a

day, which allowed him to understand their strengths in multiple subjects. In contrast, the

teacher at Bell Middle School felt extremely overwhelmed teaching two subjects.

Discussion

Education reform initiatives have placed made considerable efforts on ensuring schools

are staffed with high-quality teachers, especially schools with majority low-preforming, racially,

ethnically, and economically marginalized students populations. Additionally, how teachers are

assigned to specific classrooms has become an increasingly important issue as well. This study

provides a preliminary examination into the black box regarding how beginning teachers are

staffed and assigned to schools and classrooms, respectively.

While extant research suggests that teachers hold preferences for schools composed of

majority White, high-performing students, and schools close to their hometown or region

(Hanushek, Kain, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebrickner, 2007; Boyd, Lankford,

Loeb, & Wyckoff, 2005); I find that beginning teachers do not strategically search for or accept

positions based on their preferences. A possible explanation for teachers' behavior is a perceived

lack of options and perceived sense of urgency stemming from current economic conditions

across industries in the U.S. Lateral entry teachers appear to be the most strategic in their job

search, with greater emphasis on the distance from their homes primarily because of family obligations.

From a policy perspective, the mismatch between teachers' actual school placement and their preferred placement may have implications on school culture, instructional quality, and teacher retention. For example, teachers who take positions in schools with high concentrations for economically disadvantaged students, but have preferences for schools with high concentrations of economically advantaged students, may have a challenge understanding and adapting to the school's culture and student population. If teachers are unable to adapt and embrace a less preferred working environment, they may be more prone to leave their current school, which negatively affects students' opportunities to learn.

While beginning teachers may not be strategic in their job search, school leaders should be more strategic in their search, because they are aware of the specific staffing needs and culture of their schools. Interestingly, the findings suggest that schools do not deliberately seek to hire beginning teachers, which may be a result of accountability pressures to ensure schools are staffed with high-quality and effective teachers. In addition, prior literature indicates that strategies such as requiring candidates to conduct sample lessons and reviewing portfolios provide important signals related to the quality of a candidate (Boody, 2009; Abernathy, Forsyth, & Mitchell, 2001). However, school leaders in this study do not use information that would enable them to effectively assess candidates' quality.

From a policy perspective, schools could require beginning teachers to present a sample lesson to gauge their potential effectiveness. Although most candidates are hired during the summer months, schools could require teachers to submit video sample lessons prior to the interview. This may be less of a challenge for schools given the increased use of new portfolio

assessments such as the edTPA that require portfolio videos, by several teacher preparation programs across North Carolina.

Turing to the teacher assignment, the findings from this study are inconsistent with Cohen-Vogel's (2011) findings that school leaders use test score data to assign teachers. Similar to Hanushek and Rivkin (2010), I find no significant reassignment patterns based on test scores among the schools in the study. This might be due to middle grade licensure requirement, which prevents school leaders from moving teachers to non-tested subjects. On the contrary, Cohen-Vogel's (2011) examined elementary schools, which do not have the same licensure restricts allowing for more flexibility in reassigning teachers.

Another major finding related to teacher assignments shows that school leaders' use of a "spread the wealth" philosophy in the teacher assignment process among ELA and math teachers. This strategy is again contrary to prior research, which highlights the use of accountability and teacher tracking in the teacher assignment process. Regardless of experience, skill, or preference, school leaders assign teachers in a manner that ensures that no single teachers exclusively teaches high-performing classrooms nor are teachers assigned to only low-performing classrooms. This equity-based teacher assignment practice seemed to overshadow any federal and state accountability pressures as well as preferences. Interestingly, the underlying assumption in the "spread the wealth" practice is that all teachers are equally effective with different types student ability groups, despite the limited empirical evidence. If the argument that good teachers are good for all students, then heterogeneous teacher assignments seem appropriate. However, if this argument is false then we may want to reconsider the weight placed on equity-based teacher assignment practices. To my knowledge, there is only one study that has empirically tested the hypotheses that a good teacher is beneficial for all students; however, this

study specifically characterized "all students" as English learners versus non-English learners subgroups (Loeb, Soland, & Fox, forthcoming).

While this study adds to the paucity of research on staffing and assignment of beginning teachers, the limitations of this study must also be considered. The teachers in the study represent only those who were successful job seekers, which consequently excludes beginning teachers who were unsuccessful in the job market. These unsuccessful job seekers may have provided important information regarding the types of schools teachers apply to and the job search process. Additionally, the conclusions drawn should be taken with caution given the small sample size. As previously noted, the sample is similar to the population across racial composition and college GPA; however, there may be other demographic factors, which vastly differ. For example, if the study sample has a smaller proportion of lateral entry teachers than the actual population, the findings regarding staffing decisions may not adequately represent the lateral entry population.

Careful consideration is made in discussing the implications of the findings. The data in this study depends on what teachers and school leaders say they do, which is subject to several forms of bias such as consistency bias, moderator acceptance bias, and social acceptance bias. More qualitative work is necessary to substantiate whether there are differences in hiring and teacher assignment practices for low-performing schools. Findings here do not suggest strong differences between average and high performing schools; however, low-performing schools might experience additional pressures to improve student achievement that would cause them to be more strategic in who they hire and how they assign teachers to courses. Studies might also be designed to further examine actual reassignment patterns of beginning teachers both between and within schools using administrative data.

As states and districts continue to create school-wide policies and interventions aimed at closing test score and opportunity gaps – especially by race, ethnicity, and economic status – who teaches our children is necessary question that we must continue to examine in hopes that all students have the opportunity to receive a 'sound basic education.'

