A MULTILEVEL STRUCTURAL ANALYSIS OF PREDICTORS OF URBAN TEACHER EFFECTIVENESS

by

Akisha R. Jones

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Education

Spring 2015

A MULTILEVEL STRUCTURAL ANALYSIS OF PREDICTORS OF URBAN TEACHER EFFECTIVENESS

by

Akisha R. Jones

Approved: ____
Ralph Ferretti, Ph.D.
Director of the School of Education

Approved: ____
Lynn Okagaki, Ph.D.
Dean of the College of College of Education and Human Development

Approved: ____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Elizabeth Farley-Ripple, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
William Jordan, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Jean-Philippe Laurenceau, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____
Henry May, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Carol Vukelich, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Thank you first and foremost to my Lord and Savior Jesus Christ through which all things are possible. "Now to Him who is able to do exceedingly abundantly above all that we ask or think, according to the power that works in us, to Him be glory in the church by Christ Jesus to all generations, forever and ever. Amen. (Ephesians 3-20-21)."

Sincerest and dearest thanks to my mother – my heart and my strength, the woman who birthed an undying desire to serve and protect her family, community, and history. Thank you for inspiring me to continue the work of building through improving education. This dissertation speaks to your truth as a Detroit teacher. To the woman who sacrificed everything so that we could have the desires of our hearts, I say thank you. This, here, is for you. Love always.

To my grandparents, James and Elizabeth Foster – Thank you for all of your support. You both are the source of inspiration for the work that I do. I am forever grateful for your love and dedication to our family. Thank you and I love you.

Thank you to my advisor, Dr. Farley-Ripple, who has guided and supported me throughout this journey. Thank you for your dedication and for taking a chance on me. I've learned so much from you, so thank you for helping me reach my dreams. It is with great sincerity that I say, thank you. To Dr. Henry May, thank you for being a great teacher. Thank you for seeing the fire in me and pushing me to my limits. To Dr. Will Jordan, Dr. Jean-Philippe Laurenceau, and Dr. Carol Vukelich – thank you for your support and dedication to this project.

To my sister and best friend, thank you for always being just a phone call away and listening to my many stories along this journey. I love you. Thank you to my family for keeping me grounded and focused and always being a source of strength and release. Thank you to the hosts of friends and colleagues who have supported me in many ways throughout this process. Thank you to my Canaan Baptist Church family for your prayers and spiritual support.

Finally, I dedicate this paper and journey to the city I learned to love very early on in life – Detroit, Michigan. You taught me that we all have a voice that should be heard, never silenced. You taught me to always fight the good fight. You sparked a fire in me to challenge the norms, sacrifice for those in need, and to reach for the stars – a fire that will never die.

PREFACE

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation under Grant #DRL-0941014. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A framework for teacher quality was developed to examine attributes of urban teachers in relation to their impact on student learning, understanding the role and value teachers play in improving student learning and the unique challenges of urban schools. This framework considers direct relationships between teacher attributes and teacher effects; mediation, which may help explain relationships between teacher attributes and teacher effects; and contextual effects, which consider the nested nature of educational data. Using data from the Measures of Effective Teaching (MET) study, I establish a sample of urban teachers from the six MET districts working in schools with at least both 60% minority and 60% low-income student populations. A multilevel structural modeling (MSEM) is employed to examine urban teacher characteristics and qualifications as predictors of teacher effects estimated using value-added modeling, with measures of pedagogical content knowledge and instructional practices as mediators. Using significant attributes resulting from MSEM analyses, I predict the impact on student learning of making specific improvements to the qualifications and characteristics of the urban teacher workforce.

Findings suggest that improving instructional quality among urban teachers is likely to have the largest impact on student learning. Teachers' pedagogical content knowledge in math was also found to have significant impacts, and teachers' race is associated with student performance in English Language Arts (ELA). Contextual effects were found in ELA with schools composed of teachers with advanced degrees or higher levels of instructional quality showing positive associations with student

learning. A significant mediating effect was also identified in which White teachers, as mediated by their pedagogical content knowledge, were found to have a significant positive impact on student learning in math. No evidence was found indicating that teachers' gender, advanced degree, or years of teaching experience in their district had a direct predictive relationship with student learning, in either subject.

Significant teacher effects resulting from this study can be used to inform policymakers' and practitioners' urban teacher hiring, development, and evaluation decisions, understanding that recruiting teachers with higher instructional quality and math pedagogical content knowledge may be critical. Teachers' race may also be important to recognize in ELA instruction, with minority teachers being more effective, suggesting that policymakers and practitioners should find ways to attract minority teachers for ELA instruction that have high levels of instructional quality. In relation to significant contextual effects, reformers must better understand and build upon the social dynamics between teachers that share these significant attributes in urban schools as influenced by such facilitators as communities of practice and teacher leadership, and their likeliness to influence student learning. This study can also be used to guide more focused research on urban schools and teachers, specifically, and guide future studies of teacher quality, which should consider contextual and mediating effects.

Policymakers can use significant findings, along with future research on urban teachers, to better select and prepare teachers for urban schools. Results from this study can inform national conversations on urban teacher and school reform, drive future research on urban teachers and schooling, and ultimately help to improve the quality of education for our students who need it most.

Chapter 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Detroit, Michigan, once a thriving urban city, now faces an array of social and political challenges (i.e., poverty, crime, political corruption), with the children of Detroit suffering the most. Like many other urban school systems, Detroit public schools have deteriorated; they struggle with school funding, poor working conditions, and challenges in attracting and retaining quality educators. These conditions negatively impact student achievement among Detroit students who continue to perform below state and national averages on Michigan and national assessments (Michigan Department of Education, 2013; NCES, 2013a). Researchers have determined that improving student learning relies heavily on improving the classroom teacher (Darling-Hammond, 2003; Hanushek, 1992; Hanushek, Kain, O'Brien, Rivkin, 2005; Jacob, A., 2012; Nye, Konstantopoulos & Hedges, 2004; Rice, 2003; Rivkin, Hanushek, & Kain, 2005).

It is estimated that the value of having a quality teacher, as measured by their ability to improve student performance, is greater than any other school input (Goldhaber, 2002) and can exceed one grade-level of achievement each year (Hanushek, 1992). The role of teachers is even more critical for urban schools, which typically serve underperforming, low-income, and minority students and where teachers face a variety of troubling factors that make teaching in urban schools especially challenging (Darling-Hammond, 2000, 2003; Hanushek, Kain & Rivkin, 2004; Ingersoll, 2003b). As education reformers continue to examine ways to improve teacher quality across all schools, they must consider the unique challenges that urban teachers face. With a specific focus on teachers in academically at-risk urban schools, this study examines characteristics of effective urban teachers, and the potential impacts of having more effective teachers in urban schools. The identification of attributes of effective urban teachers can help policymakers and educators become more strategic in their efforts to recruit, prepare, support, and strengthen the urban teacher workforce, which can in turn improve urban student achievement.

While studies of teacher quality have focused on various teacher and teaching attributes within the broader context of American schooling, this dissertation focuses exclusively on the quality of urban teachers. This study examines attributes of effective urban teachers teaching in the most academically at-risk urban schools using measures of teacher characteristics, qualifications, instructional quality, and pedagogical content knowledge from a sample of schools with at least 60% minority and 60% low-income student compositions from the six large districts in the Measures of Effective Teaching database. I include teacher demographics and qualifications along with measures of teachers' pedagogical content knowledge and instructional quality as predictors of teachers' value-added to student achievement, while focusing solely on urban teachers. I also examine these relationships as contextual or compositional effects at the school level, allowing us to understand the potential impact of having a quorum of quality teachers in urban schools, above and beyond the impact of individual teachers. Finally, I use predictive analyses to estimate how making specific improvements to the qualifications and characteristics of the urban teacher workforce might impact student learning. The following research questions guide this study.

1. Which attributes of teacher quality are predictive of urban teacher effectiveness?
2. What impact would increases in the number of quality teachers in urban schools, through changes in specific teacher attributes or practices, have on student learning?

This study contributes to the field by deepening our understanding of what constitutes an effective urban teacher and informs future research that may focus on improving the quality of teachers in urban schools. Results from this study may also be used to inform hiring and teacher development practices and policies, which may likely facilitate improvements in teachers' value-added to student learning and achievement. Furthermore, the methodological approaches used in this study improve upon prior methods used in studying teacher quality by including mediation analyses, multi-level structural equation modeling (MSEM), and teachers' value-added scores as outcomes.

1.2 Background

Teachers' contribution to student learning is well recognized, and accordingly, teacher quality continues to be the focus of many reforms and policies designed to improve student achievement in schools. With this focus, researchers continue to study what teacher attributes define a quality teacher. However, researchers mostly examined teacher quality without considering the unique challenges of urban schools. I argue that in order to improve student achievement in urban schools, researchers must examine attributes of teacher quality specifically among urban teachers. In the following sections, I discuss the unique challenges of urban communities and schools, how these challenges impact urban teachers and the quality of the urban teacher workforce, as well as provide a framework with which to define teacher quality.

1.2.1 Urban Communities and Schools

The term "urban" is commonly and loosely used to describe schools in large American cities with highly concentrated populations of poor and minority persons (Chou & Tozer, 2008; Eckert, 2013). Urban schools often exist in communities challenged with an array of complex issues including high rates of poverty, unemployment, crime, political corruption, and a lack of social capital (Eckert, 2013; Jacob, 2007; Ladson-Billings, 2006).

When looking at population demographics, data from the 2000 Census indicate that a larger proportion of minority and foreign-born populations live in urban communities compared to suburban and rural communities. Minorities represent the majority of urban populations, with more than half identifying as minority (52%) and one-quarter (27%) identifying as foreign born. In addition, urban families report having lower median family income compared to families in suburban communities and higher reports of violent crime compared to non-urban communities. Data comparing urban community characteristics to rural and suburban characteristics are presented in Table 1.1 below.

Table 1.1: Community Characteristics[a]

	United States	Urban	Suburban	Rural
Race/Ethnicity[a]				
White	75%	48%	81%	85%
Black	12%	25%	8%	9%
Native American	1%	1%	1%	2%
Asian	4%	9%	4.0%	1.0%
Native Hawaiian and Other Pacific Islander	0.1%	0.1%	0.0%	0.1%
Hispanic[b]	13%	26%	11%	6%
Immigration[a]				
Native Born	89%	73%	89%	97%
Foreign Born	11%	27%	11%	3%
Income/Employment[a]				
Unemployment Rate (Civilian Labor Force)	6%	8%	5%	6%
Poverty Rate (Families)	9%	14%	6%	11%
Median Family Income (dollars)	50,046	44,035	57,655	40,490
Crime[e]				
Violent Crime Rate per 100,000 Inhabitants	387	409	380	177
Property Crime Rate per 100,000 Inhabitants	2,859	2,950	3,535	1,539

[a] U.S. Census Bureau, 2000 Census Summary Files. Retrieved using "American Fact Finder" from http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

[b] Categorized as "Hispanic or Latino and Race". One can be Black and Hispanic as Hispanic is an indicator of ethnicity.

[c] Federal Bureau of Investigation (2012) *Crime in the United States by community type*. Retrieved from http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/2tabledatadecoverviewpdf.

Urban schooling is complicated by a myriad of factors including poverty, crime, limited parental involvement, inequities in school funding, and is impacted by the communities in which these schools exist (Darling-Hammond, 2007; Kozol, 1991; Noguera, 2003). These challenges are reflected in the achievement and outcomes of urban students (Darling-Hammond, 2004; Ladson-Billings, 2006; Lee, 2002) compared to their non-urban counterparts as presented in Table 2.2 (Golding, et al., 2013; NAEP, 2014; NCES, 2013).

These data show that urban schools have higher rates of minority, low-income, and limited English proficient or English Language Learners than suburban and rural schools. Data on student achievement and outcomes demonstrate that urban students underperform relative to their suburban and rural counterparts in math and reading in both 4th and 8th grades, as indicated by much lower rates of proficiency on the National Assessment of Educational Progress assessment (NAEP, 2014). In addition, graduation and college attendance rates among urban students are lower than in suburban and rural schools (NAEP, 2014).

Table 1.2: Student Characteristics by School Type

	All Public Schools	Urban	Suburban	Rural
Schools				
Number of Students[a]	49,508,800	14,152,560	16,215,460	13,433,240
Number of Schools	90,010	23,560	24,260	29,940
Average number of students per school	550	601	668	449
% of K-12 Students				
Free or Reduced Lunch	48%	61%	38%	45%
IEP	12%	12%	12%	12%
ELL/LEP	9%	15%	9%	5%
Title I Services	37%	50%	29%	33%
African American	15%	23%	13%	10%
Hispanic	22%	34%	21%	14%
Minority	46%	67%	44%	29%
Graduated with a high school diploma (2010-11)	89%	81%	86%	93%
Went on to a four-year college (2010-11)	39%	39%	42%	40%
Reaching Proficient or Above on 2013 4th Grade Math Assessment (NAEP)[b]	42%	36%	46%	44%
Reaching Proficient or Above on 2013 8th Grade Math Assessment	36%	31%	40%	36%
Reaching Proficient or Above on 2013 4th Grade Reading Assessment	35%	30%	39%	36%
Reaching Proficient or Above on 2013 8th Grade Reading Assessment	36%	32%	41%	41%

[a] National Center for Education Statistics (NCES) (2013b). *Schools and Staffing Survey (SASS).* Retrieved from http://nces.ed.gov/surveys/sass/index.asp.
[b] *National Assessment of Educational Progress.* U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved 3/19/2014 from http://nces.ed.gov/nationsreportcard/.

Researchers have proposed several explanations for differences between urban and non-urban schools, including poverty, school crime, a lack of parental support and school funding.

1.2.1.1 Poverty

Urban communities and schools are characterized by their high concentrations of low-income populations (Chou & Tozer, 2008; Eckert, 2013). Individual and family poverty is conflated with larger communal issues of poverty including the lack of availability of jobs and housing, higher rates of crime, an unstable economy, and a lack of local power and influence over schools (Noguera, 2003). Furthermore, "concentrated poverty severely limits the ability of communities to control and improve the quality of their schools" (Noguera, 2003). Researchers have commonly found that challenges in teaching and learning in urban schools, including developmental issues in children (Duncan, Brooks-Gunn & Klebanov, 1994; Leventhal & Brooks-Gunn, 2000) and poor student behavior (Duncan, Brooks-Gunn & Klebanov, 1994; Gregory, Skiba & Noguera, 2010; Raffael Mendez, 2003), are related to high rates of poverty.

1.2.1.1.1 Cognitive development

Among infants and adolescents, researchers have found that socioeconomic status (SES), using various measures including family or neighborhood income, parental education, or occupational status, is strongly positively correlated with cognitive development (Berger, et al., 2009; Bradley & Corwyn, 2002; Duncan, Brooks-Gunn & Klebanov, 1994). Lower levels of socioeconomic status have been associated with poorer cognitive development among youth, whereas higher SES children benefit from having access to more resources useful in supporting their positive developmental growth (Berger, et al., 2009; Bradley & Corwyn, 2002; Bornstein & Bradley, 2014; Duncan, Brooks-Gunn & Klebanov, 1994). This challenge is especially important to note in urban schools, which have higher rates of low-income students, and which, in turn, may result in higher rates of students with developmental challenges associated with low SES.

1.2.1.1.2 Student Behavior

Researchers have also associated poverty with challenging student behavior, which can create a difficult environment for learning and teaching (Duncan, Brooks-Gunn & Klebanov, 1994; Gregory, Skiba & Noguera, 2010). At the family level, income and poverty status has been strongly correlated with the behavior of students; increases in family income or higher SES status are associated with lower rates of student misbehavior (Duncan, Brooks-Gunn & Klebanov, 1994). Even when looking across neighborhoods, low-SES neighborhoods have been associated with adolescent juvenile delinquency and externalizing behavior problems (Duncan, Brooks-Gunn & Klebanov, 1994). Furthermore, low-income students with histories of low achievement who reside in high-crime/high-poverty neighborhoods may be at greater risk for engaging in behavior that results in office disciplinary referrals and school suspensions which are moderate to strong predictors of dropping out and not graduating on time (Gregory, Skiba & Noguera, 2010; Raffael Mendez, 2003). This is not to suggest that poverty causes these behaviors among students. For example, researchers also note that behavior is more criminalized in schools with high concentrations of poor and minority youth (Hirschfield, 2008; Laub, 2002; Wacquant, 2001). Nonetheless, research shows an association between poverty and student behavior documented in these contexts.

1.2.1.2 School Safety

Areas with a high incidence of crime and violence also tend to have schools that experience higher rates of violence and disorder (Noguera, 2003). Data from NCES (2011) comparing rates of crime in urban, suburban, and rural schools support this claim. In Table 1.3 below, data show that urban schools have higher rates of violent, seriously violent, theft, and all other school crime incidents than suburban and rural schools.

Table 1.3: School Crime Characteristics by School Type

Crime Types[a]	All Public Schools	Urban	Suburban	Rural
Violent Incidents	25.0	28.8	22.4	22.5
Serious Violent Incidents	1.1	1.3	1.0	1.1
Theft	5.5	6.2	4.9	5.3
Other Incidents	9.2	11.7	8.0	7.8

Note: Rate per 1,000 students
[a] Neiman, S. (2011). Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2009–10 (NCES 2011-320). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S.

The educational environment is affected by high rates of school and community violence in these low-income communities; both contribute to challenges in teaching and student learning (Smith & Smith, 2006). School violence affects students' and teachers' desire to attend school (Elliott, Hamburg, & Williams, 1998; Price & Everett, 1997), students' eagerness to participate or pay attention in class

(Price & Everett, 1997; Smith & Smith, 2006), as well as teachers' eagerness to challenge or discipline students (Price & Everett, 1997). Interviews with teachers who left urban schools describe these schools as violent places where their safety was threatened (Smith & Smith, 2006). Thus, violence within urban schools can influence the teaching experience.

1.2.1.3 Parental Involvement

According to researchers, teaching and student learning in urban schools can be more difficult as a result of a lack of parental support and involvement (Cooper & Crosnoe, 2007; Desimone, 1999; Fan & Chen, 2001; Jeynes, 2005, 2007), which may be a result of time limitations among low-income parents or a lack of financial resources or awareness (Williams & Sanchez, 2011). When compared to suburban and rural schools, urban schools have been found to have lower rates of parent involvement (Cooper & Crosnoe, 2007; Desimone, 1999; Fan & Chen, 2001; Jeynes, 2005, 2007), although rates of parental involvement have been found to vary by race or ethnicity, by socioeconomic status, and by support given at home or in children's school (Desimone, 1999).

1.2.1.4 Inequitable school funding

There remains an inequitable distribution of school funding, where urban schools do not have adequate funding to support the educational needs of the students they serve. While school funding is a function of the public school funding system in America where educational costs are the responsibility of the state and are primarily supported by local taxes along with state grants-in-aid (Baker & Corcoran, 2012; Crampton, 2010; Darling-Hammond, 2004), the system fails to account for the needs of students in urban communities which require greater funding and resources, creating inequities in school funding. As shown in Table 4 below, few disparities are actually shown in the distribution of funding across school types., However, urban schools have higher proportions of students at-risk of academic failure or failure to graduate from high school, which include students who are low-income, disabled, minority, English Language Learners, students in urban environments, or students with low parental education attainment (Baker & Duncombe, 2004; Crampton, 2010; Duncombe & Yinger, 2005; Land & Legters, 2002). Researchers have concluded that urban schools with high proportions of academically at-risk students require far more additional resources than non-urban schools with fewer proportions of at-risk students (Baker & Duncombe, 2004; Crampton, 2010; Duncombe & Yinger, 2005). Nevertheless, data show that urban student expenditures, while greater, are not significantly higher than their non-urban counterparts as is needed.

Inadequacies in school funding for urban schools can negatively impact class size, books, materials, computers, course offerings, and other necessary services that enable school quality (Darling-Hammond, 2004, 2007; Kozol, 1991; Jacob, 2007).

Table 1.4: Funding and Resources by School Type*

	All Public Schools	Urban	Suburban	Rural
School Funding/Resources[a]				
Expenditure per student	7,268	7,892	7,542	6,974
Share receiving Title 1 services	54.4	57.9	49.3	60.2
Share of schools with temporary buildings	31.7	37.7	34.4	21.3
Share of schools with teacher vacancies	73.7	75.4	76.9	66.6

Note: Rate per 1,000 students

[a] Condition of Education, 2004 (Per pupil expenditure) https://nces.ed.gov/pubs2004/2004077.pdf

1.2.2 Teaching in Urban Schools and the Urban Teacher Workforce

In the prior sections, I established that urban schools and communities differ significantly from non-urban schools, and why teaching in urban schools and communities are different and often times more difficult. Below, I discuss some of the reasons why teaching can be more challenging in urban schools, how these challenges affect staffing of teachers in urban schools, and how this results in an inequitable distribution of quality teachers across schools.

1.2.2.1 Staffing Challenges of Urban Schools

Teachers' lack of preparation to teach in the urban context impacts the staffing in urban schools (Abel & Sewell, 1999; Darling-Hammond, 2003). Once hired to teach in an urban school, teachers experience poorer working conditions (Abel & Sewell, 1999; Adamson & Darling-Hammond, 2012; Darling-Hammond, 2003, 2004; Hanushek et al., 2004; Jacob, B., 2007), lower pay (Adamson & Darling-Hammond, 2012; Darling-Hammond, 2003, 2004; Hanushek et al., 2004; Jacob, B., 2007), and a lack of professional support or ability to provide input and influence in schools (Adamson & Darling-Hammond, 2012; Darling-Hammond, 2003; Ingersoll, 2003a). In turn, teaching is made more difficult, creating challenges in the hiring and retaining of quality teachers (Adamson & Darling-Hammond, 2012; Darling-Hammond, 2003; Guin, 2004; Haberman & Rickards, 1990; Hanushek et al., 2004; Ingersoll, 2003b; Ingersoll & Merrill, 2010, 2014; Jacob, B. 2007).

1.2.2.1.1 Preparation to Teach in Urban Contexts

Urban schools are often culturally complex with high rates of academically at-risk students, making it challenging and difficult for some urban teachers to make connections and engage students in the learning process (Abel & Sewell, 1999; Darling-Hammond, 2003; Eckert 2013; Haberman, 1987, 1995; Jacob, B., 2007; Matsko and Hammerness, 2014; Stotko, et al., 2007). As a result, there is evidence that teachers prefer to teach in schools with higher-achieving students and fewer minority and low-income students (Boyd, et al., 2008; Hanushek, Kain, & Rivkin, 2004; Ingersoll & Merrill, 2010; Lankford, Loeb, & Wyckoff, 2005; Loeb, et al., 2005; Quartz, et al., 2004).

Teachers' decisions to leave urban schools are likely influenced by the lower academic performance of urban students (Boyd, Lankford, Loeb, & Wyckoff, 2005; Hanushek, et al., 2004), as urban teachers report that improving the learning and achievement of their students is important and rewarding and that they are disappointed when their students fail (Brunetti, 2001; Shann, 1998). Effective teachers, for instance, are more likely to stay in the classroom when they find success with their students (Goldhaber, Gross & Player, 2007).

High proportions of minorities in urban schools also have considerable influence on teachers' decisions to transition to different schools or leave the profession altogether (Hanushek, et al., 2004; Scafidi, et al., 2007). For instance, African American teachers are more likely to remain in schools where there are large numbers of African American students, whereas White teachers are more likely to leave (Hanushek, et al., 2004; Scafidi, et al., 2007).

Researchers attribute the lack of preparation of teachers to teach in urban contexts to the curriculum and structure of teacher preparation programs (Haberman, 1987, 1995; Stotko, et al., 2007). Whereas traditional teacher preparation programs believe that their students will be able to teach in all settings (Stotko, et al., 2007), some college preparation programs have faculty who fail to relay the practical applications of knowledge of teaching to urban contexts (Haberman, 1987, 1995; Stotko, et al., 2007). It is further suggested that preparation to teach in urban schools can only be done on the job, with a teacher or coach, a support network, or specific trainings (Haberman, 1987, 1995). Some teacher preparation programs have addressed this concern by focusing curriculum on teaching or providing student teaching opportunities in urban schools, which is shown to positively influence the experience of urban teaching (Anderson & Stillman, 2012; Eckert, 2013).

1.2.2.1.2 Working Conditions

The quality of working conditions also adds to the difficulties of teaching in urban schools (Darling-Hammond, 2003; Jacob, B., 2007; Ingersoll, 2001, 2003a, 2004; Johnson, et al., 2012; Loeb, Darling-Hammond, Luczak, 2005). Working conditions influence teachers' decisions to leave schools above and beyond that of the characteristics of students discussed prior (Johnson, et al., 2012). Urban teachers are most often dissatisfied with a lack of resources, support and recognition from the school administration, large class sizes, a lack of teacher influence over school and classroom decision-making, the many intrusions on classroom time, and inadequate time to prepare (Brunetti, 2001; Darling-Hammond, 2000, 2003; Haberman & Rickards, 1990; Ingersoll 2001, 2003a, 2004; Loeb, et al., 2005; Shann, 1998). Overall, teachers desire to work in a supportive environment where they feel a part of the decision-making process (Brunetti, 2001; Darling-Hammond, 2003; Ingersoll, 2001, 2004; Johnson, et al., 2012; Shann, 1998; Haberman & Rickards, 1990).

Other challenges related to urban teachers' working conditions include weak relationships, support or involvement from parents, and job security (Haberman & Rickards, 1990; Ingersoll, 2003a, 2004; Shann, 1998). Teachers in urban schools are often frustrated with the lack of parental participation in the affairs of their children and in the school community, making it more challenging for teachers to influence students' learning (Haberman & Rickards, 1990; Ingersoll, 2003a, 2004; Shann, 1998). Teachers also expressed concern with the security of their jobs, with frequent layoff notices, and where teacher performance is not considered in decisions about retaining or firing staff (Shann, 1998).

1.2.2.1.3 Teacher Wages

Teacher salaries are sometimes lower in urban schools, although they do not universally account for the more challenging task of working in urban schools (Adamson & Darling-Hammond, 2012; Darling-Hammond, 2003, 2007; Hanushek, et al., 2004; Ingersoll, 2004; Jacob, B., 2007). According to data from NCES, in the 2007-08 school year, the average base salary of the lowest paid urban teachers was $35,000 compared to $37,900 for suburban teachers. The base salary of the highest paid urban teachers was only $58,300 compared to $76,400 for suburban teachers—an $18,000 difference. Furthermore, Adamson and Darling-Hammond (2012) concluded in an elasticity analysis that increases in teacher salaries are associated with decreases in the proportion of inexperienced, undereducated, or un-credentialed teachers in schools. Lower pay along with more challenging working environments may cause many urban teachers to seek jobs in other professions or in higher paying districts with better working conditions (Darling-Hammond, 2003; Hanushek, et al., 2004; Ingersoll, 2004; Loeb, et al., 2005; Scafidi, et al., 2007).

As demonstrated through the literature, there are many challenges to teaching in urban schools including lower pay, poor working conditions, as well as a lack of preparation and experience in teaching students from diverse populations. These challenges reflect in teachers' decisions to leave urban schools, resulting in high rates of teacher attrition and mobility.

1.2.2.2 Teacher Turnover in Urban Schools

Teachers who face challenges of teaching in urban schools are much more likely to leave or transfer to other schools, contributing to higher rates of turnover in urban schools (Boyd, et al., 2008; Darling-Hammond, 2003; Hanushek, et al., 1999; Henke et al., 2000; Ingersoll, 2004; Ingersoll & Merrill, 2010; Lankford, et al., 2002). An NCES report on teacher attrition (2007) indicated that urban schools have the highest rate of teacher turnover (20%), compared to teacher turnover in suburban and rural schools (15%). This attrition creates, according to Jacob (2007), teacher "shortages" in urban districts, "where the number of effective teachers the district wants to employ is greater than the number of effective teachers who are willing and able work at a given salary (Jacob, 2007, p.6)"; ultimately, it is harder to hire and retain qualified teachers in urban schools (Ingersoll, 2004; Jacob, 2007). When faced with these shortages, districts often assign teachers outside of their fields of qualifications, hire uncertified or inexperienced teachers, use long-term substitutes, expand class sizes, or cancel course offerings – exacerbating problems in providing quality schooling to urban students (Darling-Hammond, 2004; Jacob, 2007; Murnane & Steele, 2007). As a result, teachers in urban schools tend to be slightly less prepared or lack experience in teaching than their non-urban peers (Anderson & Stillman, 2012; Coopersmith, 2009; Darling-Hammond, 2003; Eckert, 2013; Hanushek, et al., 1999; Ingersoll, 2001). A brief look at recent data from NCES (Coopersmith, 2009) in Table 1.5 demonstrates that urban schools have teachers with less teaching experience in their current schools and are less likely to be certified than non-urban teachers. Urban teachers also have fewer advanced degrees than their suburban peers.

More significantly, a review of NCES data (U.S. Department of Education, 2006) in Table 1.6, shows that urban schools face more challenges in filling teacher vacancies in key subjects (i.e., mathematics and science) and are more likely to fill those vacancies with a substitute, by hiring a less than fully-qualified teacher, or by expanding class sizes (U.S. Department of Education, 2006).

Table 1.5: Teacher Characteristics by School Type[a]

	All Public Schools	City	Suburban	Rural
Years of Teaching Experience (Overall)				
Less than 4 years	19%	20%	19%	18%
4-9 Years	28%	27%	31%	26%
10-14 Years	16%	16%	16%	17%
15 or More Years	37%	37%	33%	39%
Average Number of Years	13.0	12.9	12.4	13.4
Years of Teaching at Current School				
Less than 4 years	36%	39%	37%	35%
4-9 Years	32%	33%	34%	31%
10-14 Years	12%	12%	13%	12%
15 or More Years	19%	16%	17%	22%
Average Number of Years	8.4	7.6	8.0	9.0
Highest Degree Earned				
Less than Bachelor's	1%	1%	1%	1%
Bachelor's Degree	47%	46%	43%	53%
Master's Degree	45%	46%	48%	40%
Higher than a Master's Degree	7%	8%	9%	6%
Certification and Major in Main Assignment				
Total Certified	80%	77%	81%	82%
Major in Main Assignment	84%	82%	85%	81$
Certified with Major in Main Assignment	84%	82%	85%	81%
Base salary of lowest paid	34,000	35,000	37,900	32,500
Base salary of highest paid	60,400	58,300	76,400	54,500

[a] Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States 2003-04 Schools and Staffing Survey (all other data) http://nces.ed.gov/pubs2006/2006313.pdf

Table 1.6: Staffing Challenges[a]

	All Public Schools	City	Suburban	Rural
Teacher Turnover				
Stayers	84%	80%	85%	85%
Movers/Leavers	17%	20%	15%	15%
Vacancies				
Share of schools with teaching vacancies	72%	75%	77%	67%
Vacancies in specific fields				
Special Education	65%	71%	69%	60%
English/Language Arts	55%	65%	58%	49%
Social Studies	47%	53%	49%	44%
Computer Science	28%	31%	29%	19%
Math	54%	64%	56%	47%
Biology	42%	46%	43%	35%
Physical Sciences	36%	39%	36%	30%
English as a Second Language	33%	43%	35%	27%
Foreign Languages	42%	47%	43%	30%
Music or Art	46%	46%	46%	44%
Methods used to Cover Vacancies				
Fully-qualified teacher	72%	75%	77%	67%
Less than fully-qualified teacher	92%	91%	94%	92%
Long or short-term Substitute	17%	19%	14%	18%
Expanded Class Size	26%	42%	30%	18%
Added sections to other teachers' normal teaching loads	10%	11%	9%	10%
Assigned a teacher of another subject or grade-level to cover vacancy	10%	12%	9%	10%
Assigned an administrator or counselor to teach those classes	4%	2%	2%	3%

[a] U.S. Department of Education (2006), National Center for Education Statistics, Schools and Staffing Survey, 2003–04, Public School, BIA School, and Private School Data Files.

1.2.2.3 Inequitable Distribution of Quality Teachers

Challenges with hiring and retaining qualified teachers in urban schools result in an inequitable distribution of quality teachers across schools and districts, with urban schools employing larger numbers of less-qualified or ill-prepared teachers as measured by various indicators of teacher quality (Adamson & Darling-Hammond, 2012; Betts, Zau & Rice, 2003; Darling-Hammond, 2003, 2004, 2007, 2010; Eckert, 2013; Ingersoll & Merrill, 2010; Jacob, B., 2007; Murnane & Steele, 2007; Ingersoll, 2004; Lankford, et al., 2002). This inequitable distribution of teachers often results from teachers' decisions to leave teaching in urban schools due to the cultural complexity and academic challenges among urban students, as well as the inadequate salaries and poor working conditions found in urban schools (Murnane & Steele, 2007; Darling-Hammond, 2003, 2004, 2007, 2010; Lankford, et al., 2002; Ingersoll, 2004). Other systems-level factors may also have an influence including the sorting of teachers based on school characteristics or performance (Lankford, et al., 2002); efficiency in hiring practices (Jacob, 2007); differences in defining quality teachers (Jacob, 2007); or teachers' own preferences regarding whether to teach, what to teach, and where to teach (Hanushek, et al., 1999; Jacob, B., 2007). In addition, urban schools find it challenging to match the salaries, benefits, and resources offered by more affluent schools. Thus, they have difficulty competing for the available supply of quality teachers (Ingersoll, 2004; Jacob, 2007; Murnane & Steele, 2007). As a result, more often than not, urban students find themselves in classrooms with less-qualified or ill-prepared teachers (Adamson & Darling-Hammond, 2012; Buddin & Zamarro, 2009; Darling-Hammond, 2004, 2007, 2010; Desimone & Long, 2010; Eckert, 2013; Lankford, et al., 2002).

In the prior sections, I establish the challenging conditions of urban communities and the specific challenges urban teachers face in urban schools. These factors produce an inequitable distribution of qualified teachers across schools, significantly disadvantaging urban schools and students. Researchers have offered many explanations or suggestions for addressing these challenges and inequities in urban schools including improving social, economic, or schooling conditions, as well as educational policies, or practices (Ladson-Billings, 2006; Lee, 2002; Rothstein & Wilder, 2005). However, the most important school-based factor in improving student learning is the classroom teacher (Darling-Hammond, 2003; Hanushek, 1992; Hanushek, et al., 2005; Jacob, A., 2012; Nye, et al., 2004; Rice, 2003; Rivkin, et al., 2005). It is estimated that the value of having a quality teacher versus a bad teacher, as measured by their ability to improve student performance, can exceed one grade-level of achievement (Hanushek, 1992) and the impact of having a good teacher on student achievement is larger than any other school input (Goldhaber, 2002). Therefore, we must consider improving the quality of the urban teacher workforce in order to improve achievement and outcomes of urban students and schools. As we continue to try to make such improvements, a central issue for research, policy, and practice is in understanding what constitutes a "good" or "quality" teacher.

1.2.3 Teacher Quality

Teacher quality is a common term used in educational policy and research in which teacher related attributes are examined to help define how well teachers influence student learning. It should be noted that this common definition of teacher quality is not the only way that teacher quality can be examined as the practice of teaching is complex and results in many different outcomes of student learning and engagement. Nevertheless, this definition of teacher quality is used throughout this study as my core inquiry focuses on improving student learning.

Studies of teacher quality at the state, district, school, and individual student level have focused on various teacher and teaching attributes, including teacher demographics, attitudes, experience teaching, preparation, practices, subjects taught, and pedagogical content knowledge. Teacher quality continues to be a complex issue discussed across the field, with many researchers and practitioners trying to find the best way to both identify what makes a quality teacher as well as the best ways to measure it. One productive way of making sense of this issue is through Goe's (2007) framework on teacher quality, which categorizes these attributes into inputs, processes, and outcomes. In this framework, Goe focuses on outcomes that use empirical measures of teacher effectiveness, which link teachers to student achievement data.

The following sections discuss research on teacher quality that examines the relationship between teacher related attributes and their effectiveness through the lens of Goe's framework that categorizes teacher attributes into inputs, processes, and outputs. It concludes with a description of the conceptual framework used in this study.

1.2.3.1 Inputs

Goe (2007) groups attributes of teacher quality, referred to as inputs, into two categories – teacher qualifications and teacher characteristics. Teacher qualifications include their education certification, credentials, and experience while teacher characteristics include teachers' race and gender, as well as their attitudes and beliefs.

1.2.3.2 Teacher Qualifications

The first set of inputs in Goe's (2007) framework on teacher quality is teacher qualifications. Measures of teacher qualifications include their education, certification, teacher test scores, experience, and the prestige ratings of their teacher education programs. These qualifications are resources that teachers bring with them into the classroom and are often used in teacher hiring decisions (Goe, 2007). Various aspects of teacher qualifications have been examined to identify their relationship or contribution to student learning. Unfortunately, researchers' findings are inconsistent (Betts, et al., 2003; Boyd, et al., 2006, 2008; Darling-Hammond, et al., 2005; Goldhaber & Brewer, 2000; Hanushek, 2003; Hanushek, et al., 2005; Nye et al., 2004; Rice 2003; Rivkin, et al., 2005). Most promising as a qualification indicator of teacher quality is teacher experience, with teachers having the most influence on student learning in their first few years (typically within the first 3-5 years) of teaching. A review of research on teacher qualifications is provided below.

1.2.3.2.1 Teacher Certification

Most teachers gain their way into the classroom through a certification process, which purports to validate teachers' knowledge and skills. Certification processes can require the assessment of teachers' knowledge of the subject matter taught; how they teach content to a wide range of learners; their ability to manage a classroom, design and implement instruction; or their ability work effectively with students, parents, and other school professionals (Darling-Hammond, et al., 2005). Teacher licensing, or the certification process, is the responsibility of each state. There are variations in state licensure and certification policies. For example, some states require prospective teachers to pass standardized or competency exams, and/or to have a minimum grade point average for entering into a teacher education program (Darling-Hammond, et al., 2005; Goldhaber & Brewer, 2000). More recently, select states require candidates to pass standardized content knowledge exams and demonstrate teaching competence through a score on a performance task, like the edTPA (i.e., New York, Minnesota, Washington).

Teachers can obtain various types of certifications (e.g., standard, private school, temporary, provisional, or emergency certifications) (Goldhaber & Brewer, 2000). However, some teachers gain entry into the profession through alternative routes, often obtaining full, temporary, or emergency certification while already in the profession (e.g., Teach for America, Teacher Fellows program, lateral-entry programs). Findings related to the effect of different types of teacher certifications on teacher quality demonstrate that teachers entering under some routes of entry may be more effective than other routes (Darling-Hammond, et al., 2005; Kane, et al., 2006).

When comparing certified teachers, of any type, to non-certified teachers, researchers have found that certified teachers produce stronger student achievement gains than non-certified teachers (Darling-Hammond, et al., 2005; Kane, et al., 2006), or do no worse than non-certified teachers (Croninger, et al., 2005; Kane, et al., 2006). An exception is a study using Early Childhood Longitudinal Study (ECLS) data which found that teachers with full certification produced significant teacher effects in first graders' reading achievement gains (effect size = 0.09), but found no such associations in the children's math achievement gains (Palardy & Rumberger, 2008). Research on subject-specific certifications will be discussed in the later section on teachers' pedagogical content knowledge.

When schools or districts experience difficulty finding fully-credentialed teachers, many states allow teachers to obtain emergency licensure in order to teach immediately and fill short-term vacancies (Goe, 2002; Goldhaber & Brewer, 2000). Studies of emergency credentialed teachers found varied relationships with student achievement. Students of emergency credentialed teachers in math and science were found to do no worse than students of teachers with standard credentials (Goldhaber & Brewer, 1999, 2000), whereas Clotfelter and colleagues (2007) found that teachers with emergency credentials had negative effects on student achievement in both math and reading. However, Betts and colleagues (2003) found that teachers with emergency credentials and 0-1 years of experience were associated with larger gains in student achievement than more experienced teachers with full credentials.

Teachers may also enter the profession through alternative certification programs (i.e., Teach for America (TFA), Teacher Fellows). Studies on the effectiveness of TFA teachers, who mostly graduated from competitive or high-ranking colleges (Decker, et al., 2004), tend to indicate that they are slightly more effective than non-TFA teachers in math instruction (Darling-Hammond, et al., 2005; Decker, et al., 2004; Heilig & Jez, 2010; Kane, et al., 2006; Xu, et al., 2007). For instance, using experimental design (Glazerman, et al., 2006) or mini-experiments (Decker, et al., 2004), studies on the effectiveness of TFA teachers have found that TFA teachers have a significantly greater impact on students' math achievement than non-TFA certified teachers, although they have no greater impact on reading achievement. This impact was equivalent to about a 0.15 standard deviation increase in math achievement or about one additional month of math instruction (Decker, et al., 2004; Glazerman, et al., 2006). In two studies of New York City TFA teachers, researchers found that TFA teachers had a slightly greater impact on students' math achievement than non-TFA teachers, although in reading, these researchers discovered no effect or a negative effect of TFA teachers on students' reading achievement, with the children of TFA teachers scoring significantly lower in reading than the students of non-TFA teachers (Boyd, et al., 2006; Kane, et al., 2006). Students assigned to classrooms taught by TFA teachers scored 0.02 standard deviations higher in math than certified, non-TFA teachers (Boyd, et al., 2006; Kane, et al., 2006).

Similar results were found when examining the impact of teachers in the Teacher Fellows program, another alternative certification program that focuses on training teachers to teach in urban schools and is used to assist with teacher shortages in New York City. Teacher Fellows were found to be no more effective at improving students' reading achievement and slightly more effective at improving students' math achievement than traditionally certified teachers (Boyd, et al., 2006; Kane, et al., 2006). In Kane and colleagues (2006) study, for instance, no difference was found in the impact of Teacher Fellows and non-Teacher Fellows in students' math achievement and students assigned to Teacher Fellows scored below certified, non-Teacher Fellows by 0.01 standard deviations in reading.

While not an alternative form of certification, the National Board for Professional Teaching Standards (NBTS), whose mission is to "advance student learning and achievement by establishing the definitive standards and systems for certifying accomplished educators" (http://www.nbpts.org/mission-history), provides experienced teachers with this certification as a distinction of accomplished teachers. These teachers, who are known as "National Board Certified Teachers" (NBCT's), were found to produce small, significant gains in student achievement and be more effective than non-NBCT's (Cavallozzo, 2004; Clotfelter, et al., 2006, 2007, 2010; Goldhaber & Anthony, 2005; Vandervoort, et al., 2004). In Miami, for example, after including school-level effects, Cavallozzo (2004) concluded that 9th and 10th grade mathematics students taught by NBCT teachers gained an average of 0.07 of a standard deviation in achievement more than they would have under the instruction of non-NBCT teachers. Some authors suggest that either the Board Certification process or the recognition associated with being Board Certified may contribute to their effectiveness (Clotfelter, et al., 2007).

1.2.3.2.2 Teacher Test Scores

Most states' certification process considers teachers' standardized licensure test scores as a proxy measure for what teachers know and can do. These test scores have been examined independent of teachers' certifications in studies on teacher quality. Findings indicate that higher teacher licensure test scores are associated with higher student achievement gains (Clotfelter, et al., 2006, 2010; Ferguson, 1998).

The Praxis, formerly known as the National Teacher Examination (NTE), is one of the most common sets of American teacher certification exams. Praxis II tests, developed by Educational Testing Service, aim to measure teachers' content knowledge. According to two studies led by Clotfelter and colleagues (2006 – cross-sectional linked data; 2007 – 10 year longitudinal), higher Praxis II content knowledge test scores were associated with higher math and reading achievement scores, with larger effects in math than reading. Strauss and Sawyer (1986) studied the impact of teacher quality using NTE scores in both math and reading, concluding that a 1% improvement in NTE scores resulted in 0.5% and 0.8% increase in student mean achievement scores, respectively.

Not all states require teachers to demonstrate their content knowledge on one of the Praxis II series test. North Carolina is one such state. Licensure test scores of North Carolina teachers were found to have a small, significant effect on students' math achievement scores in algebra or geometry, using end-of-year course assessment scores as measures of achievement (Clotfelter et al., 2010).

Tests of math and verbal ability (i.e., non-licensure tests) have also been used in studies on teacher quality. Like the Praxis II test series, these tests aimed to measure subject area knowledge and not teachers' knowledge of pedagogy for specific subjects. Ehrenberg and Brewer (1994) reanalyzed data from the 1966 report *Equality of Educational Opportunity* concluding that teachers' verbal ability scores, as measured by scores on a verbal aptitude test, were a positive predictor of student achievement gains at both the elementary and secondary levels. In Murnane and Phillips' (1981) study of urban elementary teachers, no relationship between student achievement and teachers' verbal ability scores was found using a self-administered word test measure among teachers and student vocabulary achievement scores as measured by the Iowa Test of Basic Skills (ITBS). Hanushek (1992) examined these same data to study the teachers' impact on both ITBS reading and vocabulary scores and concluded that teachers' word test scores were associated with students' greater reading score gains but not with students' vocabulary gains. As it relates to teachers' math ability, Rowan and colleagues (1997) used student math achievement data and teachers' mathematics performance on the National Educational Longitudinal Study of 1988 and found a positive effect of teachers' math test scores on students' achievement gains in math, although the effect was small (0.02 standard deviation).

1.2.3.2.3 Level of Education

Research relating the level of degree obtained by teachers in relation to their influence on student learning produced mixed results. For instance, using Texas data, several researchers concluded that teachers with a master's degree, in any area, produced very small math achievement gains (Hanushek, et al., 2005; Rivkin, et al., 2005) and very small or no additional reading achievement gains (Rivkin, et al., 2005), beyond teachers without a master's degree. Similarly, studies have found that graduate degrees, of any type, were not predictive of higher student achievement (Clotfelter, et al., 2007, 2010; Nye, et al., 2004; Palardy & Rumberger, 2008). Miami teachers with a graduate degree produced a small effect on student achievement with their students increasing test scores by 1.7% (Cavalluzzo, 2004). In addition, Harris and Sass (2007) found negative or insignificant correlations between teachers who earn advanced degrees during their teaching careers and student achievement among students in Florida, except for in middle school math. Yet, student achievement gains were found among middle school English teachers with Ph.D.'s (Betts, et al., 2003).

1.2.3.2.4 Selectivity of Teacher Preparation Programs

Research on the relationship between the selectivity of teachers' preparation programs and student achievement has produced mixed results. Two common guides or ranking systems for determining college selectivity include Barron's College Admissions Selector and Peterson's Guide to 4-Year Colleges and Universities. The college selectivity of teachers in Miami, using Peterson's Guide, was found to result in a small, negative effect suggesting that expected achievement declines with increases in the selectivity of teachers' undergraduate institutions (Cavalluzo, 2004). North Carolina teachers from competitive universities, using Barron's measure, were associated with significantly higher student reading performance (Clotfelter, et al., 2006). Ehrenberg and Brewer (1994) coded Barron's rankings into a six category rating system, concluding that as the average selectivity of the teachers' undergraduate institution increased, student gain scores increased. These results may be indicative of the selection criteria of these schools (e.g., minimum grade point average, SAT scores), which may mean that brighter students may make better teachers.

1.2.3.2.5 Teaching Experience

One of the most observable measures of teacher qualifications is the number of years of teaching experience. This measure has shown to effect student achievement, as teachers in their first few years of teaching have been associated with increased gains in student achievement (Boyd, et al., 2006; Clotfelter, et al., 2007, 2010; Hanushek, 2003; Hanushek, et al., 2005; Harris & Sass, 2007; Kane, et al., 2006; Goe, 2002; Nye et al., 2004; Rice 2003; Rivkin, et al., 2005; Rockoff, 2004), with these gains diminishing after the first few years (Boyd, et al., 2006; Clotfelter, et al., 2010; Hanushek, et al., 2005; Palardy & Rumberger, 2008; Rivkin, et al., 2005). For example, analyses of data from the Tennessee Project STAR (Student Teacher Achievement Ratio) randomized experiment found significant positive relationships between the number of years of teaching experience among teachers and student achievement gains in reading and mathematics, with effects ranging from 0.06 to 0.19 standard deviations (Nye, et al., 2004; Rivkin, et al., 2005). Another study (Kane, et al., 2006) of the relationship between New York teachers' experience and students' achievement suggest that teachers' effectiveness improves during the first few years of experience: students assigned to teachers in their first year scored 0.06 and 0.03 standard deviations lower in math and reading, respectively, than students assigned to those same teachers after they gained two years of teaching experience. It should be noted that researchers have also concluded that teachers with less than 2 years of experience had lower levels of student gains in reading and that students taught be teachers with between 2 and 5 years of experience experienced the greatest impact (Croninger, et al., 2007).

1.2.3.2.6 Pedagogical Content Knowledge

Pedagogical content knowledge is defined as knowing what instructional approaches match the content being taught while also knowing how to arrange content for better instruction (Shulman, 1986, 1987). Researchers have examined this attribute through several different means, including direct researcher-developed measures of pedagogical content knowledge, as well as teachers' subject-specific degrees and certifications.

Research on the relationship between teachers' pedagogical content knowledge and student achievement generally produced positive significant results, primarily in mathematics (Betts, et al., 2003; Cavalluzzo, 2004; Clotfelter, et al., 2010; Croninger, et al., 2007; Frome, et al., 2005; Goldhaber & Brewer, 1999, 2000; Harris & Sass, 2006, 2007; Hill, Rowan & Ball, 2005). For example, Hill and colleagues (2005) found that teachers' mathematical knowledge for teaching in first and third grade is associated with significant gains in student achievement, using a survey measure created under the Study of Instructional Improvement. Similarly, using linked student and teacher data, Harris and Sass (2006, 2007) concluded that mathematical pedagogical knowledge was positively associated with gains in math scores among elementary and middle school students.

In San Diego, Betts and colleagues (2003) found strong evidence that teachers with math specific training in high school math were the best teacher-level predictor of student achievement, with the authors suggesting that more subject-specified training might be needed as the math curriculum gets more difficult. Similarly, Cavalluzzo (2004) found that teachers in Miami with a regular state certification in high school mathematics lead to a 5.7% increase in student test scores. Frome and colleagues (2005) found that the percent of teachers within a school with pedagogy training in math (math education majors) was significantly correlated with student math achievement scores.

Researchers who studied the relationship between student achievement and teachers' certification in the specific subject area(s), found slightly positive results. Consistent with many states, in North Carolina, for example, teachers can receive a subject-specific certification — one that requires that a teacher both successfully complete an approved program of study in a subject area and receive a passing grade on the Praxis II. Clotfelter and colleagues (2010) found statistically significant large effects on student achievement among teachers with subject-specific certification, particularly in math and English/Language Arts (ELA), when compared to teachers without this certification. When examining NELS of 1998 data, Goldhaber and Brewer also concluded that the students of math teachers with subject-specific degrees or certification outperformed the students of math teachers without subject-matter preparation (Goldhaber & Brewer, 1999, 2000). According to their 2000 published report, this difference between math certified and non-math certified teacher's leads to a 0.10 standard deviation increase in 12^{th} grade students' math achievement (Goldhaber & Brewer, 2000). In their 1999 report, teachers holding both a bachelor's and a master's in mathematics represent a 0.08 standard deviation increase in 12th grade math scores, or more than a third of a year of schooling (Goldhaber & Brewer, 1999).

In summary, research relating teacher qualifications (i.e., teacher certification, test scores, education, certification, experience) to teachers' impact on student achievement is vast, however results generally show small, inconsistent effects. Research shows that teacher experience and pedagogical content knowledge are among the stronger teacher qualification predictors of teacher quality. Several studies concluded that in the first three to five years of teaching, teachers produce great gains in student achievement, with diminishing effects in later years. Furthermore, teachers with strong pedagogical content knowledge have been associated with small, positive effects on student learning, particularly in mathematics.

1.2.3.3 Teacher Characteristics

The second category of inputs based on Goe (2007) is teacher characteristics, including teachers' attitudes, beliefs, race and gender. Several researchers who examined relationships between these attributes and student achievement, found small, inconsistent, and often times even weaker effects on student achievement, when compared to teacher qualification attributes (Clotfelter, et al., 2010; Dee, 2004; Ehrenberg, Goldhaber, and Brewer, 1994; Goddard, Hoy & Hoy, 2000; Hanushek, et al., 2005).

1.2.3.3.1 Race/Ethnicity

Several studies have examined the relationship between teachers' racial or ethnic backgrounds and student achievement. Most studies found positive effects on achievement when Black teachers taught students of their own race (Dee, 2004; Clotfelter, 2010; Hanushek, et al., 2005). For example, using matched panel data from a large district in Texas, Hanushek and colleagues (2005) found inconsistent direct relationships between race and student gains and found strong evidence that minority teachers and students matched by race were associated with positive effects on student gains. Similar matched race findings were found in the evaluation of student and teacher data in the Tennessee STAR experiment, Black students assigned to Black teachers for a year resulted in a 3 to 5 percentile point increase in math scores and 3 to 6 percentile point increase in reading on the Stanford Achievement Tests (Dee, 2004). In addition, White students taught by White teachers were associated with a 4 to 5 percentile point increase in math scores (Dee, 2004). Interestingly, Clotfelter (2010) found that Black teachers had a negative impact on North Carolina White students'. There was also little association found between the race or ethnicity and gender of 10th grade teachers and student achievement in history, reading, mathematics and science using NELS data from 1988 (Ehrenberger, et al., 1994).

1.2.3.3.2 Gender

Similarly, little association has been found between teachers' gender and student achievement. In Clotfelter's (2010) study of elementary teachers and students in North Carolina, the author found, interestingly, that male teachers produce negative effects on the achievement of female students. Male and female teachers were found to be equally as effective with students of their same gender. In a previous 2007 study, elementary male teachers in North Carolina were found to produce a significantly negative effect on students' reading achievement, although no significant differences were found in math (Clotfelter, et al., 2006). Several relationships were found between both teachers' race and gender, and their 8th grade students' achievement in Ehrenberg and colleagues (1995) study using NELS 1988 data, suggesting that a teacher's race and gender is not the strongest predictor of students' gains. For example, Black male teachers were associated with higher history gain scores for Black and White male and White female students, but lower reading scores for Hispanic male students (Ehrenber, et al., 1995). In addition, White male teachers were associated with higher science scores for Hispanic female students, but lower reading and history scores for Hispanic male students (Ehrenber, et al., 1995).