Figure 3.1 Conceptual Framework

```
                    ┌─────────────────────────────┐
                    │  Geographic Preferences     │
                    │   • Student Composition     │
                    └─────────────────────────────┘
┌──────────┐
│ Supply of│                                              ┌──────────────┐
│ Teachers │                                              │  Job Search  │
└──────────┘        ┌─────────────────────────────┐       └──────────────┘
                    │   Working Conditions        │
                    │   • Student Composition     │
                    │   • Leadership              │
                    │   • Colleagues              │
                    └─────────────────────────────┘
```

Supply of Teachers

Geographic Preferences
• Student Composition

Working Conditions
• Student Composition
• Leadership
• Colleagues

Job Search

Hiring and Staffing Process

Teachers Assignment to Classrooms

Schools' Demand for Teachers

Vacancy
• Notification of Leaving
• Transfers
• Student Enrollment

Recruitment

Assignment Mechanisms
Federal and State Policies
• Value-added scores
• Evaluation ratings
Institutional Structures
• Academic Teams
• Course Offerings
• Hierarchy of Courses by rigor
Institutional Politics
• Parental Preferences
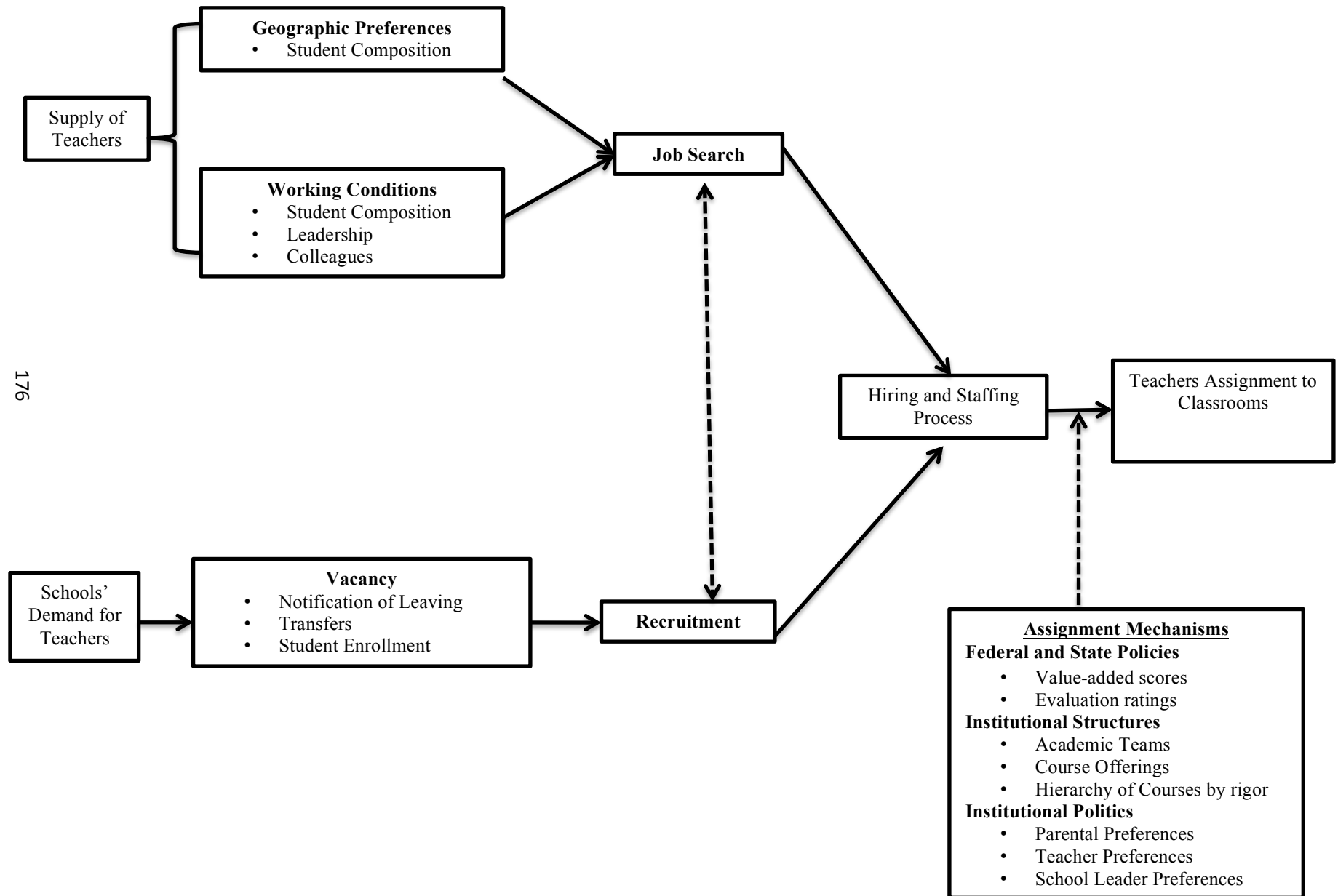• Teacher Preferences
• School Leader Preferences