1.2.3.3.3 Teachers' Attitudes and Beliefs

Researchers have also used teachers' attitudes and beliefs in defining teacher quality including teachers' self-efficacy and their motivation or expectations for students. Measures of teacher efficacy examine teachers' expectations or confidence held about their individual capability to influence student learning or outcomes (Klassen, et al., 2010), whereas measures of teacher expectations or motivation are used to measure their expectations for specific students (Rowan, et al., 1997). Researchers have suggested that higher teacher motivation and self-efficacy improves teachers' performance (Bandura, 1997; Rowan, et al., 1997), teacher expectations play a significant role in how well students learn (Bamburg, 1994), and that teachers put more effort into teaching students' whom they expect to learn more (Rowan, et al., 1997). Teachers' attitudes and beliefs have been measured using data from both teacher and student surveys.

Reading achievement gains were found to be lower among teachers with negative expectations, with a significant effect size of -0.04, using ECLS achievement data and teacher surveys (Palardy & Rumberger, 2008). Teacher efficacy was also negatively associated with math achievement gains, with lower achievement gains found among teachers with a negative sense of teacher efficacy (effect size = -0.04) (Palardy & Rumberger, 2008). Frome and colleagues (2005) used student surveys to examine the link between teachers' motivation and expectations and student achievement in Georgia. They conclude that students' ratings of teachers' expectations were significantly related to their reading, mathematics, and science achievement scores. Other researchers, using student achievement and teacher survey data from NELS of 1988, found small, but significant, effects on math achievement with increased levels of teachers' expectations for student outcomes, (Rowan, et al., 1997).

Some teacher characteristics were shown to produce positive relationships with student achievement, although their related effects were generally small and inconsistent across subjects and grade-levels. Teachers' race, when matched with the race of their students, was one of the stronger predictors of student achievement among the different aspects of teacher characteristics.

1.2.3.4 Processes

Goe (2007) describes processes related to teacher quality as the quality of their instructional practices. Teacher practices, including planning (both in and out of the classroom), classroom management, instructional delivery, and teachers' interactions with students, have also been considered in evaluations and discussions on teacher quality. Examining teacher quality not only by their qualifications on paper but also by what they do in the classroom is a critical component of current efforts to evaluate teachers (Goe, 2007).

1.2.3.5 Instructional Practices

Using scores from teacher evaluation systems, ratings of teachers' classroom assignments, or results from teacher or student surveys to explore various teacher practices, positive correlations have been found suggesting that instruction may matter in defining teacher quality (Borman & Kimball, 2005; Cohen & Hill, 1998; Frome, et al., 2005; McCaffrey, et al., 2003; Smith, Lee, and Newmann, 2001).

There are several common standards-based teacher evaluation systems used to examine the quality of teachers' instructional practices, or teaching quality. Standards-based teacher evaluation systems are those developed based on empirical and theoretical literature on effective teaching behaviors, assessed using a comprehensive set of standards and rubrics in an effort to improve teachers' instruction (Borman & Kimball, 2005; Danielson, 1996). For example, many states and districts have adopted standards-based teacher evaluations systems based on Charlotte Danielsons' (1996) *Enhancing Professional Practice: A Framework for Teaching*. Most research on teaching quality found that teachers with higher ratings on evaluation systems that use this framework generally produce higher gains in student achievement (Borman & Kimball, 2005; Gallagher, 2004; Heneman, et al., 2006; Holtzapple, 2003; Kane, et al., 2011; Kimball, et al., 2004; Milanowski, 2004). Researchers Borman and Kimball (2005) used the framework's teacher evaluation ratings to establish "high-quality" teachers – teachers in the 84th percentile or above; and "low-quality" teachers – teachers in the 16th percentile or below. Their research concluded that high-quality teachers were associated with higher levels of student achievement, with a difference of 0.01 standard deviations between the two sets of teachers (Borman & Kimball, 2005).

The Classroom Assessment Scoring System™ (CLASS) (Pianta, La Paro, & Hamre, 2008) also has been used to examine teachers' instructional practices. CLASS was developed by researchers from the University of Virginia's Curry School of Education to measure the extent to which teachers effectively support children's social and academic development, focusing on teacher/student classroom interactions that support student learning (Teachstone Training, 2014). An examination of the teachers' instructional practices using CLASS among secondary teachers concluded that classrooms with positive emotional climate sensitive to student needs and perspectives, those that use diverse and engaging learning formats, and those that focused on analysis and problem solving were positively associated with higher levels of student achievement (Allen, et al., 2013). However, using this same CLASS system, classrooms with higher teaching quality ratings were not found to be a significant predictor of reading achievement among kindergarten classrooms (Ponitz, et al., 2009).

Instructional quality has been examined using other teacher-related instructional measures including teachers' assignments and teacher surveys. Teacher surveys have shown small, significant effects, particularly in math achievement (McCaffrey, et al., 2001; Wenglinsky, 2000, 2002; Smith, et al., 2001). Wenglinsky (2002), for instance, used 1996 NAEP teacher survey data to examine the relationship between teachers' practices and students' math achievement and found that classroom practices (as measured by 21 different variables) had a positive impact on student achievement with a total effect size of .56. Using teacher questionnaires, McCaffrey and colleagues (2001) concluded that teachers' use of instructional practices in mathematics, as prescribed by the National Council of Teachers for Mathematics, produced higher student math achievement scores in integrated math, but not in algebra or geometry. Using survey data from elementary school teachers in California, Cohen and Hill (1998) concluded that teachers' use of instructional practices, using California's *Mathematics Framework* as the lens, produced modest student gains (Cohen & Hill, 1998). In examining the intellectual demands of assignments given to students, researchers concluded that classrooms with more challenging intellectual work produce greater gains in scores on both the ITBS and the Illinois Goals Assessment Program (Newmann, Bryk, and Nagaoka, 2001). Palardy and Rumberger (2008) found that the frequency of reading instruction practices among 1st grade teachers had a significant positive association with reading gains (effect size = 0.03).

Student perception surveys on specific aspects of teachers' instructional practices were developed in the Tripod Project for School Improvement (Ferguson, 2008) and used in the Measures of Effectiveness study on teacher quality. Researchers concluded that student perceptions of their teachers' instructional practices were moderately associated with student achievement gains with standard deviations of 0.35 in math and .10 in ELA using standardized state assessments (Rothstein, 2011). Students have also been used to assess the impact of teachers' instructional practices. Student reported use of instructional practices among 8th grade teachers in Georgia was associated with higher student test scores in both math and reading (Frome, et al., 2005).

Generally speaking, teachers' instructional practices are a strong indicator of students' achievement. The effects of teachers' instructional practices are slightly better than those related to teachers' characteristics or qualifications, suggesting that instructional practices may play a bigger part in improving student learning and defining teacher quality.

1.2.3.6 Outcomes

The third dimension of Goe's framework pertains to outcomes of teacher and teaching quality as measured by individual teachers' impacts on student learning or their effectiveness. As an outcome, teacher effectiveness is measured by teachers' contribution to their students' learning. A common approach to the measuring of teacher effectiveness is through Value-Added Modeling (VAM). This section first describes approaches to VAM and then provides a review of studies that use VAM to demonstrate that differences in teacher effectiveness exist, with less of a focus on which teacher attributes contribute to these differences, as discussed in prior sections.

1.2.3.6.1 Analytic Approaches to VAM

Researchers have generally used three different approaches to VAM – covariate adjustment models; gain score models; and multivariate models (McCaffrey, et al., 2003; McCaffrey, et al., 2004). The typical value-added analytical approach used to examine attributes of teachers to student achievement is that of a covariate adjustment model often presented as education production functions (McCaffrey, et al., 2003; McCaffrey, et al., 2004). Covariate adjustment models regress current test scores on prior test scores, specifying the current score as a function of the prior score and other possible covariates using separate models for each year and explicitly linking students' scores to the effects of their current teachers only.

Gain score models specify a one-year gain score (current year's test score less prior year's test score) separately for each year and link student gains to their current-year teacher effects (McCaffrey, et al., 2003; McCaffrey, et al., 2004). These gains have typically been measured from spring of one grade to spring of the next, although they can also be measured from fall of one year to spring of the same school year.

In both of these models, residual errors are assumed to be independent and normally distributed. They both also treat teacher effects as fixed or random. The main difference between these two models is that the gain score model assumes that the effects of previous teachers is permanent and unchanging while the covariate adjustment model makes no assumptions and allows the estimation of the persistence of teacher effects.

The third type, the multivariate model, specifies a joint distribution for the entire multivariate vector of scores for the student, expressing the score means as a function of time, specifying the variances and correlations between pairs of scores for different years, and links students' scores to teacher effects for multiple years (McCaffrey, et al., 2003; McCaffrey, et al., 2004). These models also account for missing data and provide more flexibility in exploring other assumptions such as the persistence of teacher effects and the residual covariance structure of student outcomes. While these models seem to be more promising in estimating teacher effects, they are computationally challenging. Recent approaches to this model include cross-classified models, layered models, and variable persistent models (McCaffrey, et al., 2003; McCaffrey, et al., 2004).

1.2.3.6.2 Studies of Teacher Effectiveness

Across the variety of approaches to VAM, researchers have determined that teachers do matter, and have small to moderate effects on student learning (Aaronson, et al., 2003Goe, 2007; Hanushek et al., 2005; Jacob, A., 2012; Rivkin et al., 2005;

Nye, et al., 2004; Rowan, et al., 2002; Stronge, et al., 2011). For example, Rivkin and colleagues (2005) used Texas matched panel data from Texas to examine the impact of teachers on student achievement and concluded that a one standard deviation increase in average teacher quality, using an empirical estimation of within-school variance of teacher effectiveness, raises average achievement in that grade by at least 0.11 standard deviations in math and 0.10 standard deviations in reading (Rivkin, et al., 2005). In Chicago, Aaronson and colleagues (2003) concluded that having a teacher who was rated one standard deviation higher than other teachers in teacher effectiveness, according to value-added scores, resulted in an increase of 22% of an average yearly gain in math achievement and that most of the variation in teacher effects is not explained by observable teacher characteristics (at most 10% of the variation is explained by teacher attributes). In examining 5th grade end-of-course assessment scores, Stronge and colleagues (2011) found that teacher effects contributed to 0.59 standard deviation student achievement gains in reading and 0.45 standard deviation student achievement gains in mathematics. Rockoff (2004) found that the impact of teachers in a New Jersey county increased reading vocabulary scores by 0.11 standard deviations, reading comprehension scores by 0.08 standard deviations, math computation by 0.11 standard deviations, and math concepts by 0.10 standard deviations, using elementary students' nationally standardized basic skill assessments. Nye and colleagues (2004) used data from the Tennessee Project STAR randomized experiment and concluded that the impact of teachers on reading achievement was 0.26 standard deviations, and 0.36 standard deviations on math achievement. Finally, Rowan and colleagues (2002) compared three different approaches to estimating teacher effects. Teachers' contribution to total variability in gains, across multiple subjects and cohorts, ranged from 0.21 to 0.42 standard deviations using covariate adjustment models, from 0.16 to 0.36 standard deviations using one-year gains (i.e., current year score minus prior year score) while adjusting for background variables (e.g., gain score models), and from 0.32 to 0.45 standard deviations using cross-classified random effects models.

1.2.4 Conceptual Framework for Present Study

Drawing on the research presented in this chapter, I have established a framework with which to understand and define teacher quality. This framework is built off of Goe's (2007) work on teacher quality, which I find to be a productive way to make meaning of such a complicated concept. Her use of inputs, processes and outcomes make it easier to understand the different aspects of teaching and how they may situate themselves in discussions and research on teacher quality. Inputs include teachers' characteristics or qualifications, processes include teachers' instructional practices, and outcomes include empirical measures of teachers' effectiveness.

The conceptual framework for this dissertation continues to group teacher attributes into inputs, processes, and outcomes, and considers pedagogical content knowledge as an attribute on its own - a combination of both inputs and processes. Goe (2007) considers this attribute a measure of teacher qualifications, although Shulman (1986) argues that good teachers should have content knowledge that "embodies the aspects of content most germane to its teachability (p. 9)", or more specifically, strong pedagogical content knowledge. This attribute represents the "blending of content (inputs) and pedagogy (processes) into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (Shulman, 1987, p. 4). Furthermore, according to research, pedagogical content knowledge has shown enough evidence of having some relationship with student gains to be considered an attribute of its own in defining teacher quality. Therefore, in this framework, pedagogical content knowledge is considered a unique, independent attribute.

The teacher quality framework for this study also considers mediating relationships between teacher characteristics and qualifications, and pedagogical content knowledge and instructional quality, in relation to teacher effectiveness measures. Under Goe's framework, inputs and processes are considered to have direct relationships with measures of teacher effectiveness. Some researchers have argued that it is what takes place in the classroom and what can be measured by their instructional practices or pedagogical content knowledge, which is likely to be the mechanism by which teachers affect student learning (Grossman, et al., 2010). Additionally, previous research has not shown strong evidence of direct relationships between teacher characteristics and qualifications, and teacher effectiveness, thus this model suggests that other mediating factors may account for significant variations in measures of teacher effectiveness.

In support of these arguments and understanding of teacher quality, I consider teacher quality through a slightly different framework than that established by Goe (2007). The conceptual framework for this study considers instructional quality, as measured by teacher practices, and teachers' pedagogical content knowledge as mediators or mechanisms through which teacher characteristics and qualifications are able to influence teacher effectiveness (Baron & Kenny, 1986). Based on previous findings, teachers' characteristics and qualifications have mixed relationships with student learning, so that there is no clear implication for making decisions about hiring and retaining teachers. Whereas, addressing issues of teachers' instructional quality or pedagogical content knowledge, independently or in combination with teachers' characteristics or qualifications, is likely to be the best approach in improving teacher quality.

Through the lens of this framework, I use data from the Measures of Effective Teaching (MET) longitudinal database to investigate teacher quality as the relationship between teacher characteristics and qualifications, and teacher effectiveness as mediated by instructional quality and teachers' pedagogical content knowledge, while focusing solely on urban teachers. A visual representation of the conceptual framework for this study is presented in Figure 1 below.

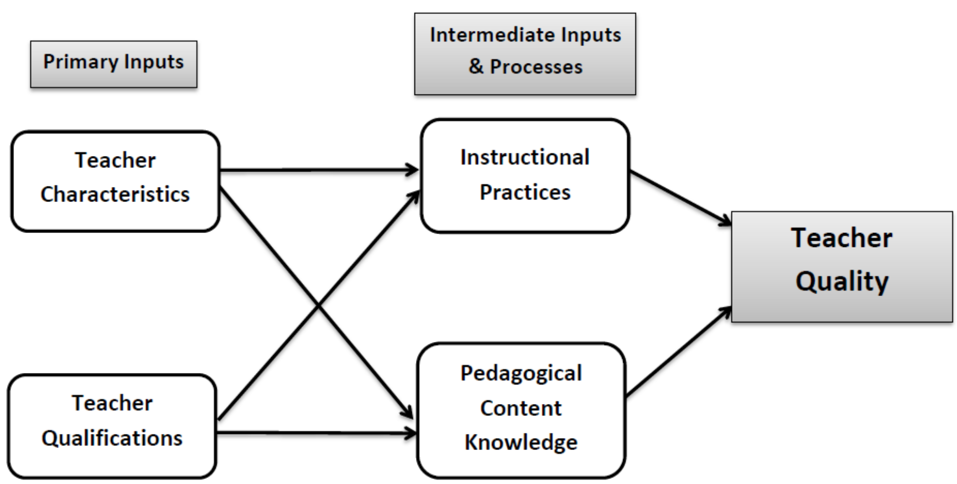


Figure 1. Teacher Quality Conceptual Framework for Present Study

1.3 Limitations to Previous Research

Researchers have concluded that teachers have an impact on student learning, yet based on existing research, variables often used to define teacher quality have done little to explain the variation in teacher effects, only showing small to moderate associations of student achievement to teacher quality. In examining the research on teacher quality, there are several limitations including the methodological approaches used in examining teacher quality, ways of estimating teacher effects, issues with measurement and sampling error, sample selection, and random error in growth models.

1.3.1 Methodological Approaches

Methodological approaches used to examine teacher quality typically involve a variation of the education production function, which is a regression model that attempts to measure to what extent changes in student achievement can be attributed to the teacher or teaching attributes that students receive. While many types of outcomes can be considered in these approaches, most value-added modeling to date focuses exclusively on scores from standardized assessments. These models vary in the types of variables controlled for and in the inclusion, or exclusion, of teacher fixed or random effects. Some studies tended to use basic regression models that accounted for typical student and school characteristics in their models (i.e., race, socioeconomic status, school size) using teacher fixed effects or correlational studies to measure the degree of the association between teacher variables and student outcomes (e.g., Harris & Sass, 2006, 2007). Later studies address the nested structure of educational data by employing multilevel regression models that included teacher random effects to account for variances across all levels of the data (e.g., Rowan, et al., 2002). Despite this, researchers have recognized that while it is nearly impossible to establish true causation due to the complexity of educational contexts and limitations in measuring unobserved effects on student learning, one of the best statistical models used to establish causation in teacher quality literature is VAM. Even still, while VAM is a promising methodological approach to studying teacher quality, researchers must use the most appropriate VAM model for the data and inquiry at hand.

1.3.2 Estimating Teacher Effects with VAM

VAM allows researchers to estimate the effects of educational inputs on student achievement. Teacher effects, under this modeling, are estimates of teachers' contributions to student learning or variations between classrooms in achievement gains, while sometimes controlling for student characteristics, family backgrounds and students' prior knowledge in subjects examined (McCaffrey, et al., 2003; Nye, et al., 2004). However, researchers must clearly consider several issues, including which teacher and student measures are used in the models, confounding factors or missing variables in estimating and isolating teacher effects, multiple measures of the same student, multiple teachers instructing each student, and the possibility that teacher effectiveness may vary over time (Hanushek & Rivkin, 2020, Jacob, 2007; Kane & Staiger, 2008; McCaffrey, et al., 2003; McCaffrey, et al., 2004).

If the idea in estimating causal effects is the comparison of students' achievement with the current teacher to the students' achievement under a plausible alternative, we must be clear about what that alternative must be (McCaffrey, et al., 2003). If it is another teacher, is it another teacher in the same school, district, or teachers who teach similar students regardless of location? The appropriate alternative is likely to depend on the purpose of estimating teacher effects. In most recent cases, teacher effects are estimated with respect to school districts, so that the plausible alternative is the average teacher in the school district.

The use of longitudinal student outcome data presents challenges for statistical modeling of teacher effects and the variability among teachers including multiple measures of the same student and multiple teachers instructing each student (McCaffrey, et al., 2004). In many cases, class groupings change annually and students are taught by a different teacher each year, which means that student outcomes do not follow the traditional nesting structure identified for use in multilevel modeling (McCaffrey, et al., 2004; Raudenbush & Bryk, 2002). As a result, it is common to see value-added modeling of teacher effects in elementary schools where few shifts in teachers and students occur over time. Not only does the use of elementary schools in VAM's avoid problems with attrition when students receive instruction from multiple teachers, as elementary students are typically in self-contained rooms with few class group changes throughout the year, but the linkage between course content and what students are tested on tends to be stronger in elementary and middle schools (McCaffrey, et al., 2009).

Truly isolating causal inference can also be a challenge in estimating teacher effects (Jacob, 2007; McCaffrey, et al., 2003). The context of the school, district, or other educational factors as well as student characteristics, their environments, neighborhoods, families and peers, along with sources we may consider residual or measurement error may also have an influence on estimates of teacher effects. If so, teacher effectiveness is no longer attributable to the teacher alone. For example, although teacher effects are partly a function of the school, some researchers may be interested in examining teacher impacts while distinguishing them from their settings or school policies, and thus inferences must consider a school effect. Others may be interested in measuring the variability of teacher effects at a given time point and under their current context, thus the teacher effect of interest, by definition, includes the indirect effects of schools that affect students through teachers. In order to control fully for unmeasured teacher effects, some value-added studies may compare teachers within the same school, although this limits the capability to measure the relative effectiveness of teachers across different schools or districts (Jacob, 2007). Although modeling gain scores might be the best metric for measuring the importance of teacher effects and is preferable to modeling cohort-to-cohort test-score gains, as is commonly done in VAM, they do not necessarily measure the effects solely attributable to teachers and do not ensure that estimates are not confounded by other factors. While some VAM models use covariates, such as prior year test scores and other available covariates to account for confounding factors, this does not guarantee that estimated teacher effects are truly causal effects. While VAM estimates can control for confounding factors, VAM estimations can still be problematic without randomization as it can be difficult to identify and separate school versus teacher effects in schools that serve very different kinds of students, as the more variable the context of the school the harder it is to assume that VAM estimates correspond to teachers' practices (McCaffrey, et al., 2004; Raudenbush, 2004). There is some concern that controlling for confounding variables, given the uneven distribution of teacher quality across schools and districts, inflates or deflates VAM estimates depending on the student or school characteristic (McCaffrey, et al., 2003). Nevertheless, Raudenbush (2004) argues for using VAM to estimate the combined effects of context and practice at the classroom and school levels. He further suggests, in efforts to truly try to estimate teacher effects from their instructional practices, the use of models that yield multiple growth curves for each child through multiple subject area tests as well as the use of multiple cohorts of students in order to increase the precision of VAM estimates (Raudenbush, 2004).

Some studies also omit measurable variables used to account for confounding factors in VAM estimations, which can also lead to biased estimates (Hanushek & Rivken, 2010). While others aggregate student achievement and teacher attributes at the school level in their VAM estimations, which can create aggregation bias and misestimation of standard errors (Raudenbush & Bryk, 2002).

Finally, one must also consider that teacher effectiveness can vary over time (Ballou 2005; Darling-Hammond, et al., 2012; Hanushek & Rivkin, 2010; Jacob, et al., 2010; Kane & Staiger, 2008; McCaffrey, et al., 2003; Schochet and Chiang 2010). Some researchers have proven that teacher effects can change over time – improving or declining for various reasons including teachers' experience, changes in class assignments, or in response to other factors outside of the school (Jacob, et al., 2010; Kane & Staiger, 2001; McCaffrey, et al., 2003; Rivkin, et al., 2000). As a result, an explicit statement detailing the causal effect of interest should be provided. Nevertheless, despite these concerns, VAM is one of the best possible and more recent approaches to estimating teacher effects and impact on student learning for low-stakes diagnostic purposes, as is intentioned under this study (McCaffrey, et al., 2003).

1.3.3 Measurement Error

As most VAM's involve the use of standardized assessments, there becomes a concern over measurement error and whether the use of school performance measures capture students' true knowledge. Measurement error can both result from the test itself as well as from the students who take them. Measurement error can result from the choice of test questions, random events or influences on students in testing situations, students' familiarity with the test, subjectivity in grading open-ended questions, and other factors which can cause measured scores to differ from students' true knowledge, creating biased value-added estimates of teacher effects or measurement error (Hanushek & Rivkin, 2010; Harris, 2011; McCaffrey, et al., 2004). Furthermore, the variance in measurement error is not constant across the range of "true scores" (McCaffrey, et al., 2004). As measurement error tends to diminish if teachers are observed over multiple years and with large numbers of students (Hanushek & Rivkin, 2010), studies using VAM models must both consider the presence of measurement error as well as the variability in measurement error in estimates of teacher effects (Schochet & Chiang, 2010).

1.3.4 Sampling Error

Studies of teacher quality must also attempt to reduce sampling error and consider whether the sample selected is representative of the larger population. As long as the samples are carefully and randomly selected so that they are representative of the population, the conclusions drawn can be applicable to the wider populations. However, there always remains some degree of sampling error as samples vary, by chance, from the larger population. Even with great attempts to produce representative samples, through random selection, sampling error still typically occurs by chance and must be noted and communicated within the results of the study (Harris, 2011).

1.3.5 Random Error in Growth Measures

When moving from snapshot measures of teacher performance to measures based on student growth over time, random error increases, as error around student assessment performance is likely to change from one time period to the next. While this error can be reduced by using teachers with large numbers of students and/or by creating value-added measures that take into account this error and shrink based on statistical confidence in the measures, random error cannot be eliminated (Harris, 2011).

1.3.6 Sample Selection

The sample of students selected for the study must be of most interest for the study at hand, as teachers may not be equally effective with all students – some teachers may be more effective with higher achieving students whereas others might be more effective with struggling learners. In addition, some students might be more challenging to teach than others (Harris, 2011; McCaffrey, et al., 2003; Raudenbush, 2004). This can affect VAM results to the extent that students are distributed differentially across classrooms. Teachers' efforts may not be proportional or linearly related to student achievement gains, causing teacher effectiveness to vary with the students' level of achievement. Thus VAM estimations become problematic without randomization when schools serve very different kinds of students as it becomes difficult to separate school versus teacher effects (McCaffrey, et al., 2004; Raudenbush, 2004). Once again, researchers need to be explicit about which effects are being considered if teacher effects are not constant across all students.

1.4 The Present Study

The role of teachers is critical in urban schools, which typically serve low-performing, low-income, and minority students, and face a variety of factors both within and outside the schools that make teaching in urban schools especially challenging (Betts, et al., 2000; Darling-Hammond, 2003, 2007; Hanushek, Kain & Rivkin, 2004; Ingersoll, 2003b). As educational reformers continue to examine ways to improve teacher and teaching quality across all schools, they must consider the unique challenges that urban teachers face. Furthermore, efforts to improve education in urban communities must consider the urban teacher.

While teacher quality is studied in a variety of ways, this dissertation focuses exclusively on the quality of urban teachers by examining characteristics of effective urban teachers and the potential impacts that having more effective teachers in urban schools would have on student learning. The study investigates the relationship between teacher characteristics and qualifications, and teacher effectiveness as mediated by instructional quality and teachers' pedagogical content knowledge, while focusing solely on urban teachers. Outcome measures of teacher effectiveness in this study utilize teachers' specific value-added scores in order to provide better estimates of teachers' impact on student's learning. It also examines these relationships in terms of contextual or compositional effects at the school level, allowing us to understand the potential impact of having a body of quality teachers in urban schools, above and beyond the impact of individual teachers. Finally, this study examines the potential impact that making workforce improvements will have on student learning.

In the subsequent chapters of this dissertation, I detail the methods used to guide this study, the results of analyses, a discussion of results, conclusions drawn from this study, and the potential impact of findings on policy and future research. In

Chapter 2, I present my research questions, as well as the data and analytical methods used in this completing this study. In Chapter 3, I detail the results of analyses, followed by Chapter 4 where I discuss these results. I conclude this dissertation with Chapter 5, in which I draw conclusions and discuss the potential implications of my work.

Chapter 2

METHODS

This study investigates the relationship between teacher characteristics and qualifications, and teacher effectiveness as mediated by instructional quality and teachers' pedagogical content knowledge, while focusing solely on urban teachers. The conceptual framework presented in Chapter 1 considers a) direct relationships between teacher attributes and teacher effects, b) mediating relationships which may help explain relationships between teacher attributes and teacher effects, and c) contextual effects which consider the nested nature of educational data.

This study has two primary goals – to identify significant predictors of teachers' value-added to student learning and to understand the impact on student learning if urban teacher workforce changes are made based on these significant predictors. The following research questions guide this study.

1. Which attributes of teacher quality are predictive of urban teacher effectiveness?
2. What impact would increases in the number of quality teachers in urban schools, through changes in specific teacher attributes or practices, have on student learning?