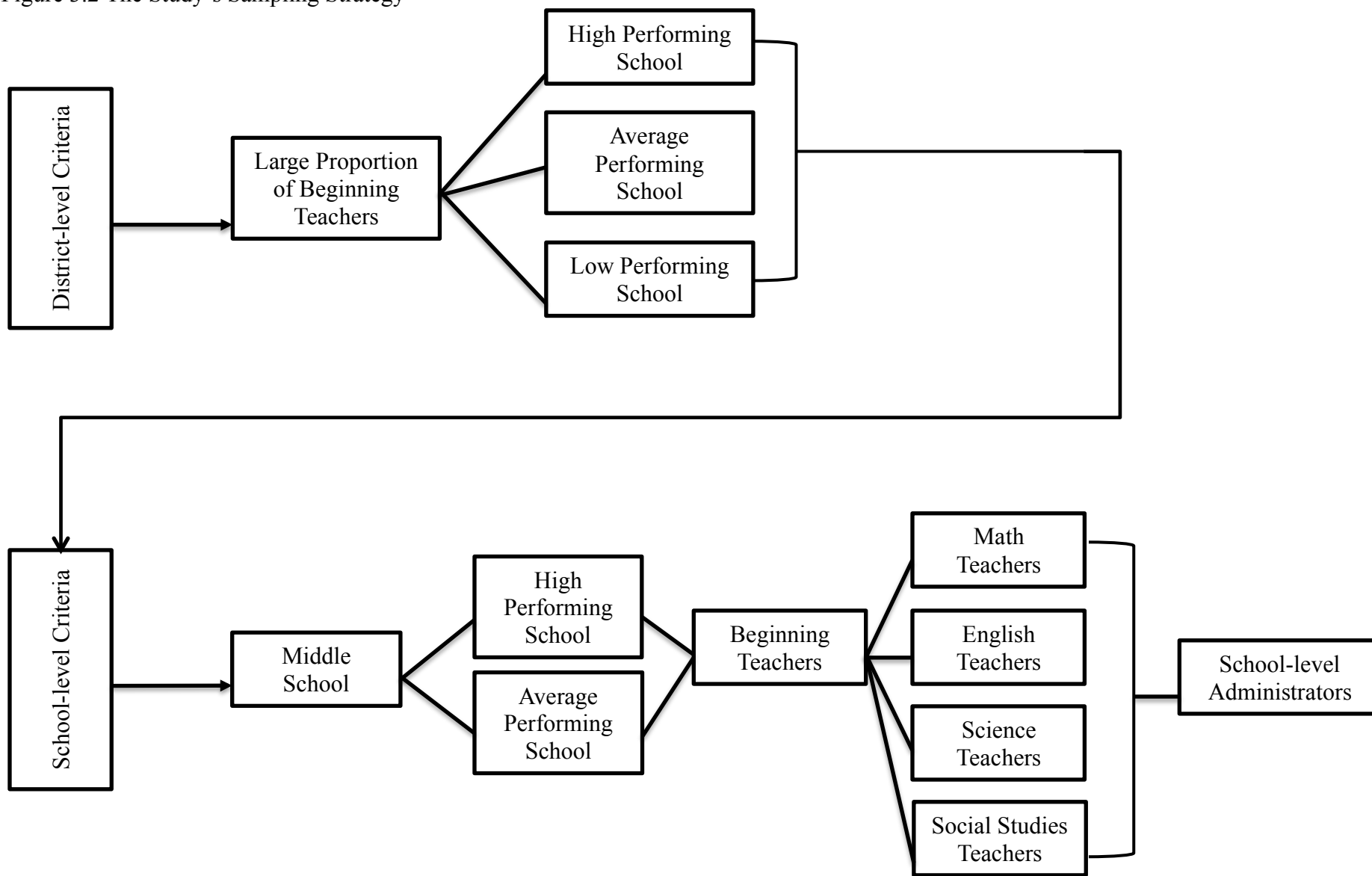Figure 3.2 The Study's Sampling Strategy

Table 3.1 Profile of Study Schools

School	Performance	Urbanicity	Title I	School Size	Number of Teachers	Turnover Rate	% SOC (2011-12)	% FRL (2011-12)
Bell Middle School	Average Performing	City	N	800-900	50-60	20-30	60-70	60-70
DuBois Middle School	High Performing	Rural	N	1100-1200	60-70	5-10	60-70	50-60
Burroughs Middle School	High Performing	Rural	N	900-1000	50-60	10-20	30-40	40-50
Bethune Middle School	Average Performing	City	Y	800-900	60-70	10-20	70-80	70-80
All Schools (actual average)	---	----	---	973	60.5	15	58.75	57.25

Note: For anonymity purposes, pseudonyms are used for school names. Data are presented in ranges to protect the identity of schools. The data come from the North Carolina School Report Card. The shaded rows represent high-performing and rural schools; non-shaded rows represent average-performing and city schools; SOC=students of color; FRL= free and reduced price lunch eligible

Table 3.2 Profile of Study Participants

School	Participants	Grade	Subject	Preparation	Gender/Race
Teachers					
Bell Middle School	Hallie	8	Math	Traditional prep	BW
Bell Middle School	Anna	7	ELA/SS	Traditional prep	WW
Bell Middle School	Lucy	8	Science	Lateral entry	BW
DuBois Middle School	Joseph	6	ELA/SS	Traditional prep	WM
Burroughs Middle School	Jennie	6	ELA	Lateral entry	BW
Burroughs Middle School	Edna	8	ELA	Lateral entry	WW
Bethune Middle School	Ethel	7	ELA	Lateral entry	WW
School Leaders					
D. Bell Middle School	Octovius	6/8	N/A	N/A	WM
DuBois Middle School	Septima	N/A	N/A	N/A	WW
Bethune Middle School	Bazoline	7	N/A	N/A	BW

Note: The shaded rows represent high-performing and rural schools; non-shaded rows represent average-performing and city schools; For anonymity pseudonyms are used for participant and school names; BW: Black woman, WW: White woman, WM: White man

Table 3.3 Bell Middle School Team Assignment – Math Teacher

Teacher 1 Team A	Teacher 2 Team B	Teacher 3 Team C
2 AG Classes	1 AG Classes	2 AG Classes
2 Regular Classes	3 Regular Classes	1 Regular Classes
--	--	1 Inclusion or EC Classes

Note: AG = Academically Gifted; EC = Exceptional Children

Table 3.4 Burroughs Middle School Team Assignment – ELA Teacher

Teacher 1 Team A	Teacher 2 Team B	Teacher 3 Team C
1 AG Classes	1 AG Classes	1 AG Classes
2 Inclusion or EC Classes	2 Regular Classes[1]	2 Inclusion or EC Classes

Note: AG = Academically Gifted; EC = Exceptional Children; [1]
Includes English language learner students and a "lower" regular class

Table 3.5 Report of Teacher Assignment Practices by School

School	Federal and State Policies		Institutional Politics		Institutional Structures		
	Uses test scores in teacher reassignment	Use of NCEES in teacher reassignment	Collects teacher assignment preferences	Parental Influence in teacher assignment	Math/ELA - Homogeneous Groups	Science/Social Studies - Heterogeneous Groups	Multiple ability leveled course assignment
Bell Middle School	N	N	N	N	Y	Y	Y
DuBois Middle School	Y	N	Y	Y	Y	Y	Y
Burroughs Middle School	N	N	Y	N	Y	Y	Y
Bethune Middle School	N	N	N	N	Y	Y	Y
Note: The shaded rows represent high-performing and rural schools; non-shaded rows represent average-performing and city schools; For anonymity purposes, pseudonyms are used for participant and school names							

References: Chapter III

Abernathy, T.V., Forsyth, A., & Mitchell, J. (2001). The bridge from student to teacher: What principals, teacher education faculty, and students value in a teaching applicant. *Teacher Education Quarterly*, 28(4), 109-119.