2.1 Data

To answer these questions, I use data from the Measures of Effective Teaching (MET) study to establish a sample of urban teachers from the six MET districts working in schools with at least 60% minority and 60% low-income student populations. Multilevel structural modeling (MSEM) is employed to examine urban teacher characteristics and qualifications as predictors of teacher effects estimated using value-added modeling, with measures of pedagogical content knowledge and instructional practices as mediators. Using significant teacher attributes resulting from MSEM analyses, I predict the impact on student learning of making improvements to the urban teacher workforce.

The Measures of Effective Teaching (MET) longitudinal database is used in the main analysis of this study to answer the two research questions – "Which attributes of teacher quality are predictive of urban teacher effectiveness?" and "What impact would increases in the number of quality teachers in urban schools, through changes in specific teacher attributes or practices, have on student learning?" The use of data from the MET study allows for the examination of teacher and teaching predictors of teacher effectiveness as well as the potential impact of having more effective urban teachers as predicted through this analysis. The MET database is restricted from general dissemination and a Confidential Data Use Agreement must be established in order to access the data. One main restriction of use of the MET database is the release of school or district level results. Access was granted to the MET database with IRB Approval in October of 2014. The approval letter is included in Appendix A.

With support from the Bill and Melinda Gates Foundation, the MET project is the largest study of classroom teaching ever conducted in the United States, collecting data on a variety of indicators of teacher quality among over 2,500 teachers from six school districts, over a two-year period (see Bill and Melinda Gates Foundation, 2013a and 2013b). Through this study, data were collected on teachers and their teaching including measures of students' achievement used in estimating teachers' value added to student learning, survey of students, video-recorded lessons, assessments of a teacher's pedagogical and content knowledge for teaching, and two different teacher surveys. In addition, principals of the schools where teachers worked also completed a survey and provided administrative data on schools, teachers, and students.

Three MET data files are used for analysis in this study: district/school, teacher and class section. Descriptions of MET data and measures are based on information in the MET database user guide provided by the Inter-Consortium for Political and Social Research (Bill and Melinda Gates Foundation, 2013b) and accompanying reports provided by The Gates Foundation.

2.2 Sample

MET researchers collected data over two two-year periods – 2010 and 2011, 2012 and 2013. The final sample used in the study included a total of six districts, 284 schools and 1,559 teachers and 1,379 classrooms, which were recruited through opportunity sampling. Opportunity or convenience sampling is a non-probability sampling technique where subjects are selected because they volunteer to participate or because they are conveniently accessible (Jupp, 2006). MET teachers received a $1,500 incentive for participating in the project. Districts were also awarded small budgets to provide thank you gifts for participating teachers.

Six large school districts volunteered to participate in the MET study. The six districts were Charlotte-Mecklenburg Schools (North Carolina), Dallas Independent School District (Texas), Denver Public Schools (Colorado), Hillsborough County

Public Schools (Florida), Memphis City Schools (Tennessee), and the New York City Department of Education (New York).[1]

This sampling process continued among elementary, middle, and high schools within each of these districts recruited into the study. As schools were recruited, opportunity sampling continued among teachers, at targeted grade levels, within these schools. Among elementary teachers, sampling was focused on elementary teachers in grades 4 and 5, of which most were subject-matter generalists and taught ELA and mathematics to a single class of students. Sampling in middle school focused on middle school teachers grades 6 through 8, of which about half were ELA teachers and the other half mathematics teachers at these grade-levels. At the high school level, sampling focused on teachers in grade 9, teaching grade 9 English, 9th grade Algebra 1, or 9th grade Biology, of which about one third of 9th grade teachers taught one of these subjects. Only schools that had more than one teacher in a grade teaching the same subject were allowed to enter the study. From the six MET school districts, a total of 2,741 teachers from 317 schools resulted from the initial sampling process.

The overall year-two sample included 2,086 teachers from 310 schools. Teacher attrition occurred in year two as some teachers were lost when their school dropped out of the study. Attrition also occurred when teachers lost interest in the study, they left their school or district, or they began teaching a different subject or grade. Additionally, in year-two the study included a randomization component in order to improve causal inferences about teacher effectiveness (Bill and Melinda Gates

---

[1] Due to the use of the restricted-use MET database, district-specific results are not included in this report.

Foundation, 2013b). Schools selected to participate in the study had to identify a group of teachers in which teachers 1) taught the same subject to students in the same grade, 2) had the necessary certification to they could all teach common classes, and 3) expected to teacher the same subject to students in the same grade in the 2010-11 school year. MET researchers identified these groups of teachers as “exchange groups” within each school. At least two members of an exchange group had to be teaching at the same school for teachers to be randomized and included in the study. One class roster of students for each teacher in an exchange group was identified and these rosters were randomly assigned to exchange group teachers. Teachers who could not be placed in an exchange group were not a part of the randomization process but continued through the study, participating with naturally formed sections. It should be noted that students were not randomly assigned to classrooms in this study, yet teachers were randomly assigned to classrooms. The randomization sample included 1,559 teachers in 284 schools. Classes were randomly assigned to teachers, with one class section per teacher, in the second year of the study – concluding with a total of 1,379 class sections. Table 3.1 below outlines MET samples by year (Bill and Melinda Gates Foundation, 2013b).

Five of the six districts are characterized as “urban” by the National Center for Education Statistics, while one district is instead characterized as “suburb, large”. Locale codes are derived from the schools within each district, where if 50% or more of the schools within a district have the same locale code, the district is assigned that locale code, and thus it was determined that some schools within this district are likely to have urban characteristics, just not all (NCES, 2014). Furthermore, the “suburb, large” district locale is often considered with other “urban” district locales. For instance, the Broad Prize in Urban Education often includes these districts in their selection criteria, therefore data from all six districts were included in this analysis.

Table 2.1: Samples of the MET Study by Year

	Full Sample *All Year One (2009-10) Teachers*	Core Study Sample *All Teachers Present in Year Two (2010-11)*	Randomization Sample *Teachers Randomized in Year Two (2010-11)*
Districts	6 Districts Participate	6 Districts Continue	6 Districts Continue
Schools	Opportunity sampling (grade by subject exchange groups required) 317 schools participate	310 schools continue in the study.	284 schools with teachers randomly assigned to classes continue in study.
Teachers	Opportunity Sampling (teacher must be in exchange group at school). 2,741 teachers participate	2,086 teachers continue in the study.	1,559 teachers randomly assigned to classes during summer continue in study.
Classrooms	Opportunity sampling (specialist teachers nominate class sections for study). 4,497 class sections in study	1,909 class sections present in second year of the study.	1,379 class sections (one per teacher) randomly assigned by MET researchers.

2.2.1 Urban School Sample

As the main focus of this study is teachers in at-risk urban schools, this analysis only includes teachers in schools within MET districts categorized in this study as urban. Urbanicity is typically defined as schools located in large, metropolitan cities with high concentrations of minority and low-income students. As I am targeting teachers in the most challenging urban schools, I use this definition of urban as a guideline, but focus only on at-risk schools within these districts, thus the sample for this study only includes teachers from schools that have high percentages of minority and low-income students within urban districts. Knowing that within districts, there can be great variability in the demographic makeup of schools, I needed to establish criteria for identifying a sample of urban schools that serve the academically at-risk within these districts – ideally those schools with both high percentages of low-income and minority students. In order to determine the criteria, I reviewed the distribution of school characteristics in a scatterplot and examined it for trends and outliers. Based on this analysis, the highest concentration of similar schools has both at least 60% minority and 60% low-income students. Thusly, only schools fitting this criterion were included in the final sample for this study. It should be noted that one district did not provide information on whether or not their students received free or reduced lunch. Schools selected from this district were selected solely on their minority student composition.

The final sample included 209 schools from which analyses were conducted. The average percentage of minority students in schools in the urban sample ranged from 60% to 100% and the average percentage of low-income students ranged from 60% to 99% across the six districts. In schools not included in the urban sample, the average percentage of minority students in schools in the urban sample ranged from 12% to 100% and the average percentage of low-income students ranged from 11% to 84% across all six districts. Be mindful that only schools that had both at least 60%

minority and 60% low-income students were included in the urban sample. The number of schools that did not meet the urban criterion within each district ranged from 0(0%) to 55(64%).

The use of data from the MET database restricts me from providing any detailed information on MET district or schools, including their student or teacher characteristics. Ranges of school characteristics across all six districts included in the final randomized sample and those not included in the sample are provided in compliance with these guidelines.

2.2.2 Teacher Sample

Based on the aforementioned criteria established for identifying urban schools – those serving in schools with both at least 60% minority and 60% low-income students – a final randomized sample of urban teachers was identified. T-tests were used to compare the characteristics of teachers in this finalized random sample by type to those teachers not selected for this study. Results are presented separately for ELA and math teachers, although 212 teachers taught both ELA and math across both samples.

2.2.2.1 ELA Sample

Using the sampling criteria for urban schools, a total of 498 teachers were identified for analyses of ELA data, with a total of 153 clusters or schools, which resulted in an average cluster size of 3.26. The majority of teachers in the final ELA randomized sample were female, 88%, compared to 78% among teachers not selected in the sample ($t(2599)= 4.701$, $p<0.001$). About 47% of the teachers in the final randomized sample were White, with 62% among non-sampled teachers

($t(2596) = -6.12$, $p<0.001$). About 48% of sampled urban teachers were Black, compared to only 28% among non-sampled teachers ($t(2596) = 8.38$, $p<0.001$). The final sample of urban teachers was comprised of 4% of Hispanic teachers, which compares to 6% among non-sampled teachers ($t(2596) = -2.140$, $p<0.05$). All other teachers in the sample were of some other race. About 45% of teachers in the sample had a master's degree or higher compared to 35% among non-sampled teachers ($t(1993) = 3.65$, $p<0.001$). The average number of years that teachers taught within their district was 5.9, which is lower than the 7.4 year average among non-sampled teachers ($t(2122) = -4.021$, $p<0.001$). The average number of years of teaching experience overall was 8.3 years, which was lower the average of 9.99 among non-sampled teachers ($t(1155) = -2.255$, $p<0.05$).

2.2.2.2 Math Sample

For the math sample, a total of 475 teachers were identified from 146 schools, resulting in an average cluster size of 3.25. This sample of urban teachers consisted of about 46% White teachers compared to 62% in among the non-sampled teachers ($t(2596) = 6.25$, $p<0.001$). About 47% of the sample urban teachers were Black compared to only 29% among non-sampled teachers ($t(2596) = 7.791$, $p<0.001$). Hispanic teachers made of 4% of the urban teacher sample compared to 6% of the non-sampled teachers ($t(2596) = -1.72$, $p<0.10$). The remainder of teachers in the urban teacher sample used for this study was of some other race or ethnicity. The majority of the teachers were female (80%), which was the percentage among teachers not included in the sample ($t(2599) = 0.172$, $p=.864$). A little less than half (49%) of the urban teachers sampled held a master's degree or higher, compared to 34% among teachers not included in the sample ($t(1993)=5.19$ $p<0.001$). The average number of years of teaching experience in their district among the sampled teachers was 5.7 compared to 7.4 years of teaching experience among non-sampled teachers ($t(2122)= -4.33$, $p=<0.001$). The average number of years teaching, overall, was 8.5 among the final urban sample of teachers compared to 9.92 among non-sampled teachers ($t(1155)= -1.80$, $p=0.071$).

2.3 Measures

MET provides data on the key components of my conceptual framework: teachers' instructional practices, pedagogical content knowledge, and teaching effectiveness. MET researchers also collected information on school and teachers characteristics and qualifications. The MET measures specific to this analysis are discussed in more detail below.

2.3.1 Instructional Practices

Observational videos were used to score instructional practices of teachers using four different measures. Two of these measures were subject specific – Protocol for Language Arts Teaching Observation (PLATO) (Grossman, et al. 2010) and Mathematical Quality of Instruction (MQI) (Hill, 2010). The PLATO system was developed for use in ELA classrooms, while the MQI system was developed for use in mathematics classrooms. The other two measures were used to assess instructional practices in multiple content areas – Framework for Teaching (*FFT*) (Danielson, 2011) and Classroom Assessment Scoring System™ (*CLASS*) (Pianta, La Paro, & Hamre, 2008). Because this study is more interested in overall instructional practices, not related to specific subject areas, this analysis will only include data collected from FFT and CLASS measures.

Video recording of teachers' instructional practices took place between February and June of 2010 in year one of the study and between October 2010 and June 2011 in year two. MET teachers were asked to schedule half of their days for video recording when they were teaching a set of "focal" topics determined by the researchers, and the other half when they were teaching a topic of their choice. Videos were scored according to the measure used for assessing teachers' instructional practices. A study of the observational instruments used in the MET study among teachers in grades 4-8 concluded that the main rater effect for these measures was generally consistent and not drastically different, with no more than 10% of the total variance in scores due to raters either scoring consistently too difficult or too easy (Kane & Staiger, 2012).

2.3.1.1 Framework for Teaching

Charlotte Danielson's (1996) *Framework for Teaching* (FFT) is a research-based set of components used to help describe and evaluate teacher practices. For the MET study, FFT was used in the development of an instrument to evaluate teachers' skills based on evidence from direct observations of classroom practice in determining which aspects of a teacher's practice are most closely related to a high degree of student learning (Bill and Melinda Gates Foundation, 2013b).

The MET protocol developed under FFT divides 22 performance standards into four domains of teaching responsibility – Classroom Environment and Instruction (the domains of planning and preparation and professional responsibilities were not coded). The Classroom Environment domain, used to determine if an environment is conducive to learning overall, includes scales for "Creating an Environment of Respect and Rapport", "Establishing a Culture for Learning", "Managing Classroom

Procedures", "Managing Student Behavior", and "Organizing Physical Space". The Instruction domain, used to determine students' actual engagement with content, includes scales for "Communicating with Students", "Using Questioning and Discussion Techniques", "Engaging Students in Learning", "Using Assessment in Instruction", and "Demonstrating Flexibility and Responsiveness".

The scoring of teacher practices using FFT was done using a detailed scoring rubric, once for each video (e.g., a video scored with the *FFT* protocol will have only 1 score per video on each of the dimensions under this *Framework*). Raters watched 15 minutes at the beginning of each video and then ten additional minutes at the 25-35 minute mark. Raters then scored the video as a single segment. Competencies were rated using a four-point scale, judging the level of performance of each instructional practice as either "unsatisfactory", "basic", "proficient", or "distinguished". The Cronbach's alpha, a reliability estimate, for the, *Framework for Teaching* measure in math are 0.45 in elementary and 0.67 in middle schools. The alpha in ELA is 0.40 in elementary and 0.68 in middle school (Mihaly, et al., 2013).

2.3.1.2 Classroom Assessment Scoring System

Classroom Assessment Scoring System™ (CLASS) (Pianta, La Paro, & Hamre, 2008) is an observational tool developed by researchers from the University of Virginia's Curry School of Education to measure the extent to which teachers effectively support children's social and academic development, focusing on classroom interactions that support student learning (Teachstone Training, 2014). The CLASS tool helps teachers identify and understand the strength of their interactions with students, has a focus on effective teaching, aligns with professional development tools, and works across age levels and subjects (Teachstone Training, 2014). Two

CLASS forms were used in the MET study – upper elementary and secondary. The upper elementary form was used to score ELA and math observational videos from grades 4-5. The secondary form was used to score ELA, math, and Biology observational videos from grades 6-9.

Within the MET study, CLASS observations categorize these interactions into four domains – emotional support, classroom organization, instructional support, and student engagement. Within each of these domains, interactions are further categorized into multiple dimensions. Emotional Support refers to the emotional tone in a classroom and is further categorized into "Positive Climate", "Negative Climate", "Teacher Sensitivity", and "Regard for Student Perspectives". Classroom Organization refers to the way a class is structured to manage students' behavior and is further categorized into "Behavior Management", "Productivity", and "Instructional Learning Formats". The support teachers' give in encouraging students conceptual understanding and their problem solving is referred to their Instructional Support and is further categorized into "Content Understanding", "Analysis and Problem Solving", "Quality of Feedback", and "Instructional Dialogue". The MET study also includes a fourth dimension of "Student Engagement", which includes only one single scoring dimension.

Scoring of teacher practices using the CLASS tool was done through observations of classroom instructional videos of participating randomized teachers using a detailed scoring rubric. Only the first 30 minutes of the videos were scored for the MET study with each instructional dimension scored at 15-minute intervals using a 7-point scale with descriptions of what is to be observed in scoring classrooms "high", "mid", and "low" points on the scale. Dimension scores were often aggregated to higher levels of analysis by averaging raters' scores to get a single segment score and then calculating the harmonic mean of segment scores across all segments for a particular target of measurement (i.e., a day, a class section, a teacher). No reliabilities were estimated for the CLASS measure in this study. The alpha for the *CLASS* measure in math is 0.30 in elementary and 0.58 in middle school. The alpha in ELA is 0.40 in elementary and 0.68 in middle school (Mihaly, et al., 2013).

2.3.2 Pedagogical Content Knowledge

In year two of the study, MET researchers administered web-based Content Knowledge for Teaching (CKT) assessments (ETS, 2011) to measure teachers' content knowledge for teaching (i.e., pedagogical content knowledge) closely tied to the teaching of ELA and mathematics. In ELA, the assessments examined knowledge for teaching such as choosing a text to support a specific teaching goal, choosing activities that highlight a particular feature of a text or literary feature of a text or literary technique, choosing an activity to assess students' understanding, and analyzing student writing for weaknesses or strengths. The mathematics assessments examined knowledge for teaching such as choosing and using appropriate mathematical representations, choosing examples to illustrate a mathematical concept, interpreting student work including use of non-standard strategies, and evaluating student understanding.

Separate CKT assessments were created and administered for grades 4-6 ELA, grades 7-9 ELA, grades 4-5 mathematics, grades 6-8 mathematics, and Algebra I in grade 9. Teachers in grades 4 and 5, teaching ELA and math, were administered two assessments (one for ELA, one for math). The mathematics assessment was administered in Fall, 2010, and the ELA assessment was administered in Winter,

2011. In Winter, 2011, teachers in all other grades were administered only one assessment in the subject area they taught in. Scores reported in the MET data files are based on the number of correct selected-response items on the assessment combined with a total score for constructed-response items. The overall scale score that is reported is a linear transformation of the total score to give a possible range of 0-100.

No detailed psychometric properties of CKT measures are provided by the MET study. Nevertheless, the CKT assessments studied through the MET project were not found to be predictors of student learning – MET project teachers who scored higher on CKT assessments were not substantively more effective in improving student achievement (Cantrell, & Kane, 2013). The present study will test the relationship between CKT with measures of teacher effectiveness under the defined sample of urban teachers to test if any significant relationships exist.

2.3.3 District Administrative Data

Each district provided administrative data on the schools, teachers, and students in year-one of the MET study. Administrative data on schools include measures of school's enrollment size, grade configuration, and student composition. Teacher administrative data included measures of a teacher's sex, ethnicity, subject and grade-level taught, years of teaching, experience, and degree status. Student administrative data included measures of students' sex, ethnicity, free or reduced lunch status, program participation status, and multiple years of scores on state achievement tests. Data on schools, teachers and students were linked in order to identify which students were in a particular teacher's class at multiple times during the MET study.

2.3.4 Value-added Measures of Teacher Effectiveness

Within the MET study, student achievement gains were measured by estimating a value-added model, using students' assessment scores and background information. MET researchers collected a variety of student assessment data by subject area to develop value-added measures, including state assessment data and data from supplemental assessments. As state assessments tended to be multiple-choice tests, MET researchers decided to also use supplemental assessments that were more cognitively challenging. ACT QualityCore end-of-course assessments – English-9, Algebra I, and Biology – were the only student achievement tests administered to students in grade 9 of the MET study.

SAT-9 open-ended (SAT-OE) reading assessment was a supplemental assessment administered in ELA classes in grades 4-8. The Balanced Assessment of Mathematics (BAM) was used as a supplemental mathematics assessment in all mathematics classes in grades 4-8. Students took the assessment corresponding to the subject of their MET section. The administration of state assessments was done according to state-specific timelines whereas the administration of supplemental assessments was done according to MET researcher timelines, but both generally occurred in the spring of each project year.

MET researchers used a VAM covariate adjustment model to specify one year gain scores linking student gains on state and supplemental assessments to their current year teacher effect, based on a single outcome measure (e.g., the state ELA test, the state mathematics test, SAT-OE, or the BAM). MET researchers estimated VAM for each outcome separately by grade and district, thus value-added estimates are test, grade, and district specific and are based on student background variables and prior year's test scores.

State tests among all six MET project districts were different and measured on different scales. To address these differences, MET researchers standardized the test scores to be normally distributed, with a mean of zero and standard deviation of one. This standardization occurred within each assessment administered within each calendar year for each grade and subject. Using these standardized scores of students' achievement on state tests, $S_{it,}$ as the dependent variable, MET researchers estimated the following value-added model for each test outcome, at each grade, in each district.

$$\boldsymbol{S}_{it} = \boldsymbol{X}_{it}\boldsymbol{b} + \bar{\boldsymbol{X}}_{jt}\gamma + \boldsymbol{iS}_{it\text{-}1} + \boldsymbol{\gamma}\bar{\mathbf{S}}_{jt\text{-}1} + \boldsymbol{e}_{it}$$

In the model, the *i* subscript denotes the student, the *j* subscript represents the teacher and the *t* subscript represents the year. X_{it} is a vector of student characteristics including race, gender, free or reduced-price lunch status, English language learner (ELL) students, and participation in gifted and talented programs. In order to capture "peer effects", the $\bar{X}_{jt}$ represents the mean of the student characteristics, prior achievement, and free or reduced-priced lunch status or the students taught by a given teacher in a given year. The students' baseline scores are represented by $S_{it\text{-}1}$, and $\bar{S}_{jt\text{-}1,}$ represent a teacher's mean student baseline score in that year, and are intended to capture the effects of peers. Finally, e_{it} is a student-specific residual.

Residuals from this model represent the degree to which each student outperformed or underperformed similar students and were used in constructing MET value-added scores. For elementary-level generalist teachers, value-added scores for teachers are the average of residuals across all students in a teacher's class. For specialist teachers at all other grade-levels, value-added scores are residuals averaged across sections. Standard errors for each teacher's VAM are the standard deviation of the residuals for a class, divided by the square root of the number of student residuals in the class average. Teacher VAM estimates, using state assessments, will be used as dependent variables in this analysis as measures of teacher effectiveness.

Table 2.2 below provides descriptive statistics of measures used based on each overall sample for ELA and math.

Table 2.2: Descriptive Statistics of MSEM Measures

Measures	ELA			Math		
	n	*M*	SD	*n*	*M*	SD
Male	471	0.12	0.33	455	0.20	0.40
White	471	0.47	0.50	453	0.46	0.50
Years of Exp. In District	378	5.87	5.71	358	5.73	5.74
Master's Degree	374	0.45	0.50	343	0.49	0.50
Ped. Content Knowledge	422	63.21	11.50	413	54.78	14.74
Instructional Quality Factor Scores	493	0.00	0.42	469	0.00	0.40
VAM - State Assessments	498	-0.001	0.22	470	-0.02	0.26
VAM - Supplemental Assessments	480	-0.003	0.31	456	-0.007	0.27
ACT Quality Scores	65	-0.007	0.34	61	0.02	0.39

2.4 Analytical Methods

The main analysis for this study answers research question #1, "Which attributes of teacher quality are predictive of urban teacher effectiveness?" and investigates the relationship between urban teacher characteristics and qualifications, and teacher effectiveness as mediated by teachers' instructional quality and pedagogical content knowledge (as illustrated in Figure 1).

Instructional quality, as measured by evaluations of teacher practices, and teachers' pedagogical content knowledge are considered mediators in this framework. As previous research has not shown strong direct relationships between teacher characteristics and qualifications, and teacher effectiveness, the model used in this study suggests that other mediating factors may account for significant variations in measures of teacher effectiveness. This study also examines these relationships as contextual or compositional effects at the school level, allowing us to understand the potential impact of having a body of quality teachers in urban schools, above and beyond the impact of individual teachers.

2.4.1 Multilevel Structural Equation Modeling

Structural equation modeling, a system of equation regression models, is used when variables of interest cannot be measured perfectly, and can include items that reflect latent variables, mediators, as well as observed measures of a variable (Rabe-Hesketh et al., 2007). An alternative approach often used is multiple regression, which models direct relationships between a dependent variable and one or more independent variables using OLS estimation, assumes perfect measurement of variables, normality of residuals, equality of variance, independence of error terms, linearity and the absence of influential cases (Bollen & Jackman, 1990; Musil, Jones, & Warner, 1998). SEM, on the other hand, is a set of regression models that allow for the testing of both direct and indirect effects simultaneously, does not assume the perfect measurement of variables or that error terms are independent, and allows for the handling of non-normal and incomplete data (Musil, Jones, & Warner, 1998;

Gefen, Straub, & Boudreau, 2000). Furthermore, whereas OLS regression gives biased estimates when the assumption of uncorrelated error terms is violated, SEM uses maximum-likelihood estimation to produce unbiased estimates in non-recursive models, or models that allow for the testing of causation in more than one direction (Musil, Jones, & Warner, 1998).

Conventional SEM assuming simple random sampling alone ignores sampling processes used to generate educational data and results in biased structural regression coefficients (Muthén & Satorra, 1989). While multilevel modeling, alone, prevents studying complex indirect and simultaneous effects within and across levels of the system (Kaplan, 2009). By definition, MSEM is "a synthesis of multilevel and structural equation modeling and is required for valid statistical inference when the units of observation form a hierarchy of nested clusters and some variables of interest are measured by a set of items or fallible instruments" (Rabe-Hesketh et al., 2007, p. 210). Ignoring either the nested structure of the data or indirect relationships between predictors and outcomes can create biases and lead to false inferences (Heck, 2001; Preacher, Zyphur, & Zhang, 2010). Thus, MSEM allows researchers to test mediation hypotheses within these hierarchically clustered data at all levels, simultaneously (Preacher, et al., 2010).

Most studies on teacher effectiveness employ regression models that either aggregate student-level data to the school level or disaggregate school-level data at the student level (Wenglinsky, 2002). The first approach can introduce bias from the aggregated data in the model not being sensitive to the multilevel nature of the data, while the second approach can seriously underestimate standard errors. Furthermore, these models assume that the variables in the model are perfectly measured by the observed data, not taking into account measurement error due to data collection efforts and the operationalization of variables in the model (Byrk & Radenbush, 1992; Wenglinsky, 2002). MSEM allows for the inclusion of latent and observed variables at different levels of the model, allows for the examination of indirect effects related to teacher effectiveness, and reduces bias and measurement error (Rabe-Hesketh et al., 2007, 2012).

The single-level SEM can be expressed in terms of a measurement and a path model. The measurement model related constructs to multiple indicators of those constructs or factors, which, according to Muthén & Asparouhov (2008) can be expressed as

$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda\eta}_i + \mathbf{KX}_i + \boldsymbol{\varepsilon}_i$$

Where i indexes individual cases; Y_i is a p-dimensional vector of measured variables; v is a p-dimensional vector of error terms; Λ is a p X m loading matrix where m is the number of random effects (latent variables); η_i is a m x 1 vector of random effects; and K is a p X q matrix of slopes for the q exogenous covariates in X_i.