Alvarez, D., & Mehan, H. (2006). Whole-school detracking: A strategy for equity and excellence. *Theory into Practice*, 45(1), 82-89.

Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, 30(1), 51-71.

Ballou, D. (1996). Do public schools hire the best applicants? *The Quarterly Journal of Economics*, 111(1), 97-133.

Balter, D., & Duncombe, W. (2005). Teacher Hiring Practices in New York State School Districts. Report prepared for the Education Finance Research Consortium.

Boe, E. E., & Gilford, D. M. (Eds.). (1992). *Teacher supply, demand, and quality: Policy issues, models, and data bases*. National Academies Press.

Boody, R.M. (2009). Career services perspectives on the use of portfolios in the teacher employment process: A survey. *Education*, 130(1), 67-71.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management*, *24*(1), 113-132.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, *31*(1), 83-117.

Boyd, D., Lankford, H., Wyckoff, J. (2008). Increasing the effectiveness of teachers in low-performing schools. In H. F. Ladd & E. B. Fiske (Ed.), Handbook of Research in Education Finance and Policy (pp.146-165). New York: Routledge.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management*, 30(1), 88-110.

Carey, N., Farris, E., & Carpenter, J. (1994). Curriculum differentiation in public high schools. National Center for Education Statistics, 95-360, Washington, D.C.

Chingos, M.M., & West, M.R. (2011). Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? *Economics of Education Review*, 30(3), 419-433.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377-392.

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

Clotfelter, C.T., Glennie, E., Ladd, H.F., & Vigdor, J.L. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, 92(5), 1352-1370.

Cohen-Vogel, L. (2011). "Staffing to the test": Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis*, 33(4), 483-505.

Cohen-Vogel, L., & Osborne-Lampkin, L. (2007). Allocating quality: Collective bargaining agreements and administrative discretion over teacher assignment. *Educational Administration Quarterly*, *43*(4), 433-461.

Cohen-Vogel, L., Osborne-Lampkin, L., & Houck, E. (2013). New data, old patterns: The role of test scores in student assignment. In D. Anagnostopoulos, S.A. Rutledge, and R. Jacobsen (Eds.), The Infrastructure of Accountability: Mapping Data Use and its Consequences. Cambridge, MA: Harvard Education Press.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, *39*(3), 124-130.

DeAngelis, K.J., Presley, J.B., & White, B.R. (2005). The distribution of teachers in Illinois. Illinois Education Research Council. Policy Report IERC 2005-1.

Dolton, P. J., & Makepeace, G. H. (1993). Female labour force participation and the choice of occupation: The supply of teachers. *European Economic Review*, *37*(7), 1393-1411.

Donaldson, M. L., & Johnson, S. M. (2010). The price of misassignment. *Educational Evaluation and Policy Analysis*, 32(2), 299-323.

George, A. L. (1979). Case Studies and Theory Development: The Method of Structured Focused Comparisons,". *Diplomacy: New Approaches in History, Theory and Policy, New York: Free Press*, *1979*, 43-68.

Glasman, N.S., & Heck, R.H. (1987). Evaluation in decision making: The case of assigning teachers to classrooms. *Administrator's Notebook,* 32(5), 1-4

Hanushek, E.A., Kain, J.F., O'Brien, D.M., & Rivkin, S.G. (2005). The market for teacher quality (Working Paper No. 11154). Cambridge, MA: National Bureau for Economic Research.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39(2), 326–354.

Hanushek, E.A., & Rivkin, S.G. (2010). The Quality and distribution of teachers under No Child Left Behind Act. *Journal of Economic Perspectives,* 24(3), 133-150.

Heck, R.H., Marcoulides, G.A., & Glasman, N.S. (1989). The application of causal modeling techniques to administrative decision making: The case of teacher allocation. *Educational Administration Quarterly,* 25(3), 253-267.

Horvat, E.M., Weininger, E.B., & Lareau, A. (2003). From Social Ties to Social Capital: Class Differences in the Relations Between Schools and Parent Networks. *American Educational Research Journal*, 40(2), 319-351.

Ingersoll, R. (2002). Out-of-field teaching, educational inequality, and the organization of schools. Center for the Study of Teaching and Policy.

Ingersoll, R., & Merrill, L., & Stuckey, D. (2014). Seven trends: The transformation of the teaching force, updated April 2014. CPRE Report (#RR-80). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

Johnson, S., Kardos, S., Kauffman, D., Liu, E., & Donaldson, M. (2004). The support gap: New teachers' early experiences in high-income and low-income schools. *Education Policy Analysis Archives*, 12(61).

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, 40(3), 581-617.

Kelly, S. (2004). An event history analysis of teacher attrition: Salary, teacher tracking, and socially disadvantaged schools. *Journal of Experimental Education,* 72(3), 195-220.

Kelly, S. (2004b). Are teachers tracked? On what basis and with what consequences. *Social Psychology of Education,* 7, 55-72.

Ladd, H. F. (2011). Teachers' Perceptions of Their Working Conditions How Predictive of Planned and Actual Teacher Movement?. *Educational Evaluation and Policy Analysis*, *33*(2), 235-261.

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

Levin, J., & Quinn, M. (2003). Missed opportunities: How we keep high-quality teachers out of urban classrooms. New York: The New Teacher Project.

Liu, E., & Moore Johnson, S. (2006). New teachers' experiences of hiring: Late rushed, and information-poor. *Educational Administration Quarterly,* 42(3), 324-360.

Loeb, S., Soland, J., & Fox, L. (Forthcoming). Is a Good Teacher a Good Teacher for All? Comparing Value-Added of Teachers with Their English Learners and Non-English Learners. Education Evaluation and Policy Analysis.

Martin, R., & Williams, J. (2012). " I Feel I'm Important": Successful Collaborative Teaching and Learning in a New Zealand Intermediate School. *Research in Middle Level Education Online*, *36*(2).

Mason, R. W., & Schroeder, M. P. (2010). Principal hiring practices: Toward a reduction of uncertainty. *The Clearing House*, 83(5), 186-193.