The path model, which relates constructs to one another, is (Muthén & Asparouhov, 2008),

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\,\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i$$

where $\boldsymbol{\alpha}$ is an $\boldsymbol{m}$ **x 1** vector of intercept terms, $\mathbf{B}$ is an $\boldsymbol{m}$ **x** $\boldsymbol{m}$ matrix of structural regression parameters, $\boldsymbol{\Gamma}$ is an $\boldsymbol{m}$ **x** $\boldsymbol{q}$ matrix of slope parameters for exogenous covariates, and $\boldsymbol{\zeta}$ is an m-dimensional vector of latent variable regression residuals. Residuals in $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\zeta}_i$ are assumed to be multivariate normally distributed with zero means and with covariance matrices Θ and Ψ, respectively, which are not allowed to vary across clusters.

A multilevel model, on the other hand, expands by allowing elements of some coefficients to vary at the cluster level. A measurement portion of Muthén & Asparouhov's multilevel general model is expressed by

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_{ij} + \mathbf{K}_j\mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$

where j indicates cluster. Likewise, the structural portion of the Muthén & Asparouhov's (2008) multilevel model is expressed by

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j\boldsymbol{\eta}_j + \boldsymbol{\Gamma}_j\mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$

where elements of parameter matrices $\boldsymbol{\nu}_j$, $\boldsymbol{\Lambda}_j$, $\mathbf{K}_j$, $\boldsymbol{\alpha}_j$, $\mathbf{B}_j$, and $\boldsymbol{\Gamma}_j$ can vary at the between level.

The multilevel part of the general model is expressed in the Level-2 structural model, which will be used to answer research question #1:

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\gamma}\mathbf{X}_j + \boldsymbol{\zeta}_j$$

It is to be noted that the $\boldsymbol{\eta}_{ij}$ in the measurement portion of the general model is different than $\boldsymbol{\eta}_j$ in the Level-2 structural model. In the model, $\boldsymbol{\eta}_j$ is a vector that contains all the random effects, stacking the random elements of all the parameter matrices with j subscripts in the measurement and structural portion equations of the general model. $\mathbf{X}_j$ is also different from the earlier $\mathbf{X}_{ij}$, where $\mathbf{X}_j$ is an s-dimensional vector of all cluster-level covariates. The vector $\boldsymbol{\mu}$ (r x1) and matrices $\boldsymbol{\beta}$ (r x r) and $\boldsymbol{\gamma}$ (r x s) contain estimated fixed effects. More specifically, $\boldsymbol{\mu}$ contains means of the random effect distributions and intercepts; $\boldsymbol{\beta}$ contains regression slopes of random effects (i.e., latent variables and random intercepts and slopes) regressed on each other, and $\boldsymbol{\gamma}$ contains regressions slopes of random effects regressed on exogenous cluster-level regressors. Cluster-level residuals in $\boldsymbol{\zeta}_j$ have a multivariate normal distribution with zero means and covariance matrix Ψ.

2.4.1.1 Present Study's Statistical Model

The MSEM model for this study includes a two-level path model with mediators. The path model examines various relationships between teacher attributes and their value-added to student learning. A nested series of structural equation models are fit to the two-level data structure at the teacher and school levels. The advantage of using SEM is that it estimates multiple relationships simultaneously and models can include measurement models, path models, or both. More importantly, this statistical model was decided as the best model to employ in efforts to better understand teacher quality in urban schools based of the framework for this study.

The MSEM model includes an examination of teacher characteristics and qualifications in relation to their VAM scores of teacher effectiveness, with instructional quality and pedagogical content knowledge measures as mediators. From the MET database, teacher characteristic variables include teachers' race and gender. Teacher qualifications include whether or not a teacher has a master's degree or above and their years' of teaching experience in the district.

Measures of instructional quality and pedagogical content knowledge were used as mediators in the model. Scale scores from the Content Knowledge for Teaching assessment, used to measure teachers' pedagogical content knowledge, were collected separately by subject - measured while observing ELA and math classes separately. Instructional quality indicators included in this analysis come from both the CLASS and Framework for Teaching measures of instructional practices. All included indicators were subject specific, therefore teachers' instructional quality scores using these measures differ by subject. The sample size in this study prohibited the inclusion of a measurement model within the MSEM model, thus attempts to reduce measurement error was done through measurement analysis outside of the MSEM framework.

VAM estimates of teacher effects using both state and supplemental assessments in both ELA and math were used separately as dependent variables in each model, for a total of four models. Two models were run using VAM scores based on student state assessment scores measured in grades 4-8 in both ELA and math. MET researchers also administered more cognitively challenging supplemental assessments in reading and math in grades 4-8. VAM estimates based on the Balanced Assessment of Mathematics (BAM) scores were used in an alternative math MSEM model. The SAT-9 reading assessment was used in the creation of supplemental VAM estimates in grades 4-8 and was used in an alternative MSEM model for ELA. The ACT QualityCore end-of-course assessments in English-9 and Algebra I were the only student achievement tests administered to ninth grade students in the MET study, thus, VAM estimates based on these assessments were used in all four models. The Mplus (Muthén, L. & Muthén, B., 2008) SEM software package is used for this analysis.

2.4.1.2 Final Analytic Models

The final analytic models used in this study were conditioned on the quality of the data provided through MET and the ability to run models through the Mplus software. Below, I detail my approaches to the measurement models I used in order to develop an overall measure of instructional quality as well as the MSEM models run for this analysis.

2.4.1.2.1 Measurement Models

Prior to employing the MSEM model, measurement analyses were conducted to reduce the 20 indicators of instructional quality into a more meaningful factor. MET researchers assessed eight different instructional practices using the CLASS measure and twelve different instructional practices using the FFT measure. Data included from the CLASS measures included scales for “Positive Climate”, “Negative Climate”, “Teacher Sensitivity”, “Regard for Student Perspectives”, “Behavior Management”, “Productivity”, “Instructional Learning Formats”, “Content Understanding”, “Analysis and Problem Solving”, “Quality of Feedback”, “Instructional Dialogue” and “Student Engagement”. Data included from the FFT measure included scales for “Creating an Environment of Respect and Rapport”, “Establishing a Culture for Learning”, “Managing Classroom Procedures”, “Managing Student Behavior”, “Communicating with Students”, “Using Questioning and Discussion Techniques”, “Engaging Students in Learning”, and “Using Assessment in Instruction”. These measures were collected both for ELA and for math separately. The FFT measure used a 4.0 scale, while the CLASS measure used a 7.0 scale. Therefore, FFT values for each teacher were converted to a 7.0 scale for this analysis. It should also be noted that the CLASS Negative Climate variable is the only negatively scaled variable in the model. The content specific scores, which were the harmonic means of the researchers’ evaluations scores, were used in each separate ELA and math measurement model.

Using the 20 indicators of instructional quality from the CLASS and FFT measures, I initially conducted an exploratory factor analysis (EFA) to identify the number of integral factors that were underlying the two instructional quality measures. EFA was employed because of the uncertainty surrounding the underlying structure of indicators from the two measures (Browne, 2001) and the potential for stronger evidence of factor structures to emerge during later CFA analyses (Goldberg & Velicer, 2006). Principal axis factor analysis was employed because of its tolerance of multivariate non-normality and its recovery of weak factors (Briggs & MacCallum, 2003; Cudeck, 2000; Fabrigar, Wegener, MacCallum, & Strahan, 1999). Squared multiple correlations allowed for the estimation of communalities, iterated to produce final communality estimates (Gorsuch, 2003). A Promax rotation was employed for both theoretical and empirical reasons, as it was assumed that retained factors would be correlated (Tataryn, Wood, & Gorsuch, 1999).

The correct number of factors to retain and rotate is one of the more critical decisions to be made when conducting an EFA (Fabrigar et al., 1999; Tabachnick & Fidell, 1996). The most common rule is to retain factors when eigenvalues are greater than 1.0. This one criterion is the default procedure in most statistical packages, yet it has a tendency to under- or over-estimate the number of true latent dimensions found in data (Gorsuch, 1983; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). As a result, each model (e.g. ELA and math) was evaluated against the following five rules: (a) eigenvalues greater than 1.0 (Kaiser, 1960), (b) minimum average parcels (MAP; Velicer, 1976), (c) Glorfeld's (1995) expansion of parallel analysis (PA; Horn, 1965), (d) yield high internal consistency (an alpha coefficient of > .70) for unit-weighted factors (Gregory, 2007; Reynolds, Livingston, & Willson, 2006; Salvia Ysseldyke, 2007), and (f) interpretability (Fabrigar et al., 1999; Gorsuch, 1983). Results from several investigations demonstrated that MAP and PA are the two best methods for determining the correct number of factors to accept, followed by the scree test (Buja &

Eyuboglu, 1992; Glorfeld, 1995; Verlicer et al., Zwick & Velicer, 1986). EFA was conducted using the SPSS statistical software.

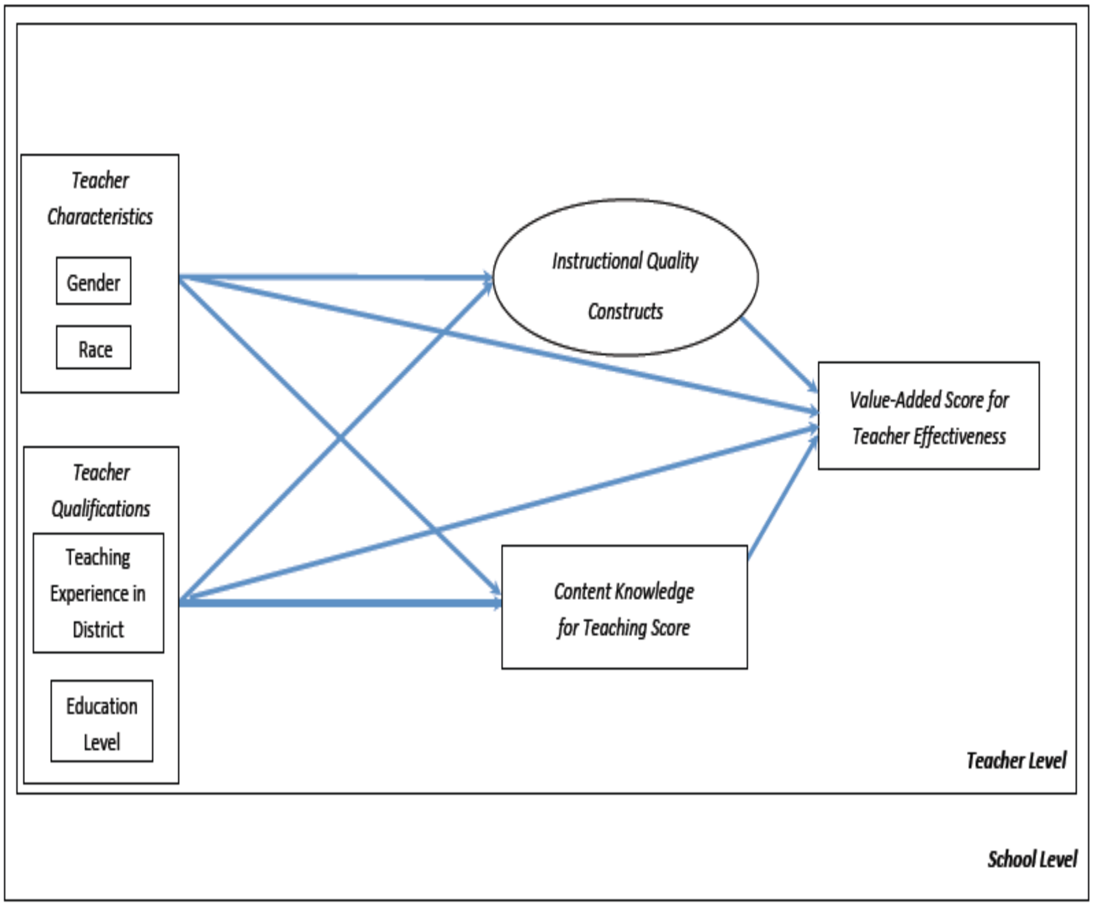
For both ELA and math, a plausible set of three factors was produced in the initial exploratory factor analysis. However, the results indicated that a higher-order factor may exist and thus a confirmatory second-order factor analysis was used to confirm the initial factor model as well as to identify a higher-order factor that provided a more general measure of instructional quality, representing indicators from both the CLASS and FFT measures. Confirmatory second-order factor analysis is the further analysis of the results of an initial factor analysis into a more generalization of the indicators using the correlations between the lower-level, or first-order factors (Marsh & Hocevar, 1988). In this analysis, the simple structure requirement that each measured variable is to load onto only one first-order factor is quite restrictive and the data might fit better if they were allowed to load onto more than one first-order factor, however, such a model may confound method and trait factors (Marsh & Hocevar, 1985). This approach was found most meaningful due to the presence of shared variance, lack of factor interpretability, and the interest in producing a more general, overall measure of instructional quality for this study. Using the factor structure resulting from EFA analyses, secondary confirmatory factor analysis was run to assess the presence of a higher order factor. Several tests of model fit were examined including the chi-square fit statistics which indicate a lack of goodness of fit (Asparouhov & Muthén, 2010); the root mean square error of approximation which measures error of approximation and taking into account sample size (RMSEA) (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kenny, 2014a); the comparative fit index (CFI) which assesses the relative improvement in fit of the model compared to the baseline model (Hu & Bentler, 1998, 1999); and the Standardized Root Mean Residuals (SRMR) which is an absolute measure of fit defined as the standardized difference between the observed correlation and the predicted correlation (Hu & Bentler, 1998,1999; Kenny, 2014a; Kline, 1998). The hypothesized one, higher-order factor model was tested and confirmed using these criteria. Continuous factor scores from the resulting second-order factor, representing an overall measure of instructional quality, were used in the final MSEM model as one overall indicator of instructional quality. Mplus uses the regression method for factor score estimation, also known as the modal posterior estimator, to estimate factor scores (Asparouhov & Muthén, 2010). Secondary confirmatory factor analysis was conducted using the Mplus software.

2.4.1.2.2 MSEM Models

The various relationships of variables that were modeled in this study are provided in Table 2.3 below, followed by a visual representation of the MSEM model for this study in Figure 2 and a description of variables in the following sections.

Table 2.3: Modeled Relationships in MSEM

Gender → Teacher Effectiveness
Race → Teacher Effectiveness
Teaching Experience in District → Teacher Effectiveness
Education Level → Teacher Effectiveness
Pedagogical Content Knowledge → Teacher Effectiveness
Instructional Quality → Teacher Effectiveness
Gender → Pedagogical Content Knowledge → Teacher Effectiveness
Race → Pedagogical Content Knowledge → Teacher Effectiveness
Teaching Experience in District → Pedagogical Content Knowledge → Teacher Effectiveness
Education Level → Pedagogical Content Knowledge → Teacher Effectiveness
Gender → Instructional Quality → Teacher Effectiveness
Race → Instructional Quality → Teacher Effectiveness
Teaching Experience in District → Instructional Quality → Teacher Effectiveness
Education Level → Instructional Quality → Teacher Effectiveness



Note: All models control for grade-level and district differences. Four separate MSEM models are examined based on content area and assessment types used in estimating VAMs.

Figure 2. MSEM Model for Present Study

2.4.1.2.3 Within-School Variables

Four predictors were modeled at the teacher level representing teachers' race, gender, the number of years of teaching experience in the district, and whether or not they held a master's degree or advanced degree. The race variable, "White" is a dummy coded variable that estimates White teachers versus minority teachers, with White as the reference category. The gender variable, "Male", is a dummy coded variable that estimates male teachers versus female teachers, with male as the reference category. The "Master" variable is also included as a dummy variable indicating whether a teacher has a master's degree or any other advanced degrees compared to those who do not. The variable for years of experience teaching in the district is a continuous variable measuring the number of years of teaching experience each teacher has in the district. This variable was highly skewed and thus the natural log of this variable was used in all four models in order to improve its normality.

In all models, I control for the grade-level and district differences by including each grade-level and district as dummy variable covariates in the model. Doing so allows for the estimation of teachers and school differences while considering differences in mediating and outcome variables that may have resulted from collecting or estimating data across the different grades and districts. As grade-level and district results were not used as unique predictors in my model, only as covariates, and to adhere to restricted data-use guidelines, district and grade-level results are not detailed or interpreted in this study.

2.4.1.2.4 Between-School Variables

The school mean for race, gender, master and years of experience in the district variables were estimated and included in each model as between-level indicators. These variables indicate the average number of teachers by each predictor that are present within each school. Including these aggregated variables at the school level allows for the examination of contextual effects. As this model does not include random slopes, contextual effects represent significant relationships between predictors and outcomes above and beyond the individual effect, with the assumption that the effect is fixed (i.e., consistent) across schools.

2.4.1.2.5 Mediators

Measures of teachers' pedagogical content knowledge and instructional quality are used as mediators in the model. The pedagogical content knowledge variable is a continuous variable that uses scores from Content Knowledge for Teaching (CKT) assessments in ELA and math separately. Continuous factor scores represent teachers' overall instructional quality, resulting from measurement analyses, and are also content specific. Continuous factor scores of instructional quality are used as a means to understand where teachers fall on the overall latent construct of instructional quality.

2.4.1.2.6 Dependent Variables

VAM estimates were used as dependent variables in MSEM analyses. Four separate MSEM models were employed in this study. The first model uses VAMs estimated using state standardized ELA assessments in grades 4-8 and the ACT QualityCore end-of-course English-9 for grade 9. The second model examines ELA performance using a supplemental reading assessment, the SAT-9, in grades 4-8 and the ACT QualityCore end-of-course English-9 in grade 9. The first math model uses VAMs estimates using state standardized math assessments in grades 4-8 and the ACT QualityCore end-of-course Algebra I assessment in grade 9. An alternative math model was also examined using the BAM supplemental math assessment in grades 4-8 and the ACT QualityCore end-of-course Algebra I in grade 9. It should be noted that ACT QualityCore end-of-course assessments were the only assessments administered in grade 9, thus these scores were used in all models.

2.4.1.2.7 Centering

All predictor and mediator variables at the individual and between-levels were included as grand-mean centered. By doing so, interpretations of model results will reflect the comparisons of the average teacher or the average school, after controlling for grade-level and district differences. This follows the theoretical inquiry at hand, which is guided by an interest in making improvements in the urban teacher workforce as a whole, thus the choice to grand-mean center at both levels. Furthermore, Hoffman and Gavin (1998) suggest that grand-mean centering reduces the correlation between the intercept and slope estimates across groups which can help alleviate potential level-2 estimation problems due to multicollinearity. The between-level regression coefficients represent the group level relationship between the between-level predictor and the dependent variable, minus the influence of the individual level predictor (Hoffman & Gavin, 1998).

Missing data ranged from 0% to 25% across all variables included in the ELA models and ranged from 1% to 28% in the variables included in the math models. Examination of missing data suggest that the variables with higher rates of missing data were teacher demographic data. These data were found to be missing completely at random or at random and appears to be related to the quality or access to the administrative data provided by the districts.

2.4.1.2.8 Handling of Missing Data

In order to handle the missing data in the model, I included my predictors in the model with missing data. This approach allows Mplus to use Full-Information Maximum Likelihood (FIML) estimation to handle missing data in the model. This method estimates coefficients that maximize the probability that observed values have occurred (Field, 2009). It allows the model to keep its statistical power and preciseness by having a sufficient number of cases on which to run the analysis. In this study, the MLR option for Maximum Likelihood estimation was used as it is best used with the possibility of having robust standard errors.

2.5 Predictive Analyses

Finally, to answer the second research question, "What impact would increases in the number of quality teachers in urban schools, through changes in specific teacher attributes or practices, have on student learning?", I make predictions based on findings from the prior MSEM analyses to estimate the potential impact that increasing levels of significant predictors across the urban teacher workforce has on student learning. I replace the VAM scores of the lowest performing teachers across the workforce with the VAMs of the highest performing 5%, 10%, and 20% of teachers in the workforce as identified by significant attributes resulting from MSEM modeling. The original average VAM across the workforce is then compared to the average after replacing 5%, 10%, or 20% of urban teachers with the highest levels of significant attributes. The change in VAM distribution is interpreted as the effect of improving the teacher workforce.

In order to better interpret the resulting effect sizes and understand policy implications of making improvements in the urban teacher work force, I convert these effect sizes into months of student learning. These conversions are based on Hill and colleagues (2008) analysis, which converts effect sizes into annual gains for better interpretation. Seven nationally normed standardized assessments in ELA and six nationally normed standardized assessments in math were used to make these effect size conversions by grade level. For the purposes of this study, I average these annual effect sizes across grade transitions from fourth through ninth grade, grades analyzed in this study, concluding that a standardized effect size of 0.27 in ELA and 0.34 in math is equivalent to about one year of learning. I convert predicted average effect sizes based on significant attributes into months of student learning, which is interpreted as additional months of learning that may be gained by specific improvements in the teacher workforce. This translation makes findings more meaningful and easily understood for policy implications. However, these analyses are not intended to suggest that hiring and firing decisions should be based solely on VAM scores. This predictive analyses is intended to explore the potential results from changing the composition of the urban teacher workforce based on attributes associated with effectiveness. Research still concludes that VAM scores are not strong enough measures of teachers' impacts on student learning to be used in high-stakes decision-making in schools. It is more likely and appropriate that VAM scores will be used in conjunction with other measures of teachers' knowledge, practices, or qualifications in decisions to hire or retain teachers.

2.6 Summary of Methodology

The methods discussed here are comprehensive and complex enough to draw sound conclusions on the trends of urban teacher characteristics and the relationships between teacher and teaching attributes, and teacher effectiveness based on the framework established for this study.

MSEM allows researchers to test mediation hypotheses within multileveled-clustered data by multiple mediators simultaneously and at all levels (Preacher, et al., 2010). According to the framework for this study, the use of MSEM allows for the study of mediated relationships between teacher characteristics and qualifications, their pedagogical knowledge and instructional practices, and their value-added to student achievement. The multilevel structure of the analysis sufficiently deals with the consistent problems in educational data of nesting and clustering and thus will be able to tease out relationships across different levels (i.e., individual students, teacher or school levels). Structural equation modeling allows me to build on previous studies that consider multiple measures of teacher and teaching characteristics, yet introduces the idea of mediation in which there may be indirect relationships between teacher attributes and student achievement as mediated through other characteristics (i.e., instructional quality, pedagogical content knowledge).

Comprehensive value-added metrics, as included in the MET data, are based on the expected growth of students with similar characteristics selected through a randomization process. Teachers' value-added scores are estimates of their students' learning growth in comparison to the expected growth of similar students in the school. These estimates are developed by subject and grade-level while controlling for prior performance and student demographics, which is an improved value-added model as recommended by researchers (Harris, 2011). By creating estimates that control for variation in students and their performance as well as consider expected growth separately by grade and content, systematic error is greatly reduced allowing for more precise estimates of teacher effects.

Finally, the random assignment of classrooms to teachers allows for the best possible estimates of causal effects on student learning and can be linked to some of the most promising measures of teacher characteristics, qualifications, pedagogical content knowledge, and instructional practices.

Significant attributes of effective urban teachers identified through this analysis are manipulated in predictive analyses of teacher effects on student learning to understand how teacher quality in urban schools might improve if we increased the numbers of teachers with these critical attributes. These effects are translated into months of student learning in order to further understand the impact of increasing levels of significant predictors across the urban teacher workforce. While recent studies using the MET database to examine teacher effectiveness have found weak relationships between teacher or teaching attributes and teacher effectiveness (Gates Foundation, 2010; Polikoff & Porter, 2014), the methods used in this study examine teacher effectiveness from a different angle among a strictly defined sample of urban schools. The strength of this model is its grounding in its conceptual framework which does not isolate teacher or teacher characteristics in the conversation on teacher quality, yet looks at them more comprehensively, examining multiple relationships that might better explain effects on student achievement. This study also restricts the sample to urban schools with high percentages of academically at-risk students, which allows for more definitive conclusions about schools of this type.

Chapter 3

RESULTS

Two research questions guide this study:

1. Which attributes of teacher quality are predictive of urban teacher effectiveness?
2. What impact would improvements in the quality of teachers in urban schools, through changes in specific teacher attributes and practices, have on student learning?

A MSEM approach was used to investigate the first research question, using a sample of teachers from the MET study representative of teachers teaching in urban schools serving the most academically at-risk students. This approach allows for the identification of significant predictors of urban teacher effectiveness. As most measures were collected separately by content area, two separate sets of analyses were conducted for ELA and math for a total of four separate model identifications. To follow, significant results from MSEM analyses were used in regression models to predict the impact of improving the urban teacher workforce by increasing these attributes among urban teachers. Resulting teacher effects were translated into months of student learning in order to understand the potential policy impact of making urban teacher workforce improvements based on MSEM findings. Below, I describe the results from measurement, MSEM and predictive analyses.

3.1 Measurement Analyses

A total of 20 indicators of teacher quality were included in the MET database from the CLASS and FFT frameworks. I conducted measurement analyses among ELA and math samples using exploratory and confirmatory factor analyses in order to reduce these 20 indicators into an overall measure of instructional quality, I.

3.1.1 ELA Measurement Results

Table 3.1 presents means (*M*s) and standard deviations (*SD*s) for the 20 instructional quality indicators from CLASS and FFT measures, ELA only, measures submitted to the EFA. Results from Bartlett's Test of Sphericity (Bartlett, 1954) indicated that the correlation matrix was not random (χ^2 = 9,581.48 *df* = 190, *p* = 0.001). The Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) statistic was 0.956 and well above the 0.60 minimum suggested by Kline (1994). Kaiser's criterion for eignevalues greater than one (i.e., 11.5, 1.7, and 1.0, respectively) and review of the scree plot suggested that three factors should be retained. Results from parallel analysis indicated that two factors should be retained while MAP results indicated that four factors should be retained. The three-factor solution satisfied requirements for simple structure in that all variables show appreciable factor loadings, although three variables loaded on more than one factor (Field, 2005; Tabachnick & Fidell, 2007). Alpha coefficients revealed substantial internal consistency for all three dimensions (0.951, 0.924, and 0.773, respectively), all above the .70 criterion.

Table 3.1: Means and Standard Deviations for ELA Scored Instructional Quality Indicators from CLASS and FFT Measures

Variable	M	SD
CLASS: Positive Climate	4.5	0.7
CLASS: Negative Climate	1.4	0.4
CLASS: Teacher Sensitivity	4.0	0.6
CLASS: Regard for Student Perspectives	3.3	0.6
CLASS: Behavior Management	5.8	0.6
CLASS: Productivity	5.6	0.5
CLASS: Instructional Learning Formats	4.1	0.6
CLASS: Content Understanding	3.8	0.6
CLASS: Analysis and Problem Solving	2.7	0.6
CLASS: Quality of Feedback	3.5	0.7
CLASS: Instructional Dialogue	3.3	0.6
CLASS: Student Engagement	4.8	0.6
FFT: Creating an Environment of Respect and Rapport	4.6	0.7
FFT: Using Questioning and Discussion	3.8	0.7
FFT: Establishing a Culture for Learning	4.4	0.7
FFT: Managing Classroom Procedures	4.7	0.7
FFT: Communicating with Students	4.6	0.6
FFT: Managing Student Behavior	4.8	0.7
FFT: Engaging Students in Learning	4.2	0.7
FFT: Using Assessment in Instruction	3.9	0.7

Note: M = mean, SD = standard deviation, N= 493.