McGrath, D.J., & Kuriloff, P.J. (1999). "They're Going to Tear the Doors Off this Place": Upper-Middle-Class Parent School Involvement and the Educational Opportunities of Other People's Children. *Educational Policy*, 13, 603-629

McGrath, D., & Kuriloff, P. (1999b). The Perils of Parent Involvement: Tracking, Curriculum, and Resource Distortions in a Middle School Mathematics Program. *Research in Middle Level Education Quarterly*, 22(3), 59-83.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Mickelson, R. A. (2001). Subverting Swann: First-and second-generation segregation in the Charlotte-Mecklenburg schools. *American Educational Research Journal*, *38*(2), 215-252.

Mickelson, R., & Everett, B. (2008). Neotracking in North Carolina: How high school courses of study reproduce race and class-based stratification. *The Teachers College Record*, *110*(3), 535-570.

Neubig, M. (2006). Essential scheduling practices for high-performing schools with career academies/SLCs. *Techniques*, *81*, 42-43.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher*, 12-21.

Oakes, J. (2005). *Keeping track: How schools structure inequality*. Yale University Press.

Parker, C., Goe, L., Hicks, B., & McCreadie, J. (2007). Factors that contribute to students' equitable access to highly qualified teachers at the secondary level in the state of Delaware. Prepared for the Delaware Department of Education in collaboration with the Mid-Atlantic Comprehensive Center, The National Comprehensive Center for Teacher Quality and the Regional Education Laboratory Mid-Atlantic.

Radcliffe, R.A., & Mandeville, T.F. (2007). Teacher preferences for middle grades Insights into attracting teacher candidates. The Clearing House, 80(6), 261-266.

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

Reed, D. K., & Groth, C. (2009). Academic teams promote cross-curricular applications that improve learning outcomes. *Middle School Journal*, *40*(3), 12-19.

Rutledge, S.A., Harris, D., & Ingle, W.K. (2010). How principals "bridge and buffer" the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring. *American Journal of Education*, 116(2), 211-242.

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

Scafidi, B., Sjoquist, D.L., & Stinebrickner, T.R. (2007). Race, poverty and teacher mobility. *Economics of Education Review,* 26(2), 145-159.

Schweig, J. (2013). Cross-Level Measurement Invariance in School and Classroom Environment Surveys Implications for Policy and Practice. *Educational Evaluation and Policy Analysis*.

Southworth, S., & Mickelson, R. A. (2007). The interactive effects of race, gender and school composition on college track placement. *Social Forces*, *86*(2), 497-523.

Stearns, C. J. (1999). A middle school venture into cooperative learning: Successes and dilemmas. *Theory into Practice*, *38*(2), 100-104.

Talbert, J.E. & Ennis, M. (1990). Teacher tracking: Exacerbating inequalities in high school. Stanford, CA: Stanford University, Center for Research on the Context of Secondary Teaching.

Tinker, C., & Armstrong, N. (2008). From the outsider looking in: How an awareness of difference can benefit the qualitative research process. *The Qualitative Report,* 13(1), 53-60.

U.S. Department of Education, Office of Postsecondary Education (2011). Teacher shortage areas nationwide listing 1990-91 thru 2011-12, 1840-0595, Washington, D.C.

Watanabe, M. (2008). Tracking in the era of high stakes state accountability reform: Case studies of classroom instruction in North Carolina. *Teachers College Record*. 110(3), 489-534.

Wells, A. S., & Serna, I. (1996). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review*, *66*(1), 93-119.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage.

Appendix A: Teacher Interview Protocol

Interviewee_____ Date_____

Subject Taught_____ Grade Taught_____

School_____ District_____

Introductory Protocol

To facilitate note-taking, I would like to audio tape our conversations today. Please sign the release form. For your information, only I will be privy to the tapes, which will be eventually destroyed after they are transcribed. In addition, you must sign a form devised to meet our human subject requirements. Essentially, this document states that: (1) all information will be held confidential and your participation is voluntary. Thank you for your agreeing to participate.

I have planned this interview to last no longer than one hour. During this time, I have several questions that I would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.
Introduction
You have been selected to speak with me today because you have a great deal of knowledge about the hiring and assignment process of new teachers. My study does not aim to evaluate your classroom instructional practices. Rather, I am trying to learn more about the placement and assignment of new teachers more generally, and hopefully learn about how this process influences student performance.

A. General Background Questions
Let's begin by telling me a little about yourself.
1. How long have you been a teacher?
2. How long have you been a teacher at _____ school?
3. Talk about what inspired you to become a teacher?
    a. Why did you decide to teach middle grades?
4. What grade(s) and subject(s) do you teach?
    a. Are you licensed to teach this/these subject(s)?
5. How would you describe _____school?
6. What kind of relationships do you have with the principal and APs?
    a. How was this relationships developed?

B. Questions about Vacancies
    Let's begin by talking about job openings
1. When did you begin looking for job openings?
2. In your search, did you notice that some districts had more openings than others?
    a. If so, how were these schools similar? Different?
3. How did you find out about job openings?

4. Aside from formal job postings, did you receive any informal advertising for open positions?
5. Where there any positions available at the school where you completed your student teaching?
    a. If so, did you apply to that school?
6. What things were you looking for in a teaching position? What would be your ideal position?
7. What things were you looking for in a school?
8. Are there things about certain schools that made them unattractive?
9. What things made ____school particularly attractive?
10. Did you receive any type of incentive?
11. About how many schools did you apply to?
12. [If out of state] What influenced you to apply to jobs out of your home state? Particularly, North Carolina?

C. Questions about the Application Process
Now I want to get a general sense the application process.
1. About how many applications did you submit?
2. Tell me about the schools you applied to?
    a. Student composition (socio-demographics, test performance, etc.)
    b. Teacher composition
    c. Location (urban, rural)
3. What type of application materials did you have to submit to the district/school?
4. What do you think made you stand out?
    a. Probe: advanced degree, college prestige, referrals, academic performance, etc.
5. Who is primarily responsible for screening the applications?
6. What is the typical time between submitting the application and being notified?

D. Questions about the Interview Process
Now I'd like to ask you about the actual interview process
1. About how many interviews did you have during your job search?
2. Walk me through a typical interview?
    a. Who was typically involved in the interview process? District HR, Principal, AP, other teachers
    b. On average, how long did the interview last?
    c. What materials were you asked to bring to the interview (i.e. portfolios, lesson plans, etc.)
    d. What did you do with these materials?
    e. What topics were covered during the interview?
    f. Were behavioral-based interviewing techniques used (i.e. "tell me about a time when you…")
    g. Were you asked to perform any task, such as teach a sample lesson?
3. What did you hope to learn about the position? School?
4. During the interview process, what kinds of signals were you looking for to determine whether the school would be a good fit?

E. Questions about School Placement
      1. How many job offers did you receive?
      2. [If more than one] How did you decide between the offers?
      3. What challenges did you have
            a. finding jobs?
            b. The application process?
            c. Interviewing for jobs?

F. Questions about the Assignment Process
Now I'd like to ask you a few questions about how teachers are assigned to classrooms?
      1. Tell me about the way your classes are structured? Length of classes (Block schedule), levels (tracking), etc.
      2. How are teachers assigned to classes? [BTs]
            a. Probe: Current course load, leadership roles held, use of EVASS or McREL data, rigor
      3. Teacher's influence in schools (listen for issues regarding seniority, retention, social networks)
            b. In general, how would you describe your relationship with the school leaders at ____school?
            c. As a new teacher, how did you build relationships with school leaders, staff, and other teachers?
            d. How much influence do you think teachers, have over the assignment process?
      4. Parent's influence in schools
            a. What do you think parents' role in schooling should be?
            b. How much influence do you think parents, have over the teacher assignment process?
The next questions pertain your experiences during your first year only.
      1. As a beginning teacher, what was the process used to assign you to a specific grade, subject, and level in your school?
      2. When were you made aware of your assignment? Who informed you of your assignment?
      3. Did you discuss the assignment with your school leader?
      4. Overall, how satisfied were you with your assignment during your first year?
      5. What would be your ideal assignment?
      6. Were there any changes to your assignment this school year?
            1. Are you satisfied with the current assignment?

G. Final Question
      a. Is there anything else that you would like to talk about that we haven't covered about the hiring process and assignment process of new teachers?
Well those are all my questions. Thank you for your time!

Appendix B: School Leaders Interview Protocol

Interviewee_____Date_____Title
_____Years of Service_____

School_____District_____

Introductory Protocol

To facilitate note-taking, I would like to audio tape our conversations today. Please sign the release form. For your information, only I will be privy to the tapes, which will be eventually destroyed after they are transcribed. In addition, you must sign a form devised to meet our human subject requirements. Essentially, this document states that: (1) all information will be held confidential and your participation is voluntary. Thank you for your agreeing to participate.

I have planned this interview to last no longer than one hour. During this time, I have several questions that I would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

Introduction

You have been selected to speak with me today because you have a great deal of knowledge about the hiring and assignment process of new teachers. My study does not aim to evaluate your classroom instructional practices. Rather, I am trying to learn more about the placement and assignment of new teachers more generally, and hopefully learn about how this process influences student performance.

    A. General Background Questions

       Let's begin by telling me a little about yourself

       1. How long have you been a principal?

       2. How long have you been a principal at _____ school?

         1. Where were you before as a principal?

       3. Talk about how you became a principal? What inspired you?

       4. How would you describe this school?

       5. What kind of relationships do you have with the teachers?

         1. How build relationships with new teachers?

           a. Ho: white principals tend to develop better relationships with white teachers which influence assignment decisions

    B. Questions about Vacancies

       Let's begin by talking about job openings

       1. When do you know that you'll have to hire a teacher to fill a vacancy in the next school year?

         a. Probe for reasons for the vacancy (i.e. increased student enrollment, dismissal, maternity leave, transfer, retirement, accountability pressures, etc.)

       2. When do you notify the district about the vacancy?

       3. How do you recruit teachers? Tell me about that process.

         a. Are there challenges?

       4. When do official efforts to recruit a teacher usually begin?

         a. Do you have to do much to attract teachers?

5. Are there student teachers at your school?
   a. How often do you extend offers to student teachers if there meet vacancy qualifications?
6. Aside from formal job postings, do you do any formal or informal advertising for open positions? (i.e., targeted hires)
7. What attracts teachers to this school?
8. Are there things about your school that makes it harder to attract candidates?
9. Do you provide any type of incentive to attract candidates?

C. Questions about the Application Process
Now I want to get a general sense of what you look for in a candidate, especially new teacher candidates.
   7. On average, how many applications do you receive for a job opening?
   8. What type of application materials do candidates have to submit to the district/school?
   9. [If district] Do you receive a copy of the materials submitted to central office?
   10. What makes a candidate stand out to you?
      a. Probe: advanced degree, college prestige, referrals, academic performance, etc.
   11. Where do you get your best candidates (i.e. out of state, student teachers, other schools etc.)?
   12. Tell me about the application screening process.
      a. Who's involved?
      b. What criteria for determining which candidates to interview?
   13. What is the typical time between reviewing the applications and notifying potential candidates?

D. Questions about the Interview Process
Now I'd like to ask you about the actual interview process
   5. What do you hope to get out of the interview or learn?
      a. During the interview process, what kinds of signals are you looking for to determine whether the candidates will be a good fit for this school?
   6. Walk me through a typical interview.
      a. How long are the interviews?
      b. What materials do you ask the candidates to bring to the interview (i.e. portfolios, lesson plans, etc.)
         i. What do you do with these materials?
      c. What topics do your interview questions cover?
      d. Do you use behavioral-based interviewing techniques (i.e. "tell me about a time when you…")
      e. Do you use a scorecard or rubric to rate the candidates?
      f. Do you ask the candidates to perform any task, such as teach a sample lesson?
      g. Do you look for different qualities from new teacher candidates than candidates with prior experience?
      h. What happens after the interview is over?

        i.   What process do you use to decide who to hire?

    7. As you work to hire teachers for the school, are there any other schools competing to get the candidates you want?

        a. How do you handle this competition?

E. Questions about the Quality of Applicants

    a. When you think about some of you most effective teachers at this school, what characteristics do/did they have that made them effective?

The next few questions are about the quality of the applicants that you get here at ____*school.*

    1. Overall, how would you rate the quality of the applicants that you get at ____school?

        a. Why do you think that is? (listen for timing, budget, school characteristics)

    2. How many offers do you make before you find a candidate who accepts?

    3. What do you do when you are unable to find a candidate to fill a position?

    4. Is there anything that gets in the way of your being able to hire teachers?

        a. Hire effective teachers?