Table 3.2 presents the rotated pattern matrix for the resulting three-factor solution. All coefficients greater than .35 were considered appreciable. The first factor explained 58% of the variance. Items loading onto this factor were "Instructional Dialogue", "Quality of Feedback", Regard for Student Perspectives", "Content Understanding", "Positive Climate", "Teacher Sensitivity", "Instructional Learning Formats", "Analysis and Problem Solving", and "Student Engagement". The second factor explained 9% of the variance and included "Engaging Students in

Learning", "Using Assessment in Instruction", "Using Questioning and Discussion", "Communicating with Student", and "Establishing a Culture for Learning". Finally, items loading onto the third factor were "Behavior Management", "Negative Climate", "Productivity", "Managing Student Behavior", "Managing Classroom Procedures" and "Creating an Environment of Respect and Rapport", explaining 5% of the variance. These factors reflected weak structure interpretability as evidence of instructional formats and establishing classroom culture are evident among the first two factors and there were cross loadings of variables on the second two factors.

Table 3.2: ELA Rotated Pattern Matrix from CLASS and FFT Instructional Quality Indicators

	Factor		
	I	II	II
CLASS: Instructional Dialogue	0.92		
CLASS: Quality of Feedback	0.89		
CLASS: Regard for Student Perspectives	0.81		
CLASS: Content Understanding	0.80		
CLASS: Positive Climate	0.79		
CLASS: Teacher Sensitivity	0.77		
CLASS: Instructional Learning Formats	0.77		
CLASS: Analysis and Problem Solving	0.74		
CLASS: Student Engagement	0.71		
FFT: Engaging Students in Learning		0.79	
FFT: Using Assessment in Instruction		0.82	
FFT: Using Questioning and Discussion		0.87	
FFT: Communicating with Students		0.71	
FFT: Establishing a Culture for Learning		0.71	
CLASS: Behavior Management			1.00
CLASS: Negative Climate (Negatively Scaled)			-0.75
CLASS: Productivity			0.66
FFT: Managing Student Behavior			0.67
FFT: Managing Classroom Procedures		0.41	0.53
FFT: Creating an Environment of Respect and Rapport		0.38	0.51

Note: N=493. Pattern coefficients great than or equal to .35 are considered salient. Items are organized in descending order according to the magnitude of pattern coefficients within factors. Interpretation was simplified through the presentation of only coefficients identified as salient.

Table 3.3, below, presents the strength of relationships among the three factors. The correlation between the first two factors was appreciable ($r = 0.649$) and indicates that 42% of their variance was common (i.e., $r^2 = 0.42$). Correlations between the second and third factor were similar (highest $r = 0.670$), in which 45% of the variance was common. Among the first and third factors, 38% ($r=0.670$) of the variance was common. This evidence of redundancy suggested the presence of a higher-order factor.

Table 3.3: ELA Intercorrelatoins Among the Three Retained Factors

	Factor I	Factor II	Factor III
Factor I	--		
Factor II	.649	--	
Factor III	.617	.670	--

Note: N=493.

Due to the evidence of redundancy across factors, the lack of structural perspective among initial factor loadings, a confirmatory higher-order factor analysis was conducted on these three factors.

In examining model fit indices, results of the best possible second-order factor model gave adequate results. In this model, chi-square fit statistics, which are sensitive to sample size and violations of normality assumptions, indicate a lack of goodness of fit ($\chi^2 = 1225.70$, df=167, $p < 0.001$). The RMSEA was 0.11, although RMSEA is considered good if less than or equal to 0.08 (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kenny, 2014a). The CFI was 0.89, where the CFI criterion is 0.90 for an acceptable model fit (Hu & Bentler, 1998, 1999). Finally, the SRMR, which is considered good if less than or equal to 0.08, was 0.05, indicating a good model fit. Tests were run to identify evidence of quadratic relationships between the factors and the dependent variables, although none were found. Other suggested remedies to reduce the variance around the individual indicators in an effort to more solidly define structures of the factors, were considered including treating the factors as binary variables. None of these remedies strengthened the model fit. Model fit and number of parameters for the second-order factor order model and the first-order factor model were the same, therefore the first-order factor model was not statistically distinguishable from the second-order factor model.

I then examined the strength of the second-order factor. The three latent variables indicated that a good amount of variance was accounted for in the observed variables by the higher-order factor, with significant R^2's's of 0.61 (p<0.001), 0.84 (p<0.001), and 0.78 (p<0.001), respectively. The parameter estimate between the second factor and the higher-order factor was significant with an estimate of 1.27 (p<0.001). The parameter estimate between the third factor and the higher-order factor was significant with an estimate of 0.809 (p<0.001). The estimate for the relationship between the first factor and the higher-order factor is set to 1 by Mplus. The residual or unexplained variances of the first-order factors by the second-order factor were 0.13, 0.06, and 0.04, respectively. Furthermore, each of the first-order factors loaded strongly onto the second-order factor, with significant standardized factor loadings of 0.782 for the first factor, 0.918 for the second factor, and 0.882 for the third factor, suggesting that the second-order factor was useful as an overall measure representing instructional quality to be used in this model.

The theoretical framework for this study assumed evidence of one overall factor of instructional quality to be used as a predictor of instructional quality. The use of this higher-order factor was justified not only by this theoretical assumption but also by the evidence of shared variance among the first-order factors and the significant parameter estimates and loadings of the first-order factors onto the second-order factor. Factor scores from this second-order factor were used in this model and represent an overall measure of instructional quality.

3.1.2 Math Measurement Results

The exact same factors were produced by analyses of the math data. Table 3.4 presents means (*M*s) and standard deviations (*SD*s) for the 20 instructional quality indicators from both the CLASS and FFT. Data included in this analysis were derived from CLASS and FFT scores measured under math instruction only. Using Principal Axis Factoring with a Promax rotation, I used the same criterion for identifying underlying constructs with the ELA data. Results from Bartlett's Test of Sphericity (Bartlett, 1954) indicated that the correlation matrix was not random ($\chi^2 = 8915.75$ $df = 190$, $p = 0.001$). The KMO statistic was 0.945 – above the .60 suggested minimum (Kline, 1994). The three eigenvalues were greater than one (i.e., 10.98, 1.98, 1.2), which under Kaiser's criterion suggests that three factors be retained. Scree plots and MAP analysis results also indicated that three factors be retained, while parallel analysis indicated that two factors be retained. All but two variables showed appreciable factor loadings, thus the three-factor solution satisfied requirements for simple structure (Field, 2005; Tabachnick & Fidell, 2007). Alpha coefficients revealed adequate internal consistency for all three dimensions (respectively, 0.943, 0.909, and 0.758) all above the 0.70 criterion.

Table 3.4: Means and Standard Deviations for Math Scored Instructional Quality Indicators from CLASS and FFT Measures

Variable	M	SD
CLASS: Positive Climate	4.4	0.7
CLASS: Negative Climate	1.5	0.4
CLASS: Teacher Sensitivity	4.0	0.6
CLASS: Regard for Student Perspectives	3.0	0.6
CLASS: Behavior Management	5.7	0.6
CLASS: Productivity	5.6	0.5
CLASS: Instructional Learning Formats	4.1	0.6
CLASS: Content Understanding	3.8	0.6
CLASS: Analysis and Problem Solving	2.6	0.6
CLASS: Quality of Feedback	3.5	0.7
CLASS: Instructional Dialogue	3.2	0.6
CLASS: Student Engagement	4.8	0.6
FFT: Creating an Environment of Respect and Rapport	4.5	0.7
FFT: Using Questioning and Discussion	3.7	0.6
FFT: Establishing a Culture for Learning	4.3	0.7
FFT: Managing Classroom Procedures	4.5	0.7
FFT: Communicating with Students	4.4	0.6
FFT: Managing Student Behavior	4.6	0.7
FFT: Engaging Students in Learning	4.1	0.7
FFT: Using Assessment in Instruction	4.0	0.7

Note: M = mean, SD = standard deviation, N= 469.

The rotated pattern matrix that resulted from the three-factor solution is presented in Table 3.5. The exact same variables loaded onto the three factors as in the ELA measurement analysis. Similar to the findings in the ELA measurement analyses, the first factor explained 55% of the variance and included “Instructional Dialogue”, “Quality of Feedback”, Regard for Student Perspectives”, “Content Understanding”, “Positive Climate”, “Teacher Sensitivity”, Instructional Learning Formats”, “Analysis and Problem Solving”, and “Student Engagement”. The second factor, which explained 10% of the variance, included the items "Engaging Students in Learning", "Using Assessment in Instruction", "Using Questioning and Discussion", "Communicating with Students", and "Establishing a Culture for Learning" and explained 10% of the variance. Finally, variables in the third factor, which explained 6% of the variance, were "Behavior Management", "Negative Climate", "Productivity", "Managing Student Behavior", "Managing Classroom Procedures" and "Creating an Environment of Respect and Rapport". Once again, these variable loadings gave weak factor interpretability with practices related to instructional formats and classroom culture loading on to the first two factors and cross-loadings on the second two factors.

Table 3.5: Math Rotated Pattern Matrix from CLASS and FFT Instructoinal Quality Indicators

	Factor		
	I	II	II
CLASS: Instructional Dialogue	0.96		
CLASS: Regard for Student Perspectives	0.88		
CLASS: Analysis and Problem Solving	0.87		
CLASS: Quality of Feedback	0.86		
CLASS: Instructional Learning Formats	0.77		
CLASS: Teacher Sensitivity	0.73		
CLASS: Content Understanding	0.73		
CLASS: Positive Climate	0.73		
CLASS: Student Engagement	0.67		
FFT: Engaging Students in Learning		0.81	
FFT: Using Assessment in Instruction		0.81	
FFT: Establishing a Culture for Learning		0.68	
FFT: Using Questioning and Discussion		0.85	
FFT: Communicating with Students		0.70	
CLASS: Behavior Management			1.03
CLASS: Productivity			0.74
CLASS: Negative Climate (Negatively Scaled)			-0.73
FFT: Managing Student Behavior			0.72
FFT: Managing Classroom Procedures		0.43	0.58
FFT: Creating an Environment of Respect and Rapport		0.43	0.53

Note: N=469. Pattern coefficients great than or equal to .35 are considered salient. Items are organized in descending order according to the magnitude of pattern coefficients within factors. Interpretation was simplified through the presentation of only coefficients identified as salient.

Table 3.6 presents the strength of relationships among the three factors. The correlation between the first two factors was appreciable ($r = 0.606$) and indicated that 37% of the variance was common. The correlation between the first and third factor was 0.580 (34% shared variance), while the correlation with the second factor and the third factor was 0.643 (41% shared variance). Once again, there appeared to be evidence of redundancy, suggesting that a higher-order factor may exist. Thusly, second-order confirmatory factor analysis of these three factors was conducted.

Table 3.6: Math Intercorrelatoins Among the Three Retained Factors

	Factor I	Factor II	Factor III
Factor I	--		
Factor II	.606	--	
Factor III	.580	.643	--

Note: N=469.

I used confirmatory secondary factor analysis of the best possible model for the identification of a second-order factor. In examining model fit indices, results of a second-order factor model gave adequate results. In this model, chi-square fit statistics were significant (χ^2 = 1384.07, *df*=167, $p < 0.001$). The RMSEA was 0.13, CFI was 0.86, and the SRMR was 0.06. Due to the variance found around the individual instructional quality indicators, other ways of structuring the resulting factors, including the use of binary or quadratic factors, were considered but did not improve model fit. Model fit and number of parameters for the second-order factor order model and the first-order factor model were the same, therefore a first-order factor model was not statistically distinguishable from the second-order factor model.

I then examined the strength of the second-order factor. The r-squared's of the three latent variables indicated that a good amount of variance was accounted for in the observed variables by the higher-order factor, with significant R^2's of 0.60 ($p<0.001$), 0.81 ($p<0.001$), and 0.70 ($p<0.001$), respectively. The parameter estimate between the second factor and the higher order factor was significant with an estimate of 1.04 ($p<0.001$). The parameter estimate between the third factor and the higher-order factor was significant with an estimate of 1.00 ($p<0.001$). The estimate for the relationship between the first factor and the higher-order factor is set to 1 by Mplus. The residual or unexplained variances of the first-order factors by the second-order factor were 0.12, 0.05, and 0.08, respectively. Furthermore, each of the first-order factors loaded strongly onto the second-order factor, with loadings of 0.773 for the first factor, 0.901 for the second factor, and 0.839 for the third factor, suggesting that the second-order factor was useful as an overall measure representing instructional quality to be used in this model.

Thusly, I proceeded with using a second-order factor, just as in the ELA analyses, as EFA results suggested that a higher-order factor may exist, and CFA results confirmed that there was indeed a higher order factor to consider. The evidence of shared variance among the first-order factors, the theoretical assumption of an overall higher-order factor, identical model fit indices, and the significant parameter estimates and loadings of the first-order factors onto the second-order factor all justified the use of this second-order factor in this model. Factor scores from the second-order factor were used to represent an overall measure of instructional quality.

The second-order factor model from both ELA and math measurement analysis is presented in Figure 3 below.

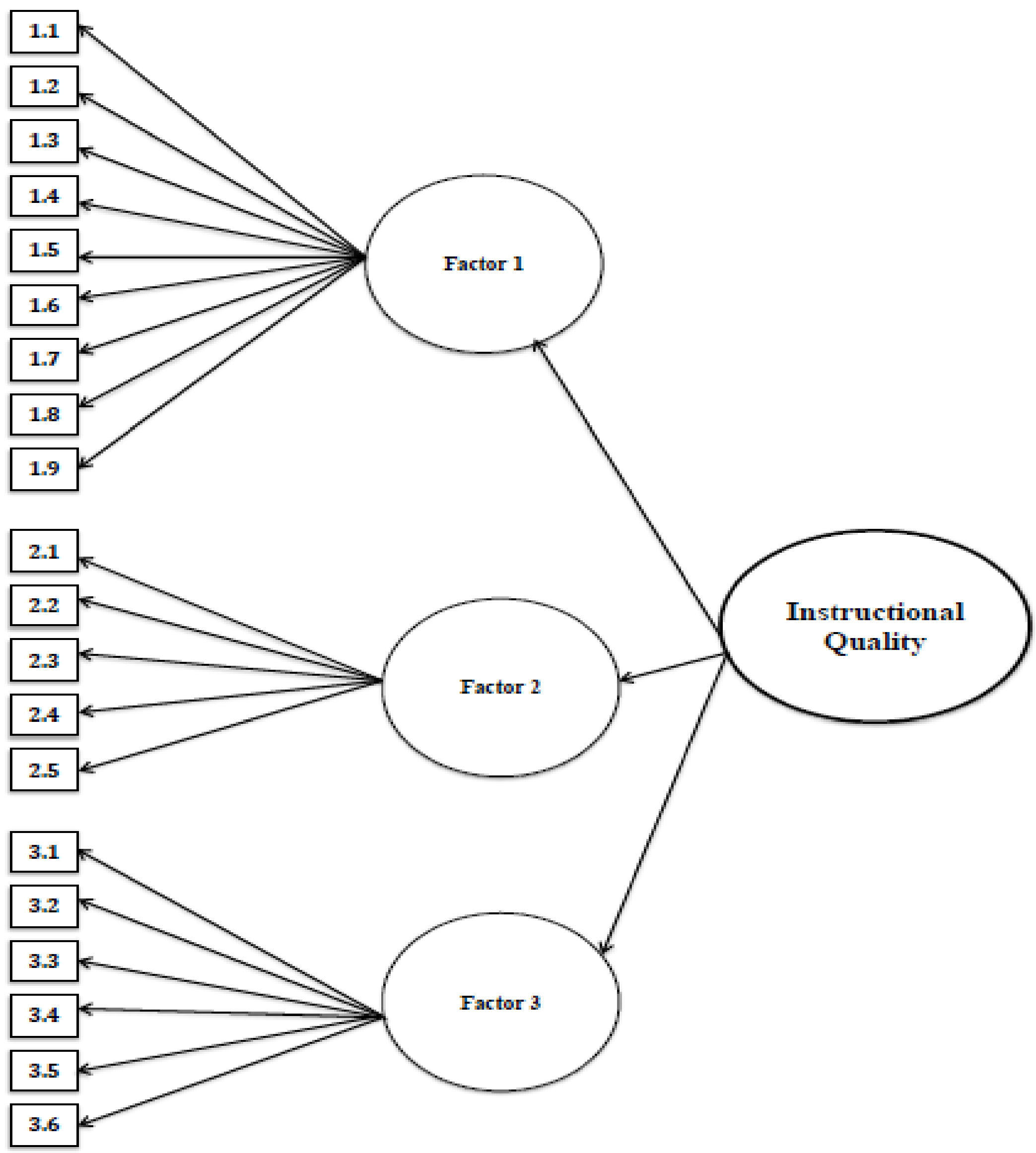


Figure 3. Second-Order Factor Analyses Results for Both ELA and Math.

3.2 MSEM Results

I employed a two-level random intercepts MSEM model to examine predictors at the teacher and school levels as well as mediating relationships between predictors and teachers' value-added to student learning. The multilevel structural equation model combined both within-school and between-school model into one unified model with the assumption that the intercepts and means vary across teachers on the within-level and schools on the between-level. It should be noted that cluster sizes (e.g., 3.25 – ELA; 3.26 – math) were too small for random effects to be estimated. This is confirmed by other research, which states that small cluster sizes limit the power to test for random slope variances at the school level (Preacher, et al., 2010; Snijders, 2005). Therefore all school-level outputs reflect fixed parameters – the effect above and beyond the individual (teacher) effect, assuming that the effect is fixed across schools. Four separate models were run depending on content area assessed (i.e., ELA, math) and assessment type used in estimating VAMs (i.e., state standardized assessments, SAT-9 supplemental reading assessment, BAM supplemental math assessment).

3.2.1 ELA Results (State Assessment VAMs)

The first MSEM model examined teacher attributes in relation to VAM measures based on ELA state standardized assessments in grades 4-8 and the ACT QualityCore end-of-course English-9 assessment in grade 9. I refer to this model as the ELA model using state assessment VAMs for simplicity purposes as all four models use ACT QualityCore end-of-course assessments in grade 9.

3.2.1.1 Model Fit Indices

The overall model fit indices for this model gave sound results. The intra-class correlation (ICC) for the dependent variable measuring teachers' value-added to student learning using state assessment data was 0.243; instructional quality was 0.103, and for pedagogical content knowledge was 0.025. The ICC is a measure of the proportion of variance in the outcome that is between groups or schools (Raudenbush & Bryk, 2002). In this model, all three ICCs fell within an acceptable range for concluding that the data were not independent and multilevel modeling was necessary (Bliese, 2000; Cheung & Au, 2005; Cheung, Leung & Au, 2006; Muthén, 1994). The SRMR for the within-group part of the model at the teacher level of analysis was 0.06, and the SRMR for the between-group part of the model was 0.02, with 0.10 as the conventional value for SRMR. I n addition, the model produced a non-significant chi-square result (χ^2=43.250, *df*=40, *p*=0.33), which indicated good model fit. The CFI for this model was 0.964, above the conventional cutoff point of 0.90, which also indicated good model fit. Finally, the RMSEA was 0.013, which was smaller than the conventional value of 0.08, demonstrating good model fit (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kenny, 2014a). The means, standard deviations and correlations between variables for this specific model are included in Appendix B.

Below I detail direct, indirect, and contextual effects resulting from this analysis. All results related to teacher effects reflect the use of state standardized ELA assessments in VAM estimates in grades 4-8 and the ACT QualityCore end-of-course English-9 assessment in grade 9.

3.2.1.2 Within-Level Direct Effects

The main inquiry of this study was on understanding significant predictors of teachers' value-added to student learning. Model results indicated that in ELA, only two predictors significantly estimated this relationship – race and instructional quality. MSEM output signified that the average White teacher in urban schools has a lower effect on student's ELA learning than the average minority teacher in urban schools (β= -0.150, p<0.05), while controlling for grade-level and district differences.

Therefore, students of the average minority urban teacher score 0.15 standard deviations above students of the average White urban teacher. Teachers' overall instructional quality was also found to be a significant predictor of teachers' value-added to student learning of ELA in urban schools (β=0.180, p<0.01), controlling for grade-level and district differences. A one standard deviation increase in the average teachers' instructional quality score is associated with a 0.18 standard deviation increase in students' learning of ELA. Teachers' gender, level of education, and years of experience were not significant predictors of urban teachers' value-added to student learning.

I also examined the direct relationship between predictors and mediators, although this was not the core inquiry of this study. Results indicated that race, gender, level of education, and the number of years of teaching experience in the district were not significant predictors of teachers' instructional quality in urban schools, while controlling for grade-level and district differences.

Race was the only significant predictor of teachers' pedagogical content knowledge in ELA. The average White urban teacher has significantly higher scores of content knowledge for teaching ELA above and beyond the average minority urban teacher (β=0.265, p<0.001), controlling for grade-level and district differences – a difference of 0.27 standard deviations. Table 3.7 below provides unstandardized and standardized estimates of the direct relationships between predictors, mediators, and outcomes at the teacher-level.

Table 3.7: Direct Effects of Within-School Model (ELA State Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=498

	Unstandardized	Standardized
VAM (ELA using State Assessments)		
Male	-0.048 (0.04)	-0.077 (0.06)
White	-0.062 (0.03)	-0.150 (0.06)*
Years Exp. In District	0.025 (0.02)	0.091 (0.06)
Master	-0.058 (0.03)	-0.141 (0.07)
Instructional Quality (ELA)	0.093 (0.04)	0.180 (0.07)**
Ped. Content Knowledge (ELA)	0.000 (0.00)	0.011 (0.05)
Instructional Quality (ELA)		
Male	-0.001 (0.06)	-0.001 (0.05)
White	-0.017 (0.05)	-0.021 (0.06)
Years Exp. In District	-0.018 (0.03)	-0.033 (0.06)
Master	-0.021 (0.05)	-0.026 (0.06)
Pedagogical Content Knowledge (ELA)		
Male	0.159 (1.99)	0.005 (0.06)
White	5.814 (1.52)	0.265 (0.07)***
Years Exp. In District	-0.301 (0.96)	-0.021 (0.07)
Master	0.978 (1.40)	0.044 (0.06)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.1.3 Contextual Effects

At the school-level, the average urban school with high percentages of teachers with master or advanced degrees was more effective at improving student learning than the average urban school with fewer teachers of this type, while controlling for differences at the teacher level (β=0.905, p<0.01). The difference between the two school types was 0.905 standard deviations in teacher effects on student learning.

Instructional quality also had a significant contextual effect. The average urban school with higher levels of instructional quality among their teachers is associated with higher levels of teacher effectiveness than the average urban school lower levels of instructional quality (β=0.540, p<0.05), while controlling for teacher-level differences. This corresponds to a difference in teacher effects of 0.54 standard deviations. Urban schools with higher rates of White or male teachers, or teachers with more years of teaching experience in the district do not significantly differ from urban schools with lower rates of teachers having these characteristics in terms of their value-added to student learning.

When examining the direct relationships between school-level predictors and mediators, one significant predictor was identified. The average urban school with higher rates of female teachers has significantly higher scores of instructional quality than the average urban school with more male teachers (β= -0.467, p<0.05), above and beyond the individual effect – a difference of 0.467 standard deviations. Urban schools having a predominance of White teachers, teachers with advanced degrees or teachers with more experience within their districts were not predictive of teachers' levels of instructional quality. Furthermore, no school-level predictors were associated with significant differences in urban teachers' pedagogical content knowledge in ELA. Below, Table 3.8 provides unstandardized and standardized estimates of these direct relationships between predictors, mediators, and outcomes at the school-level.

Table 3.8: Direct Effects of Between-Level Model (ELA State Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=498

	Unstandardized	Standardized
VAM (ELA using State Assessments)		
% Male	0.147 (0.12)	0.300 (0.25)
% White	0.104 (0.06)	0.338 (0.2)
Avg. Years Exp. In District	0.005 (0.01)	0.154 (0.34)
% Master	0.278 (0.10)	0.905 (0.27)**
Avg. Instructional Quality (ELA)	0.470 (0.26)	0.540 (0.25)*
Avg. Ped. Content Knowledge (ELA)	0.013 (0.03)	0.229 (0.51)
Instructional Quality (ELA)		
% Male	-0.263 (0.11)	-0.467 (0.20)*
% White	-0.008 (0.09)	-0.022 (0.26)
Avg. Years Exp. In District	-0.001 (0.01)	-0.032 (0.23)
% Master	-0.053 (0.11)	-0.150 (0.30)
Pedagogical Content Knowledge (ELA)		
%Male	-1.214 (2.98)	-0.145 (0.36)
%White	0.529 (2.40)	0.101 (0.45)
Avg. Years Exp. In District	-0.287 (0.17)	-0.538 (0.36)
% Master	-2.082 (2.32)	-0.397 (0.44)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.1.4 Mediating Effects

Results from this model indicated that no predictors (i.e., race, gender, advanced degrees, years of teaching experience in the district) had significant relationships with VAM that were mediated by either teachers' instructional quality or their pedagogical content knowledge in ELA, while controlling for grade-level and district differences. This was also the case when examining school-level predictors, controlling for teacher-level effects. Table 3.9 displays results from mediated analyses.

Table 3.9: Mediating (Indirect) Effects (ELA State Assessment VAMs)
Unstandardized Estimates (Standard Errors in Parentheses); N=498

Within (Teacher) Level	
Male→ Instruct. Quality (ELA)→ VAM (ELA)	0.000 (0.01)
White→ Instruct. Quality (ELA)→ VAM (ELA)	-0.002 (0.01)
Years Exp. In District → Instruct. Quality (ELA)→ VAM (ELA)	-0.002 (0.00)
Master→ Instruct. Quality (ELA)→ VAM (ELA)	-0.002 (0.00)
Male→ Ped. Content Know. (ELA)→ VAM (ELA)	0.000 (0.00)
White→ Ped. Content Know. (ELA)→ VAM (ELA)	0.001 (0.01)
Years Exp. In Dist.→ Ped. Content Know.(ELA)→ VAM (ELA)	0.000 (0.00)
Master→ Ped. Content Know. (ELA)→ VAM (ELA)	0.000 (0.00)
Between (School) Level	
% Male→ Instruct. Quality (ELA)→ VAM (ELA)	-0.124 (0.09)
% White→ Instruct. Quality (ELA)→ VAM (ELA)	-0.004 (0.04)
Avg. Years Exp. In District → Instruct. Quality (ELA)→ VAM (ELA)	-0.001 (0.00)
% Master→ Instruct. Quality (ELA)→ VAM (ELA)	-0.025 (0.05)
% Male→ Ped. Content Know. (ELA)→ VAM (ELA)	-0.016 (0.06)
% White→ Ped. Content Know. (ELA)→ VAM (ELA)	0.007 (0.03)
Avg. Yrs Exp. In Dist.→ Ped. Cont. Know. (ELA)→VAM (ELA)	-0.004 (0.01)
% Master→ Ped. Content Know. (ELA)→ VAM (ELA)	-0.028 (0.07)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.2 ELA Results (Supplemental Reading VAMs)

Similar to the previous model, an MSEM model was run among the ELA sample using ELA indicators, while this time using ELA supplemental assessments rather than state standardized assessments. VAMs used in this model were estimated based on SAT-9 supplemental reading assessment scores in grades 4-8 and ACT QualityCore end-of-course English-9 assessment scores in grade 9 to see examine any differences or similarities in outcomes compare to the previous model.