F. Questions about the Assignment Process

    1. But before we go into the assignment process, can you tell me about the way your classes are structured? Length of classes (Block schedule), levels (tracking), etc.

    2. How are teachers assigned to classes?

    3. How would you describe your staffing philosophy?

        a. Listen for: equity, accountability (test scores), student ability

    4. Tell me about how you make assignment decisions. Who's involved in the process?

        a. Probe: Current course load, leadership roles held, use of EVASS or McREL data, rigor

    5. Are there any characteristics that you look for when assigning teachers to certain grades, subjects or levels?

    6. Do you feel that you have the flexibility you need to make good staffing decisions?

Teacher's influence in schools (listen for issues regarding seniority, retention, social networks)

        a. How much influence do you think teachers, have over the assignment process?

            1. Does this influence affect your decision-making ability?

       ii. *Parent's influence in schools*

            1. What do you think parents' role in schooling should be?

            2. How much influence do you think parents, have over the teacher assignment process?

            3. Does this influence affect your decision-making ability?

The next questions pertain to first year teachers only.

    1. For first year teachers without a prior teaching record, what is the process through which they are assigned to grades or subjects in your school?

    2. What is a typical course load for a first year teacher?

    3. Are first year teachers allowed to teach advanced courses, such as Algebra 1?

4. In terms of first year teachers, how do you determine if changes need to be made in their assignment for the next school year?
5. How do vacancies influence teacher assignments?

7. Final Question
   a. When you were a teacher how were you evaluated? What did you think about the process?
   b. Is there anything else that you would like to talk about that we haven't covered about the hiring process and assignment process of new teachers?

Well those are all my questions. Thank you for your time!

Appendix C: Glossary of Pseudonyms

<u>**Pseudonyms of the Schools in the Study (in alphabetical order)**</u>

Derrick Bell (1930-2011). After his military service as a lieutenant in the United States Air Force, Bell entered law school at the University of Pittsburgh School of Law, where he was the only Black student in his class of 140, and only one of three Black students in the school. In 1969, Bell joined the faculty of Harvard Law School and later became the first Black tenured professor on the faculty of the law school.

During his academic career, Derrick wrote prolifically, integrating legal scholarship with parables, allegories, and personal reflections that illuminated some of America's most profound inequalities, particularly around the pervasive racism permeating and characterizing much of American law and society. Bell is often credited as a founder of Critical Race Theory, a school of thought and scholarship that critically engages questions of race and racism in the law, investigating how even those legal institutions purporting to remedy racism can more profoundly entrench it. Source: http://professorderrickbell.com/about/

Mary Mcleod Bethune (1875-1955). Mary McLeod Bethune was an extraordinary educator, civil rights leader, and government official who founded the National Council of Negro Women and Bethune-Cookman College. Bethune's background as a teacher inspired her to open the Daytona Educational and Industrial Training School for Negro Girls in Daytona Beach, Florida in 1904. The school became the co-educational Bethune-Cookman College in 1929 after merging with Cookman Institute and was fully accredited in 1943. Source: http://www.ncnw.org/about/bethune.htm

Nannie Helen Burroughs (1879-1961). Burroughs formed women's industrial clubs throughout the South teaching night classes in typing, stenography, bookkeeping, millinery, and home

economics to Black women. Through her powerful oratory she became secretary of the National Baptist Woman's Convention and, building on her teaching experience and grassroots network among Baptist women, she founded the National Training School for Women and Girls. She maintained her own publishing house, trained women missionaries, and educated African American women to be self-sufficient wage earners. She was a power player among both Black and White women. Source: https://www.nwhm.org/education-resources/biography/biographies/nannie-helen-burroughs/

William Edward Burghardt (WEB) DuBois (1868 – 1963) Scholar and activist W.E.B. Du Bois was born in Great Barrington, Massachusetts. After earning his bachelor's degree at Fisk, Du Bois entered Harvard University, where he became the first African American to earn a Ph.D. from Harvard University. Du Bois wrote extensively and was the best known spokesperson for African-American rights during the first half of the 20th century. He co-founded the National Association for the Advancement of Colored People in 1909. Du Bois's life and work were an inseparable mixture of scholarship, protest activity, and polemics. All of his efforts were geared toward gaining equal treatment for Black people and toward marshaling and presenting evidence to refute the myths of racial inferiority. Source: http://www.biography.com/people/web-du-bois-9279924 and http://www.naacp.org/pages/naacp-history-w.e.b.-dubois

Pseudonyms of the Participants in the Study (in alphabetical order)

Hallie Quinn Brown (1850-1949). Teacher, writer and women's activist Hallie Quinn Brown attended Wilberforce College and received a degree in 1873. She then taught in freedman's schools in Mississippi before moving to Columbia, South Carolina in 1875 where she served briefly as an instructor in the city's public schools. She later joined the faculty at Allen

University and later served as Dean of the University. Brown also served as Dean of Women at Tuskegee Institute before returning to Ohio where she taught in the Dayton public schools. Throughout her career, Brown published four significant works during her lifetime – Bits and Odds: A Choice Selection of Recitations, "Elocution," "Physical Culture," and "Homespun Heroines and Other Women of Distinction." Source: http://www.blackpast.org/aah/brown-hallie-quinn-1850-1949#sthash.I46jkn7y.dpuf

Octavius Catto (1839-1871). Born in pre-Emancipation South Carolina, Catto moved North with his father (a former slave) and eventually became a renowned educator at the elite Institute for Colored Youth in Philadelphia. During the Civil War, he raised 11 regiments of African American volunteers, rising to the rank of major in the U.S. Army. Catto campaigned aggressively for the desegregation of the transportation network by sitting on the streetcars and refusing to move. The campaign was successful. In 1867, a lawsuit by Le Count forced the city to enforce a newly passed state law desegregating Philadelphia's streetcars. Catto's work in defense of freedom was validated by the ratification of the 15th Amendment in February 1871, which guaranteed African Americans the right to vote. But the 1871 Philadelphia mayoral election—the first since the 15th Amendment's passage—was marred by mob violence, as opponents tried to prevent African Americans from exercising their franchise. On his way back from the polls, Catto, who had spearheaded a get-out-the-vote drive for Black voters, was shot in the back by a political opponent. Source: http://blog.constitutioncenter.org/2012/02/octavius-catto-philadelphia%E2%80%99s-%E2%80%9Cforgotten-hero%E2%80%9D/

Septima Poinsette Clark (1889-1987). Septima Poinsette Clark was born on in Charleston, South Carolina. Clark was qualified as a teacher; however, Charleston did not hire African Americans to teach in its public schools. Instead, she became an instructor on South Carolina's Johns Island.