3.2.2.1 Model Fit Indices

The overall model fit indices for this model using supplemental reading assessment VAMs gave good model fit. For this model, the school mean variable for years of teaching experience in the district was transformed with the variable representing the square root of its original values in order to improve model fit. The model produced a non-significant chi-square result (χ^2=19.604, *df*=40, *p*=0.997). The SRMR at the teacher level of analysis was 0.06, and the SRMR at the school-level of the model was 0.02, both indicating good model fit. The CFI for this model was 1.0 and the RMSEA was 0.000. The intra-class correlation was 0.216 for the dependent variable measuring teachers' value-added to student learning using SAT-9 assessment data; 0.097 for instructional quality; and 0.028 for pedagogical content knowledge, signifying the need for multilevel analysis. Appendix C includes means, standard deviations and correlations between variables from this model.

3.2.2.2 Within-Level Effects

Instructional quality (β=0.218, p<0.001) was once again found to be a significant predictor of urban ELA teachers' value-added to student learning as measured by SAT-9 supplemental reading assessments, while controlling for grade-level and district differences. A one standard deviation increase in the average urban ELA teachers' instructional quality is associated with a 0.22 standard deviation increase in teacher effects on reading. Urban teachers' race, advanced degrees, or years of experience teaching within their districts were not found to be predictive of urban ELA teachers' value-added to student learning.

There were once again no significant predictors of instructional quality among ELA teachers identified in this model. Race remained a significant predictor of urban teachers' content knowledge for teaching ELA (β=0.273, p<0.001). This result indicates that the average White urban teacher has scores 0.27 standard deviations higher on assessments of ELA pedagogical content knowledge than the average minority urban teacher, controlling for grade-level and district differences. Teachers' gender, advanced degrees, or years of teaching experience in their district were not significant predictors of teachers' pedagogical content knowledge in ELA. The unstandardized and standardized estimates of these direct relationships between predictors, mediators, and outcomes at the teacher-level are presented in Table 3.10 below.

Table 3.10: Direct Effects of Within-Level Model (SAT-9 Reading Supplemental Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=498

	Unstandardized	Standardized
VAM (ELA using SAT-9 Reading Assessments)		
Male	-0.104 (0.06)	-0.122 (0.07)
White	0.017 (0.04)	0.030 (0.07)
Years Exp. In District	0.016 (0.02)	0.043 (0.05)
Master	0.001 (0.05)	0.002 (0.09)
Instructional Quality (ELA)	0.152 (0.04)	0.218 (0.06)***
Ped. Content Knowledge (ELA)	0.000 (0.00)	-0.016 (0.06)
Instructional Quality (ELA)		
Male	0.001 (0.06)	0.001 (0.05)
White	-0.016 (0.05)	-0.020 (0.06)
Years Exp. In District	-0.018 (0.03)	-0.034 (0.06)
Master	-0.010 (0.05)	-0.012 (0.06)
Pedagogical Content Knowledge (ELA)		
Male	0.115 (1.99)	0.003 (0.06)
White	5.994 (1.51)	0.273 (0.07)***
Years Exp. In District	-0.233 (0.99)	-0.016 (0.07)
Master	1.204 (1.40)	0.055 (0.06)

Note:, **p*<.05, ** *p*<.01, ****p*<.001. This model controlled for grade-level and district differences.

3.2.2.3 Contextual Effects

When considering contextual effects, no school-level predictors were significantly predictive of teacher effects across schools. This was unlike results from the previous ELA model, which found that the average urban school with more teachers with advanced degrees and the average urban school with more teachers of high instructional quality were more effective at improving student learning.

However, when examining their direct relationship with mediators, the average urban school with higher rates of female teachers had significantly higher levels of instructional quality among its teachers than schools with more male ELA teachers ($\beta = -0.483$, $p < 0.05$) – a difference of 0.48 standard deviations, after controlling for teacher level effects. This relationship was also found to be significant in the previous ELA model. No school-level predictors were directly associated with significant differences in urban teachers' ELA pedagogical content knowledge. This result was similar to results from the previous ELA model. Table 3.11 below provides unstandardized and standardized estimates of the direct relationships between predictors, mediators, and outcomes at the school-level.

Table 3.11: Direct Effects of Between-Level Model (SAT-9 Reading Supplemental Assessment VAMs)
Unstandardized Estimates (Standard Errors in Parentheses); N=498

	Unstandardized	Standardized
VAM (ELA using SAT-9 Reading Assessments)		
% Male	0.018 (0.27)	0.029 (0.44)
% White	-0.022 (0.12)	-0.058 (0.31)
Avg. Years Exp. In District	-0.042 (0.11)	-0.230 (0.58)
% Master	-0.033 (0.35)	-0.085 (0.91)
Avg. Instructional Quality (ELA)	0.369 (0.54)	0.332 (0.47)
Avg. Ped. Content Knowledge (ELA)	-0.034 (0.07)	-0.500 (0.97)
Instructional Quality (ELA)		
% Male	-0.264 (0.1)	-0.483 (0.20)*
% White	-0.017 (0.09)	-0.049 (0.26)
Avg. Years Exp. In District	0.000 (0.04)	-0.002 (0.26)
% Master	-0.101 (0.12)	-0.289 (0.34)
Pedagogical Content Knowledge (ELA)		
%Male	-1.389 (2.96)	-0.156 (0.33)
%White	0.039 (2.37)	0.007 (0.42)
Avg. Years Exp. In District	-1.52 (0.88)	-0.568 (0.32)
% Master	-3.134 (2.40)	-0.549 (0.42)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.2.4 Mediating Effects

Among the modeled mediated relationships tested in this model, none were significantly predictive of teachers' value added to student learning. These results mirror results from the previous ELA model which also found no signficant mediating effects. The results from mediated relationships are presented in Table 3.12 below.

Table 3.12: Mediating (Indirect) Effects (SAT-9 Reading Supplemental Assessment VAMs) Unstandardized Estimates (Standard Errors in Parentheses); N=498

Within (Teacher) Level	
Male→ Instruct. Quality (ELA)→ VAM (ELA)	0.000 (0.01)
White→ Instruct. Quality (ELA)→ VAM (ELA)	-0.002 (0.01)
Years Exp. In District → Instruct. Quality (ELA)→ VAM (ELA)	-0.003 (0.01)
Master→ Instruct. Quality (ELA)→ VAM (ELA)	-0.002 (0.01)
Male→ Ped. Content Know. (ELA)→ VAM (ELA)	0.000 (0.00)
White→ Ped. Content Know. (ELA)→ VAM (ELA)	-0.002 (0.01)
Years Exp. In District→ Ped. Cont. Know. (ELA)→ VAM (ELA)	0.000 (0.00)
Master→ Ped. Content Know. (ELA)→ VAM (ELA)	0.000 (0.00)
Between (School) Level	
% Male→ Instruct. Quality (ELA)→ VAM (ELA)	-0.098 (0.15)
% White→ Instruct. Quality (ELA)→ VAM (ELA)	-0.006 (0.03)
Avg. Yrs Exp. In District → Inst. Quality (ELA)→ VAM (ELA)	0.000 (0.02)
% Master→ Instruct. Quality (ELA)→ VAM (ELA)	-0.037 (0.07)
% Male→ Ped. Content Know. (ELA)→ VAM (ELA)	0.047 (0.14)
% White→ Ped. Content Know. (ELA)→ VAM (ELA)	-0.001 (0.08)
Avg. Yrs Exp. In Dist.→ Ped. Cont. Know. (ELA)→ VAM (ELA)	0.052 (0.11)
% Master→ Ped. Content Know. (ELA)→ VAM (ELA)	0.107 (0.25)

Note: $^{\pm}$ p<.10, *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.3 Math Results (State Assessment VAMs)

In this MSEM model, the dependent variable consisted of math VAM values estimated using math state standardized assessments in grades 4-8 and the ACT QualityCore Algebra I end-of-course assessment in grade 9. Below I discuss its model fit as well as results of direct and indirect relationships.

3.2.3.1 Model Fit Indices

The overall model fit indices for the final math model gave sound results. The model produced a non-significant chi-square result (χ^2=18.964, *df*=40, *p*=0.998), which indicated a relatively good model fit. The SRMR for the within-group part of the model was 0.06, indicating good model fit, while the SRMR for the between-group part of the model was 0.01. The CFI was 1.00, above the conventional value of 0.90; and the RMSEA was 0.000 – smaller than the conventional value of .05, demonstrating a good model fit. ICCs demonstrate the need for multilevel modeling with an ICC of 0.241 for the dependent variable measuring teachers' value-added to student learning, 0.095 for instructional quality, and 0.072 for pedagogical content knowledge in math. The means, standard deviations, and correlations between variables from this model are included in Appendix D.

3.2.3.2 Within-Level Direct Effects

As in both ELA models, instructional quality was the only significant predictor of teachers' value-added to student learning in math (β= 0.186, p<0.05), controlling for grade-level and district differences. A one standard deviation increase in the average urban math teachers' instructional quality score is associated with a 0.18 standard deviation increase in teachers' value-added to student learning in math.

While teacher effects were the main focus of this study, it was also interesting to examine other significant relationships resulting from this analysis. While there were no significant predictors of urban math teachers' instructional quality, race was once again a significant predictor of teachers' pedagogical content knowledge (β= 0.288, p<0.001), while controlling for grade-level and district differences. The average White urban teacher performs significantly higher on math pedagogical content knowledge assessments compared to the average minority urban teacher, a difference of 0.29 standard deviations. The number of years of teaching experience in the district was also a significant predictor of teachers' math pedagogical content knowledge (β= -0.129, p<0.05), controlling for grade-level and district difference. The average urban math teacher with fewer years of teaching experience in the district is associated with a 0.13 standard deviation increase in math pedagogical content knowledge. Table 3.13 below provides unstandardized and standardized estimates of these direct relationships between predictors, mediators, and outcomes at the teacher-level.

Table 3.13: Direct Effects of Within-School Model (Math State Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=475

	Unstandardized	Standardized
VAM (Math using State Assessments)		
Male	0.007 (0.04)	0.011 (0.06)
White	-0.018 (0.03)	-0.038 (0.07)
Years Exp. In District	0.016 (0.02)	0.054 (0.07)
Master	0.004 (0.03)	0.009 (0.06)
Instructional Quality (Math)	0.116 (0.05)	0.186 (0.07)*
Ped. Content Knowledge (Math)	0.001 (0.00)	0.089 (0.07)
Instructional Quality (Math)		
Male	-0.105 (0.06)	-0.111 (0.06)
White	-0.030 (0.05)	-0.040 (0.07)
Years Exp. In District	-0.012 (0.03)	-0.025 (0.05)
Master	-0.015 (0.06)	-0.020 (0.09)
Pedagogical Content Knowledge (Math)		
Male	-0.254 (1.57)	-0.007 (0.05)
White	8.125 (1.86)	0.288 (0.06)***
Years Exp. In District	-2.305 (1.02)	-0.129 (0.06)*
Master	0.331 (2.01)	0.012 (0.07)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.3.3 Contextual Effects

Schools' composition by teachers' race, gender, years of experience in their district, and having of advanced degrees did not have contextual effects on student learning or in relation to any mediators. The results of the direct relationships between predictors, mediators, and outcomes at the school-level are presented in Table 3.14 below, with standardized and unstandardized estimates.

Table 3.14: Direct Effects of Between-Level Model (Math State Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=475

	Unstandardized	Standardized
VAM (Math using State Assessments)		
% Male	-0.023 (0.09)	-0.055 (0.22)
% White	0.118 (0.07)	0.342 (0.19)
Avg. Years Exp. In District	-0.002 (0.01)	-0.052 (0.21)
% Master	0.195 (0.11)	0.568 (0.29)
Avg. Instructional Quality (ELA)	0.490 (0.33)	0.455 (0.28)
Avg. Ped. Content Knowledge (ELA)	0.002 (0.01)	0.057 (0.30)
Instructional Quality (Math)		
% Male	-0.035 (0.10)	-0.09 (0.25)
% White	-0.004 (0.09)	-0.014 (0.27)
Avg. Years Exp. In District	0.006 (0.01)	0.186 (0.18)
% Master	-0.123 (0.12)	-0.386 (0.37)
Pedagogical Content Knowledge (Math)		
%Male	2.769 (3.43)	0.223 (0.27)
%White	-1.692 (2.91)	-0.166 (0.28)
Avg. Years Exp. In District	0.122 (0.24)	0.120 (0.23)
% Master	-1.616 (3.49)	-0.159 (0.34)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.3.4 Mediating Effects

No significant mediating effects were identified between teachers' characteristics and qualifications, and instructional quality or pedagogical content knowledge as mediators, in relation to teachers' value-added to learning math on the state assessment. This also holds true at the school-level. Table 3.15 below displays results from mediated analyses estimated in this model.

Table 3.15: Mediating (Indirect) Effects (Math State Assessment VAMs)
Unstandardized Estimates (Standard Errors in Parentheses); N=475

Within (Teacher) Level	
Male→ Instruct. Quality (Math)→ VAM (Math)	-0.004 (0.01)
White→ Instruct. Quality (Math)→ VAM (Math)	-0.012 (0.01)
Years Exp. In District → Instruct. Quality (Math)→ VAM (Math)	-0.001 (0.00)
Master→ Instruct. Quality (Math)→ VAM (Math)	-0.002 (0.01)
Male→ Ped. Content Know. (Math)→ VAM (Math)	0.000 (0.00)
White→ Ped. Content Know. (Math)→ VAM (Math)	0.012 (0.01)
Years Exp. In District→ Ped. Cont. Know. (Math)→ VAM (Math)	-0.003 (0.00)
Master→ Ped. Content Know. (Math)→ VAM (Math)	0.000 (0.00)
Between (School) Level	
% Male→ Instruct. Quality (Math)→ VAM (Math)	-0.017 (0.05)
% White→ Instruct. Quality (Math)→ VAM (Math)	-0.002 (0.04)
Avg. Years Exp. In District → Inst. Quality (Math)→ VAM (Math)	0.003 (0.00)
% Master→ Instruct. Quality (Math)→ VAM (Math)	-0.060 (0.08)
% Male→ Ped. Content Know. (Math)→ VAM (Math)	0.005 (0.03)
% White→ Ped. Content Know. (Math)→ VAM (Math)	-0.003 (0.02)
Avg. Yrs. Exp. In Dist.→ Ped. Cont. Know. (Math)→ VAM (Math)	0.000 (0.00)
% Master→ Ped. Content Know. (ELA)→ VAM (Math)	-0.003 (0.02)

Note: **p*<.05, ** *p*<.01, ****p*<.001. This model controlled for grade-level and district differences.

3.2.4 Math Results (Supplemental Math VAMs)

The final MSEM model used VAM scores based on BAM supplemental assessment in grades 4-8 and the ACT QualityCore end-of-course Algebra I assessments in Grade 9 as its outcome variable. Results from this model were compared to results from the math model using VAM estimates based on the state standardized assessments. The model fit indices and results from direct, indirect and mediating effects are presented below.

3.2.4.1 Model Fit Indices

The overall model indicated good model fit. The model produced a non-significant chi-square result (χ^2=34.336, *df*=40, *p*=0.723); a SRMR on the within-group level of 0.06; a SRMR on the between-group level of 0.01; a CFI of 1.00; and an RMSEA of 0.000, all demonstrating a good model fit. The intra-class correlation for the dependent variable measuring teachers' value-added to student learning was 0.209, 0.94 for instructional quality, and 0.073 for math pedagogical content knowledge, all demonstrating that multilevel modeling was necessary. Appendix E includes means, standard deviations and correlations between variables from this model.

3.2.4.2 Within-Level Effects

At the teacher-level, instructional quality (β= 0.156, p<0.05) was positively associated with teachers' value-added to student learning in math. Thus, a one standard deviation increase in the average urban math teachers' instructional quality is associated with a 0.16 standard deviation increase in teacher effects. This finding was similar to results from the previous math model, although one additional significant predictor was found in this model. Urban math teachers' pedagogical content knowledge was also associated with larger teacher effects (β= 0.160, p<0.05). A one standard deviation increase in the average urban math teachers' math pedagogical content knowledge is associated with 0.16 standard deviation increase in teacher effects.

Being a White teacher (β= 0.291, p<0.001) was significantly predictive of urban math teachers' pedagogical content knowledge with the average White urban teachers scoring 0.29 standard deviations above the average minority urban teacher.

The number of years of teaching experience in the district was also a significant predictor of urban math teachers' pedagogical content knowledge (β= -0.124, p<0.05). A one standard deviation increase in urban math teachers' years of experience in the district is associated with a 0.12 standard deviation decrease in their pedagogical content knowledge. Table 3.16 below provides unstandardized and standardized estimates of these direct relationships between predictors, mediators, and outcomes at the teacher-level.

Table 3.16: Direct Effects of Within-School Model (BAM Math Supplemental Assessment VAMs) Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=475

	Unstandardized	Standardized
VAM (Math using BAM Math Supplemental Assessments)		
Male	0.002 (0.03)	0.003 (0.05)
White	-0.012 (0.03)	-0.024 (0.07)
Years Exp. In District	0.014 (0.02)	0.046 (0.06)
Master	-0.021 (0.04)	-0.042 (0.09)
Instructional Quality (Math)	0.100 (0.05)	0.156 (0.08)*
Ped. Content Knowledge (Math)	0.003 (0.00)	0.160 (0.06)*
Instructional Quality (Math)		
Male	-0.105 (0.06)	-0.111 (0.06)
White	-0.030 (0.05)	-0.040 (0.07)
Years Exp. In District	-0.011 (0.03)	-0.024 (0.05)
Master	-0.017 (0.07)	-0.023 (0.09)
Pedagogical Content Knowledge (Math)		
Male	-0.227 (1.58)	-0.007 (0.05)
White	8.195 (1.85)	0.291 (0.06)***
Years Exp. In District	-2.213 (1.04)	-0.124 (0.06)*
Master	0.402 (2.07)	0.014 (0.07)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.4.3 Contextual Effects

Similar to the results in the previous math model, there were no significant contextual effects on student learning or in relation to any mediators. Table 3.17, below, provides that standardized and unstandardized estimates of direct relationships between predictors, mediators, and outcomes at the school-level.

Table 3.17: Direct Effects of Between-Level Model (BAM Math Supplemental Assessment VAMs)
Unstandardized and Standardized Estimates (Standard Errors in Parentheses); N=475

	Unstandardized	Standardized
VAM (Math using BAM Math Supplemental Assessments)		
% Male	-0.048 (0.1)	-0.122 (0.25)
% White	0.012 (0.09)	0.036 (0.28)
Avg. Years Exp. In District	0.006 (0.01)	0.195 (0.27)
% Master	0.04 (0.17)	0.127 (0.52)
Avg. Instructional Quality (ELA)	0.308 (0.39)	0.307 (0.40)
Avg. Ped. Content Knowledge (ELA)	-0.016 (0.01)	-0.497 (0.37)
Instructional Quality (Math)		
% Male	-0.040 (0.10)	-0.102 (0.26)
% White	-0.008 (0.09)	-0.026 (0.27)
Avg. Years Exp. In District	0.006 (0.010)	0.178 (0.18)
% Master	-0.134 (0.14)	-0.419 (0.41)
Pedagogical Content Knowledge (Math)		
%Male	2.451 (3.41)	0.197 (0.27)
%White	-2.109 (2.89)	-0.206 (0.28)
Avg. Years Exp. In District	0.104 (0.24)	0.102 (0.24)
% Master	-2.403 (3.71)	-0.236 (0.36)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.2.4.4 Mediating Effects

Unlike in the previous math model, a mediating, or indirect, effect was identified on the teacher level with the average White urban teacher, as mediated by his/her math pedagogical content knowledge, having significantly higher teacher effects than the average minority urban teacher with less pedagogical content knowledge (β= 0.023, p<0.05), controlling for grade-level and district differences.

According to MacKinnon, Fairchild, and Fritz (2007), this effect was identified as an inconsistent mediating effect as the direct effect of White on teachers' value-added to student learning was negative – White teachers are less effective at improving student learning in math, controlling for grade-level and district differences. However, the effect of race on teachers' pedagogical content knowledge was positive and the effect of teachers' pedagogical content knowledge was positive, making the indirect or mediating effect positive. The unstandardized total effect (e.g., total effect = direct effect (-0.012) + indirect effect (0.023)) was 0.011. Therefore, pedagogical content knowledge acts as a suppressor variable, establishing causality between teachers' race and its value-added to student learning – increasing the predictive validity of race on teachers' value-added (Thompson & Levine, 1997). Mediating effects are presented in Table 3.18, below.

Table 3.18: Mediating (Indirect) Effects (BAM Math Supplemental Assessment VAMs) Unstandardized Estimates (Standard Errors in Parentheses); N=475

Within (Teacher) Level	
Male→ Instruct. Quality (Math)→ VAM (Math)	-0.003 (0.01)
White→ Instruct. Quality (Math)→ VAM (Math)	-0.010 (0.01)
Years Exp. In District → Instruct. Quality (Math)→ VAM (Math)	-0.001 (0.00)
Master→ Instruct. Quality (Math)→ VAM (Math)	-0.002 (0.01)
Male→ Ped. Content Know. (Math)→ VAM (Math)	-0.001 (0.00)
White→ Ped. Content Know. (Math)→ VAM (Math)	0.023 (0.01)*
Years Exp. In District→ Ped. Content Know. (Math)→ VAM (Math)	-0.006 (0.00)
Master→ Ped. Content Know. (ELA)→ VAM (Math)	0.001 (0.01)
Between (School) Level	
% Male→ Instruct. Quality (Math)→ VAM (Math)	-0.012 (0.03)
% White→ Instruct. Quality (Math)→ VAM (Math)	-0.003 (0.03)
Avg. Yrs. Exp. In District → Inst. Quality (Math)→ VAM (Math)	0.002 (0.00)
% Master→ Instruct. Quality (Math)→ VAM (Math)	-0.041 (0.07)
% Male→ Ped. Content Know. (Math)→ VAM (Math)	-0.038 (0.06)
% White→ Ped. Content Know. (Math)→ VAM (Math)	0.033 (0.05)
Avg. Yrs. Exp. In Dist.→ Ped. Cont. Know. (Math)→ VAM (Math)	-0.002 (0.00)
% Master→ Ped. Content Know. (ELA)→ VAM (Math)	0.037 (0.07)

Note: *p<.05, ** p<.01, ***p<.001. This model controlled for grade-level and district differences.

3.3 Predictive Analyses Results

Significant predictors identified in MSEM model results were used to predict the impact on teacher effectiveness of replacing teachers with these attributes among the current urban teacher workforce. An alternative approach to this kind of predictive analysis would be to examine the impact of improving effectiveness of the current teacher workforce based on these significant predictors (e.g., improving math pedagogical content knowledge through professional development), instead of replacing teachers. However, this approach makes strong assumptions about the effects of professional development and is out of the scope of this study. It should also be noted that this predictive analysis was meant to be purely hypothetical, with the understanding that VAMs may not be valid or reliable enough indicators of teachers' value-added to student learning to be the only measure used in hiring and firing decisions. Instead, VAM estimates should be used in addition to other measures of teachers' instructional practices and qualifications in order to make sound decisions on the quality of teaching in urban schools (Darling-Hammond, et al., 2003; McCaffrey, et al., 2003; McCaffrey, et al., 2004). Furthermore, I acknowledge that replacing percentages of teachers in the workforce with new, more effective teachers is not likely to improve student improvement dramatically by itself, but rather it is likely to require other shifts within these schools systems, including changes in school and administrative leadership and culture, among many others (Goldhaber, 2015). Instead, this analysis gives us a simplistic view of the impact on student learning if we were to hypothetically make these changes in the urban teacher workforce.

In predictive analyses, I identified the most qualified 5%, 10%, and 20% of urban teachers by significant attributes identified in MSEM models (e.g., those with the highest pedagogical content knowledge assessment scores). I selected the VAM scores of these teachers and used them to replace the VAM scores of the most ineffective 5%, 10%, and 20% of teachers (i.e., those with the lowest VAM scores) across the urban teacher workforce. This is analogous to recruiting additional teachers whose VAM scores are similar to those of existing teachers with the most desirable qualifications and characteristics. In doing so, I am able to discuss policy and practice implications of increasing the number effective teachers based on significant characteristics resulting from each model.

In order to make the resulting teacher effects more meaningful, I converted these average teacher effects into months of student learning. A standardized effect size, as in teacher effects, of 0.27 in ELA and 0.34 in math is roughly equivalent to one additional year of learning across grades 4-9 – grades used in models in this study (Hill, et al., 2008). For example, if increasing the levels of ELA instructional quality among the most ineffective 5% of the urban ELA teacher workforce results in a gain of 0.024 standard deviations, it converts to an annual gain of 1.1 months of student learning in ELA. Below I discuss the results of predictive analyses based on each model. It should be noted that these annual gains are over the course of 12 months.

3.3.1 ELA State Assessment Model

The first ELA model, which used standardized state assessment scores, identified race and instructional quality as significant predictors of teacher effects. Results also identified contextual effects, where teachers were more effective if they taught at schools with high levels of instructional quality or at schools with more teachers with advanced degrees. Below, I discuss the impact that increasing teachers with these attributes across the urban teacher workforce has on the average teacher effect and on student learning. It should be noted that predictions made using significant predictors of contextual effects are those at the teacher level and are not true predictions of contextual effects. These predictions are likely to underestimate the true contextual effects but give an idea of how average urban teacher workforce effects would increase by replacing the most ineffective urban teachers with the highest performing urban teachers based on significant attributes.

Minority urban ELA teachers had significantly higher VAMs than White teachers in this model. While I understand there are ethical and legal concerns in making hiring decisions based solely on a teacher's race, it is important to discuss the impact of increasing minority urban ELA teachers across the workforce in relation to previous literature where teachers' race matters, as is done in the next chapter. Starting with the replacement of the most ineffective 5% of ELA teachers, I predicted that the average teacher effect improved by 0.044 standard deviations. This result converted to an annual gain of 1.9 months of student learning in ELA. The replacement of 10% of the lowest performing urban ELA teachers resulted in an average teacher effect of 0.069 standard deviations across the urban teacher workforce, which converted to 3.1 months of additional student learning annually. An average teacher effect of 0.104 standard deviations, or an annual gain of 4.6 months of student learning, resulted when replacing 20% of the lowest performing urban teachers in the workforce with the highest performing minority urban teachers.

Instructional quality was also found to be a significant predictor of teacher effectiveness among urban ELA teachers in this model, having both direct and contextual effects. I found that replacing VAM scores among 5% of the most ineffective urban ELA teachers with VAM scores of urban ELA teachers with the highest 5% of scores of instructional quality resulted in an average teacher effect gain of 0.025 standard deviations. This increase translates into 1.1 additional months of student learning in ELA annually. Replacing 10% of ineffective urban teachers resulted in an average teacher effect of 0.041 standard deviations or an annual gain of 1.8 months of student learning in ELA. When increasing levels of instructional quality across the urban teacher workforce, the replacement of 20% of the most ineffective urban ELA teachers resulted in an average teacher effect of 0.066 standard deviations, an annual gain of 2.9 months of student learning in ELA.