In 1919, Clark returned to Charleston to teach at the Avery Institute. She also joined with the NAACP in trying to get the city to hire African-American teachers. By gathering signatures in favor of the change, Clark helped ensure that the effort was successful. Clark worked with the NAACP and Thurgood Marshall on a case that sought equal pay for Black and White teachers. She described it as her "first effort in a social action challenging the status quo." Her salary increased threefold when the case was won. Clark later became the director of the Highlander's Citizenship School program. These schools taught basic literacy and math skills to African Americans. One particular benefit of this teaching was that more people were then able to register to vote (at the time, many states used literacy tests to disenfranchise African Americans). Clark then joined the SCLC as its director of education and teaching. Under her leadership, more than 800 citizenship schools were created. Source: http://www.biography.com/people/septima-poinsette-clark-38174#teaching-and-early-activism&

Edna Meade Colson (1888-1985). Educator Edna Meade Colson struggled to make it easier for African Americans to obtain high quality education in Virginia. Colson received a B.A. from Fisk University in 1915, and she became an assistant in pedagogy at the Virginia Normal and Industrial Institute. Four years later she was appointed director of the new department of education. Colson guided the program as it grew into the school of education early in the 1950s. In the meantime, she attended Teachers College, Columbia University, and received a Ph.D. in 1940. Source: http://www.lva.virginia.gov/public/trailblazers/2008/index.htm?id=4

Anna Julia Haywood Cooper (1858-1964). Anna Julia Haywood Cooper spent her life redefining the limitations and opportunities for women of color in a society. A distinguished scholar and educator, Cooper saw the status and agency of Black women as central to the equality and progress of the nation. Describing her own vocation as "the education of neglected

196

people," Cooper saw education, and specifically higher education, as the means of Black women's advancement. Her accomplishments and vision have helped not only make Cooper one of the most noted African-American intellectuals in the history of the nation, but have helped reframe the understanding of intersections of race and gender and their political, cultural and personal implications in pursuit of a better nation. Source: http://cooperproject.org/about-anna-julia-cooper/

Jennie Serepta Dean (1848-1913). Jennie Serepta Dean founded the Manassas Industrial School for Colored Youth. A former slave, Dean attended schools in Fairfax County and Washington, D.C., and in 1878 began to establish a series of Sunday schools. She was a skilled fund-raiser, securing money from African American and White donors in Virginia and in northern cities to support her plan to open a school that would teach skilled trades to young African Americans. The Manassas Industrial School for Colored Youth opened in 1894 after nearly six years of fundraising. Dean served on the school's board of directors and executive committee. Source: http://www.encyclopediavirginia.org/Dean_Jennie_Serepta_1848-1913#start_entry

Lucy Craft Laney (1854-1933). Lucy Craft Laney was an educator, school founder, and civil rights activist. At the age of fifteen, she joined Atlanta University's first class where she graduated from the teacher's training program at the University. After teaching for ten years in Macon, Savannah, Milledgeville, and Augusta, she opened her own school, Haines Institute, in the basement of Christ Presbyterian Church in Augusta, Georgia in 1883. Originally intended only for girls, when several boys appeared she accepted them as pupils. In the 1890s, the Haines Institute was the first school to offer a kindergarten class for African American children in Georgia. By 1912 it employed thirty-four teachers, and had over nine hundred students enrolled.

Among the graduates of Haines Institute were Mary McLeod Bethune and Nannie Helen Burroughs.

In Augusta in 1918, Lucy Laney helped to found the Augusta branch of the National Association for the Advancement of Colored People (NAACP). She was also active in the Interracial Commission, the National Association of Colored Women, and the Niagara Movement. Laney helped to integrate the community work of the YMCA and YWCA. Source: http://www.blackpast.org/aah/laney-lucy-craft-1854-1933#sthash.dOndhUYn.dpuf

Joseph Charles Price (1854-1893). Joseph Charles Price, studied at St. Cyprian Episcopal School founded for the children of ex-slaves by Boston educators. He later attended Shaw University in Raleigh, but transferred to Lincoln University in Pennsylvania. Price graduated as valedictorian after winning several oratorical prizes. Price went to London, England, to represent the African Methodist Episcopal Zion Church at the Ecumenical Conference. He stayed in England for a year and raised about $10,000 for the founding of Zion Wesley Institute (now Livingstone College). Source: http://www.blackpast.org/aah/price-joseph-charles-1854-1893#sthash.NJlF6y1O.dpuf

Bazoline Estelle Usher (1885-1992). Usher served for 50 years as a professional educator, she rose through the ranks of the public school system to be supervisor of Black education initiating seven new elementary schools. Usher was also responsible for beginning the first African American Girls Scout troop. Source: http://www.atlantaintownpaper.com/2014/03/look-back-month-atlanta-history-4/ and *http://www.georgiawomen.org/2014/04/usher-bazoline/#sthash.v9KKP6zp.dpuf*

Ethel Carr Watson upon graduation from Sumner School, Watson entered Howard University. While at Howard, Watson became one of the founding members of Delta Sigma Theta sorority. After graduating from Howard University, she entered the teaching profession as a fifth and sixth grade teacher. In addition to regular classroom instruction, Watson also taught acrobatics and classical ballet. After over thirty years in the teaching profession, she retired in 1948 and began a second career as a dramatic performer. Source: Parks, G. (Ed.). (2008). *Black Greek-letter organizations in the twenty-first century: our fight has just begun*. University Press of Kentucky and Leavengood, B. (2002). Wood County, West Virginia. Arcadia Publishing. Chicago.