Schools having high levels of teachers with master's degrees or above was predictive of teachers' value-added to student learning. Knowing this, I predicted the impact of increasing the number of teachers in the workforce, not in the schools, having advanced degrees. This reflects a prediction based on the direct effect of teachers with advanced degrees versus teachers without advanced degrees, and not a prediction of the true contextual effect as it was statistically too difficult to estimate. The VAM scores of the most ineffective 5% of ELA teachers in the urban teacher workforce were replaced with the VAM scores of the most effective teachers that indicated that they hold an advanced degree, based on their VAM scores. In doing so, I predicted that the average urban ELA teacher effect increases by 0.038 standard deviations, which converted to an annual gain of 1.7 months of student learning in ELA. The replacement of 10% of the lowest performing urban ELA teachers, resulted in an average teacher effect of 0.060 standard deviations across the urban teacher workforce and converted to 2.7 months of additional student learning in ELA annually. An average teacher effect of 0.085 standard deviations resulted when replacing 20% of the lowest performing urban teachers in the workforce with scores of the highest performing minority urban teachers, which converted to an annual gain of 3.8 months of student learning in ELA.

3.3.2 SAT-9 Supplemental Reading Assessment Model

In the ELA model using SAT-9 supplemental reading assessments, instructional quality was identified as the only significant predictor of urban teacher effectiveness. Knowing this, I predicted that replacing VAM scores among 5% of the most ineffective urban ELA teachers with VAM scores of urban ELA teachers with the highest 5% of scores of instructional quality increased the average teacher effect by 0.038 standard deviations, which converted to an additional 1.7 months of student learning in ELA annually. Replacing 10% of the most ineffective urban teachers resulted in an average teacher effect of 0.061 standard deviations. This converted into an annual gain of 2.8 months of student learning in ELA. Increases in the level of instructional quality across the urban teacher workforce when replacing 20% of the most ineffective urban ELA teachers resulted in an average teacher effect of 0.099 standard deviations – an annual gain of 4.5 months of student learning in ELA.

3.3.3 Math State Assessments Model

The model using VAMs based on math state standardized assessments identified instructional quality as the only predictor of urban teacher effectiveness. Starting with the replacement of the most ineffective 5% of math teachers with the math teachers with the highest levels of instructional quality, I found that the average teacher effect increased by 0.031 standard deviations. This average teacher effect converted to an annual gain of 1.1 months of student learning in ELA. The replacement of 10% of the most ineffective urban math teachers resulted in an average teacher effect of 0.034 standard deviations across the urban teacher workforce, which converted to an annual gain of 1.9 months of student learning. An average teacher effect of 0.061 standard deviations, or an annual gain of 2.8 months of student learning, resulted when replacing 20% of the most ineffective urban teachers in the workforce with teachers having the highest levels of instructional quality.

3.3.4 BAM Supplemental Math Assessment Model

The fourth and final model used math VAMs based on BAM supplemental math assessments and found that instructional quality and pedagogical content knowledge were significant predictors of teachers' effectiveness. A mediating effect was also identified where being a White teacher, as mediated by their math pedagogical content knowledge, was a significant predictor of urban math teacher effectiveness. The causal factor in this relationship was math teachers' pedagogical content knowledge; therefore I focused my predictions on increasing the number of urban math teachers with high levels of pedagogical content knowledge across the urban teacher workforce and not on race in presenting predictions based on this effect. Overall, I predicted the impact of increasing the number of math teachers with instructional quality and pedagogical content knowledge across the urban teacher workforce using the results from this model.

Starting with instructional quality, I predicted that replacing VAM scores among 5% of the most ineffective urban math teachers with VAM scores of urban math teachers having the highest 5% of instructional quality scores increased the average teacher effect by 0.031 standard deviations. This converted to an additional 1.1 months of student learning in math, annually. The replacement of 10% of the most ineffective urban math teachers resulted in an average teacher effect of 0.047 standard deviations, which converted to an annual gain of 1.9 months of student learning in math. Increases in the level of instructional quality across the urban teacher workforce when replacing 20% of the most ineffective urban math teachers, resulted in an average teacher effect of 0.073 standard deviations, or an annual gain of 2.8 months of student learning in math.

In replacing 5% of the most ineffective urban math teachers with urban teachers having the highest levels of math pedagogical content knowledge, I found that the average teacher effect increased by 0.035 standard deviations, which converted to an annual gain of 1.2 months of student learning. Replacing 10% of the most ineffective urban math teachers across the urban teacher workforce resulted in an average teacher effect of 0.050 standard deviations. This effect converted to an additional 2.0 months of student learning in math annually. When increasing levels of math pedagogical content knowledge across the urban teacher workforce, the replacement of 20% of the most ineffective urban math teachers resulted in an average teacher effect of 0.070 standard deviations, an annual gain of 2.7 months of student learning in math.

3.4 Results Summary

The main inquiry for this study was in understanding significant predictors of urban teacher effectiveness. MSEM results indicated that urban teachers' race and instructional quality were predictors of urban teacher effectiveness among ELA teachers. In math, teachers' instructional quality and pedagogical content knowledge were significant predictors of urban teachers' value-added to student learning. A mediating effect was identified in math, with urban White math teachers as mediated by their pedagogical content knowledge predictive of their value-added in improving math learning. Contextual effects were identified in ELA, where schools with higher rates of instructional quality or higher rates of teachers with advanced degrees were associated with teachers having greater impacts on student learning. Table 3.19 below summarizes significant teacher effects resulting from all four models.

Table 3.19: Summary of Significant Teacher Effects Resulting from MSEM Models

Outcomes	Predictors	Standardized Estimates of Teacher Effects
ELA Model Using State Assessment VAMs		
	Direct (Teacher-Level Effects)	
	White	-0.150 (0.06)*
	Instructional Quality	0.180 (0.07)**
	Contextual (School-Level) Effects	
	% Master	0.905 (0.27)**
	Avg. Instructional Quality	0.540 (0.25)*
ELA Model Using SAT-9 Reading Assessment VAMs		
	Direct (Teacher-Level Effects)	
	Instructional Quality	0.218 (0.06)***
Math Model Using State Assessment VAMs		
	Direct (Teacher-Level Effects)	
	Instructional Quality	0.186 (0.07)*
Math Model Using BAM Math Assessment VAMs		
	Direct (Teacher-Level Effects)	
	Instructional Quality	0.156 (0.08)*
	Pedagogical Content Knowledge	0.160 (0.06)*
	Mediating (Total) Effect	
	White→ Pedagogical Content Knowledge→ VAM	0.011 (0.01)*

Note: *p<.05, ** p<.01, ***p<.001. Models controlled for grade-level and district differences. The coefficient of the mediating effect given by Mplus is unstandardized.

Results from predictive analyses showed that making increases in the levels of significant predictors in the urban teacher workforce was likely to have moderate impacts on teacher effects on student learning. I first predicted the impact on the average teacher effect by replacing the most ineffective 5%, 10%, and 20% of teachers in the workforce with the highest performing teachers based on these significant attributes. I then converted this average teacher effect into months of student learning in order to make the predicted teacher effects more meaningful.

Table 3.20 below summarizes significant predictors of teacher effectiveness and their impact on student learning. Results demonstrate that the largest impact among ELA teachers was found when increasing the number of ELA teachers in schools who are minority and with advanced degrees. The largest impact in math on student learning is made when increasing instructional quality.

Table 3.20: Potential Impact on Student Learning in Months for Significant Predictors of Teacher Effectiveness by Assessments used in VAM Estimates (Annual Gains in Student Learning)

	Potential Impact on Student Learning (Months)		
	Replacing 5%	Replacing 10%	Replacing 20%
ELA State Assessment Model			
Minority Teachers	1.9	3.1	4.6
Instructional Quality	1.1	1.8	2.9
Master's or Advanced Degrees	1.7	2.7	3.8
SAT-9 Supplemental Reading Assessment Model			
Instructional Quality	1.7	2.8	4.5
Math State Standardized Assessment Model			
Instructional Quality	1.1	1.9	2.8
BAM Supplemental Math Assessment Model			
Instructional Quality	1.1	1.9	2.8
Pedagogical Content Knowledge	1.2	2.0	2.7

Chapter 4

DISCUSSION

Urban schools continue to face ongoing academic and contextual challenges (Ingersoll, 2003; Darling-Hammond, 2003; Hanushek, Kain & Rivkin, 2004; Darling-Hammond, 2000). Improving the quality of teachers is one means by which we can improve student learning in urban schools, as research has shown that teachers are the most important school-based factor in improving student learning (Darling-Hammond, 2003; Hanushek, et al., 2005; Jacob, A., 2012; Nye, et al., 2004; Rivkin, et al., 2005; Rice, 2003; Hanushek, 1992). This study of urban teacher quality examines attributes of urban teachers most predictive of their value-added to student learning. The framework for this study is based on a framework of teacher quality that organizes teacher attributes into inputs, intermediate processes, and empirical outcomes of teacher effects, building off of Goe's (2007) framework on teacher quality. The study also considers mediating effects that may help explain relationships between teacher attributes and their effectiveness as well as the multilevel nature of educational data. This study builds on the foundation of understanding of urban schools and communities and the instructional challenges teachers face in these contexts. Therefore the study uses a sample of teachers from schools most reflective of those in urban contexts.

The MSEM analytical approach used in this study allowed for an empirical exploration of the study's framework on teacher quality. Analyses used data from the MET longitudinal database on urban teachers serving in schools with at least 60%

minority and 60% low-income students. In this model, teachers' inputs included gender, race, years of teaching experience in their districts, and level of education. Intermediate, or mediating, processes included measures of teachers' overall instructional quality and their pedagogical content knowledge. VAM measures were used to estimate teachers' effectiveness in improving student learning. In total, four models were estimated by content area (i.e., ELA, math) and by assessment type used in VAM estimations (i.e., state standardized or supplemental assessments). I then estimate how student learning may be impacted if the number of teachers in the urban teacher workforce with significant attributes identified in these analyses increased.

Results from this study show that by focusing on specific attributes of the urban teacher workforce, we can likely make meaningful improvements in teachers' value-added to student learning, albeit, I also recognize that the study has several limitations. Below, I interpret the findings and discuss limitations from this study.

4.1 Interpretation of Findings

I interpret the results from this study following the study's conceptual framework for understanding teacher quality in urban contexts. Findings indicate that not only are there attributes of teacher quality that may have direct relationships with urban teachers' effectiveness, there is also a mediating relationship that should be considered along with contextual effects that influence student learning.

4.1.1 A Framework for Understanding Urban Teacher Quality

The conceptual framework considers teacher inputs and processes in relation to empirical measures of their influence on student learning. Using this framework, I was able to understand and explain which attributes characterize quality urban teachers.

4.1.1.1 Teacher Inputs

Teacher inputs, characteristics and qualifications, examined in this study include teachers' gender, race, years of experience teaching in their district, and their educational level. In previous research investigating teacher quality, these teacher inputs were found to have weak, inconsistent effects on student achievement (Clotfelter, et al., 2010; Dee, 2004; Ehrenberg, et al., 1994; Goddard, et al., 2000; Hanushek, et al., 2005). These findings are similar to the results of this study.

This study revealed that gender is not a significant attribute that contributes to variation in urban teacher quality, regardless of VAM assessment type or subject content area. Even more, while significant results were not found, the relationship between gender and teacher effects on student learning were mixed depending on the assessment used and the content area, demonstrating the weakness of using gender as an indicator of urban teacher effectiveness. While I examined the direct relationship of teachers' race with their effectiveness in the classroom, other researchers have examined teachers' gender in combination with their race, concluding with some slightly stronger relationships (Ehrenber, et al., 1995). Race and gender combinations were not modeled in this study and should be considered in future research on urban teacher quality.

Results were also mixed when considering race as a predictor of teachers' effectiveness in urban schools – differing by assessment type and content area. Nevertheless, one significant direct relationship between teachers' race and their effectiveness was identified. Using ELA state assessments, minority teachers were found to be significantly more effective at improving student learning of ELA than White urban teachers. This direct relationship indicates that race may matter in improving students' learning of ELA in urban schools serving the most academically at-risk, and can result in about 4.6 months of student learning annually.

Previous research has shown that Black teachers have positive effects on student learning when teaching students of their own race (Dee, 2004; Clotfelter, 2010; Hanushek, et al., 2005). Given the direct significant relationship of minority urban ELA teachers' impact on student learning observed in this study, these findings suggest that the race of the teacher matters when instructing students of their same race. Teachers included in this study taught in schools with high percentages of minority students (i.e., at least 60%). Therefore, it is likely that the relationship between teachers and students is influenced by shared race and cultural experiences. Based on this assumption, and the findings from the previous studies referenced above that identified this potential relationship, it is important that future studies of teacher quality in urban contexts use matched student-teacher race comparisons to examine the strength and nature of teachers' race on student learning when racial identities are shared.

Teachers' years of experience teaching in their district was also examined as a predictor of urban teacher quality and did not have a significant direct relationships with teacher impacts on student learning. It is possible the relationship between teachers' experience and teachers' value added may differ if a measure of total years of teaching experience was used rather than only using a measure of teaching experience only within the district where the teacher was currently employed.

Furthermore, in previous research the first few years of teaching experience overall have been associated with increased gains in student achievement (Boyd, et al., 2006; Clotfelter, et al., 2007, 2010; Goe, 2002; Hanushek, 2003; Hanushek, et al., 2005; Harris & Sass, 2007; Kane, et al., 2006; Nye et al., 2004; Rice 2003; Rivkin, et al., 2005; Rockoff, 2004) with these gains diminishing after the first few years (Boyd, et al., 2006; Clotfelter, et al., 2010; Hanushek, et al., 2005; Palardy & Rumberger, 2008; Rivkin, et al., 2005). Therefore, it would have been more meaningful to examine teachers' years of experience overall, as it is more reflective of their overall experience in the classroom. Data on this measure were included in the MET database but were mostly missing across the teacher sample and not included in this analysis. Nevertheless, as previous research has demonstrated a relationship between teachers' experience in the classroom and their impact on student learning, I suggest that urban teachers' overall teaching experience should be further examined in order to understand whether or not teachers' experience really matters in improving student learning in urban schools.

Teachers' level of education also did not have a significant direct relationship with their value-added to student learning when considering teaching in urban contexts across all models. In addition, the direction of the relationship differed across models. This finding is similar to findings from other teacher quality studies that included teachers' level of education in the model and found mixed results (Cavalluzzo, 2004; Clotfelter, et al., 2007, 2010; Hanushek, et al., 2005; Harris & Sass, 2007; Nye, et al., 2004; Palardy & Rumberger, 2008; Rivkin, et al., 2005). While the effect of having an advanced degree was not evident at the teacher level, a contextual effect was identified with this attribute, which I discuss later. Overall, having an advanced degree has little direct association with teachers' value-added to student learning.

The present study's results indicate that teacher inputs (i.e., teacher characteristics or qualifications) play a small and mostly insignificant part in explaining urban teacher effectiveness. Previous research on teacher quality suggests that some of these inputs could be examined together to strengthen their causal explanation of teacher quality. Nevertheless, I conclude that teacher inputs generally have weak, direct relationships with teachers' impact on student learning in urban contexts. This is not to overlook the finding that minority urban ELA teachers are more effective in improving student learning than White urban teachers, which may be attributed to the shared backgrounds of teachers and students of the same race or ethnicity. Future research should continue to examine urban teacher effectiveness when student and teachers' race are shared.

4.1.1.2 Teacher Processes

The conceptual framework for this study also allowed for the examination of urban teachers' processes in the classroom in relation to their impact on student learning. The overall measure of instructional quality used in this study resulted from a confirmatory second-order factor using indicators from both the CLASS and FFT teacher evaluation frameworks. This overall measure was the most consistent predictor of urban teacher effectiveness across all models having a significant, positive relationship with teacher effects in all four MSEM models. This finding mirrors results from other studies of teacher quality, which also find these positive relationships using FFT (Borman & Kimball, 2005; Gallagher, 2004; Heneman, et al., 2006; Holtzapple, 2003; Kane, et al., 2011; Kimball, et al., 2004; Milanowski, 2004) and CLASS (Allen, et al., 2013) frameworks, separately. Thus, we must consider that urban teachers' instruction in the classroom may contribute to their value-added to student learning among our most academically at-risk urban schools. Increasing levels of instructional quality across the urban teacher workforce is likely to have an impact of up to 2.8 months of student learning in math and up to 4.5 months of student learning in ELA, annually.

4.1.1.3 Teacher Inputs/Processes

Under the study's conceptual framework, pedagogical content knowledge is understood to be both an input and a process. Due to previous research that finds moderate relationships between this attribute and student learning, and my understanding of the attribute as both an input and process, I examined pedagogical content knowledge independently under this framework.

In the model using BAM supplemental math assessments, pedagogical content knowledge was found to be a significant predictor of urban teachers' value-added to student learning. Making improvements in urban teachers' math pedagogical content knowledge can contribute to an additional 2.7 months of learning, annually. In previous studies of teacher quality, some of the larger teacher effects have been associated with high levels pedagogical content knowledge in math (Harris & Sass, 2006, 2007; Hill, Rowan & Ball, 2005). Therefore it is not surprising to find the significant relationship between urban math teachers' pedagogical content knowledge associated and teacher effects. However, this attribute was not significant in either ELA model. This is not surprising either, as no previous research has concluded that teachers' ELA pedagogical content knowledge is predictive of their effectiveness.

It is interesting that PCK was only significant in the BAM assessment model and not in the math state assessment model. State assessments are aligned to classroom curricula and in most cases connected to high-stakes decisions within schools. As a result, there is an incentive for teachers and students to game the system by teaching to the test or pressuring students in preparation for taking the test, among other strategies (Darling-Hammond, 2007; Darling-Hammond & Rustique-Forrester, 2005). These practices are more likely to occur in grade levels where high-stakes decisions are related to test results (Stecher & Barron, 1999) and where there are more underperforming students at risk of not passing the test (Herman & Golan, 1993). As a result, in some cases state assessment scores are more likely reflect students' mastery of taking state assessments and teachers' ability to teach to the test (Centolanza, 2004). As a result, math teachers' pedagogical content knowledge may not be as evident as a predictor of teacher effects using state assessments. On the other hand, as the BAM supplemental assessment was not aligned to state curricula, it may be that it is likely to be more of a measure of students' actual knowledge of math concepts. Therefore, math teachers' pedagogical content knowledge is more likely to be predictive of their value added under the BAM model, measuring students' true knowledge of math, than in the state assessment model, where the focus of instruction may be on passing these high-stakes assessments.

4.1.2 A Mediation Hypothesis

This study's conceptual framework uniquely considered mediators. One significant mediating relationship was identified – White teachers were found to be more effective at improving student learning than minority teachers when considering the mediating effect of their pedagogical content knowledge using math supplemental

BAM assessments. This indirect effect was shown to act as a suppressor, indicating that urban White math teachers' pedagogical content knowledge was causing the relationship between teachers' race and their value-added to student learning. Since teachers' pedagogical content knowledge is driving the mediated relationship, this finding demonstrates that pedagogical content knowledge may be a powerful mechanism for improving the quality of math teachers in urban schools. This is not to dismiss the conversation around race in this mediating relationship. White teachers may be receiving stronger instruction in how to teach math than minority teachers, which may then speak to the teacher preparation programs they attend or from which they are recruited. White teachers may likely be attending educational institutions or getting trained in teacher preparation programs with strong instruction and curriculum designed to provide teachers with knowledge in math instruction. However, this is not ascertainable in the present study but warrants further inquiry. Findings do suggest however that future studies must consider teacher attributes as potential mediators helping to explain teacher quality. Other promising mediating relationships may surface from these studies, particularly in urban contexts.

4.1.3 The Contextual Influence on Student Learning

The framework for this study also considered the naturally nested nature of educational data by examining contextual effects at the school level. Two contextual effects were identified in this study.

Positive effects on urban teachers' value-added were found among teachers working in schools having more teachers with advanced degrees and higher levels of instructional quality when modeled using VAM scores based on ELA state assessments as outcomes. This result is interpreted quite differently than a teacher level effect, as it signifies that teachers' effectiveness in the classroom is influenced by having larger percentages of teachers' with master's degrees or higher levels of instructional quality. Increasing the number of teachers in the urban teacher workforce with high levels of instructional quality can lead to about 2.9 months of student learning in ELA, annually. Even more, increasing the number of highly effective urban ELA teachers with advanced degrees can lead to about 3.8 months of student learning, annually. The true impact of the contextual effects on student learning are likely to be greater than these estimates, yet these estimates give us an idea of the impact of increasing the numbers of these types of teachers in the urban teacher workforce.

These contextual effects may be evidence of of a social or interaction component of teacher networks in schools based on either their level of education or their instructional quality. Schools may benefit from building a community of practice based on shared characteristics, skills, or knowledge, which researchers suggest lead to quality instruction in urban schools (Oakes, et al., 2002). This often occurs in collaborative efforts among educators in schools such as professional development efforts, mentoring or coaching, professional learning communities, co-teaching, or other models of collaboration. Thus, urban teacher learning, as a social process, may be evident in these findings of contextual effects, which likely influences teachers' impact on student learning (Oakes, et al., 2002).

Alternatively, the contextual effect related of having higher percentages of teachers with advanced degrees may reflect other aspects of school conditions. For example, schools with larger percentages of teachers with advanced degrees may offer financial incentives or reimbursement for advanced degrees. Similarly, those schools may engage in hiring practices based on levels of education or have other characteristics that enable them to attractgreater numbers of teachers with advanced degrees. While these explanations are beyond the scope of this project, future research must tease out the contextual effect based on teachers' levels of education to understand what is truly driving its relationship between with teachers' value-added to student learning in order to make solid policy or practical implications based on this finding.

While a growing number of studies have considered contextual effects in their examination of teacher quality (Borman & Kimball, 2005; Cohen & Hill, 1998; Croninger, et al., 2007; Eckert, 2012; Palardy & Rumberger, 2008; Wenglinsky, 2000, 2002), we must continue to grow our understanding of significant contextual influences on student learning in urban schools – those extending above teacher level effects. What might we learn about teacher collaborations and communities and how might they impact urban teacher effectiveness? While research on social dimensions of teaching and learning communities is extensive, little research connects that to effects on student learning (see Vescio, et al., 2008 for a discussion).

Overall, results from this study suggest that a new framework for studying teacher quality is warranted, particularly when examining urban teachers. Significant direct, contextual, and mediating effects were identified which indicates the need for a more detailed understanding of urban teacher quality. These results also indicate the need for further examination of significant effects, which resulted from this study. I discuss directions for future research in more detail in the Chapter 5.

4.2 Study Limitations

The results presented give us promising leads in efforts to understand urban teacher effectiveness, although there are several limitations to this study that should be discussed. Study limitations include issues with external validity, the validity of VAM estimates, unobserved or uncollected data as well as measurement issues within the MET database. Addressing these issues requires more robust databases with more valid and reliable measures on urban teachers and teaching from which to conduct future studies on urban teacher quality.

4.2.1 External Validity

External validity is the degree to which results can be generalized across educational settings. Threats to external validity in this study can result from not having an ample or representative sample size. The randomization process employed in the MET study lessens threats to external validity by assigning teachers to classrooms, but does not eliminate them. Students were not randomly assigned to classrooms, therefore some bias may be introduced due to students sorting or tracking – where students are placed into certain classrooms based on their characteristics or schools track students into particular courses based on course difficulty. However, VAM estimates control for prior test scores, along with other student demographics, which allows estimates of value-added to not be influenced by students' prior knowledge and instead allows for VAM estimates to be mostly influenced by teacher and teaching attributes.

The use of opportunity sampling (volunteers) and incentives among MET participants is a threat to external validity as certain types of teachers with similar teacher attributes may have opted into the study, making the sample less representative of a broader sample of teachers. The teachers included in this study volunteered to participate, which may be indicative of a particular characteristic of these types of teachers unknown to us. Nevertheless, we must consider that the sample of teachers used in this study is not truly representative of teachers teaching at-risk urban youth, thus these results are not generalizable.

Furthermore, the MET study was conducted only among six districts making it even further more difficult to generalize these findings. While I tried to restrict the sample to schools in districts with similar characteristics, the sample used is still restricted to the six districts studied by MET. Just as there is variation in the schools across districts, there is variation in the contextual influences of urban schools across America; thus not all urban districts or schools look the same. In order to comprehensively study urban schooling, researchers must amass a much larger sample of urban districts across the country in order to draw sound conclusions.

4.2.2 Validity of VAMs

VAM has become one of the more common methods used to estimate teacher effects on student learning, although there continues to be some concern over the validity of different VAM approaches. MET researchers used a covariate-adjustment approach to VAM in estimating teacher effects. Researchers suggest that VAMs control for both prior test scores and mean peer characteristics perform best and can be used to measure teacher effects, particularly when examining short-term effects (Kane & Staiger, 2008). While MET researchers used this approach, some question whether VAMs accurately measure teacher effects on student learning. This question appears to be of most concern when VAMs are tied to high stakes teacher evaluation systems and less when VAMs are used in research on teacher quality, although the question remains.

Researchers have expressed concerns that VAMs only partially control for differences in students assigned to teachers (Rothstein, 2011). While using the covariate-adjustment approach to VAM is easier to specify and interpret (i.e., all students starting at the same level of achievement), fitting models separately for each year of data ignores information on student performance in other years that can account for individual student factors and reduces sampling error and bias in the current year teacher effect estimate (Darling-Hammond, et al., 2012; McCaffrey, et al., 2003; McCaffrey, et al., 2004). The random assignment of teachers to classrooms, as conducted in the MET study, intends to reduce biases related to the assignment of teachers to classrooms with higher or lower achieving students. Although, in this study the randomization process is somewhat questionable due to the limited number of teachers meeting the randomization criterion in schools and the fact that students were not randomly placed in classes. Nevertheless, there was an attempt to improve causal inferences through this process.

Furthermore, some argue that VAMs are unable to control for unobserved student, teacher, or school characteristics in their estimations (Harris & Sass, 2011). Relying on observed, measurable characteristics alone contribute to inconsistent or biased estimates of teachers value-added (Harris & Sass, 2011). This argument, although valid, is very difficult to overcome in VAM estimations, particularly within the MET study.

Measurement error is also of concern under this approach of estimating teacher effects, as there is an assumption that prior year test scores are free of measurement error. Ignoring such measurement error creates biased model coefficients and can potentially result in systematic errors in teacher effects (McCaffrey, et al., 2003). This is likely the case under the MET model as the distribution of prior achievement varies across classes, which probably has created some bias in teacher effects and should be considered in interpreting results from this study.

The use of VAM also presents issues with missing data or incomplete records (McCaffrey, et al., 2003). Under the covariate-adjustment model, students with missing prior or current year student scores are excluded. The discarding of partial cases can create biased coefficient estimates. In covariate-adjustment models, such as the MET model, valid estimates of teacher effects are produced when the missing data are missing at random (MAR) or missing completely at random (MCAR) but not when the data are missing not at random (MNAR).

One notable argument is that VAMs tend to be unstable from year to year (Ballou, 2005; Darling-Hammond, et al., 2012; Schochet and Chiang, 2010), with correlations between 0.2 and 0.5 (Goldhaber and Hansen 2013; McCaffrey et al. 2009). Polikoff (2015) argues that these instabilities can lead to other potential problems with the use of teacher effects wrongly classifying teachers as effective or ineffective as well as their lack of ability to inform policy and practice.

The results of this study use VAM estimates of teacher effects in both ELA and math. While there is still debate over the strength of these estimates in truly measuring teachers' value-added to student learning, I consider these measures valuable in exploring teacher quality in urban contexts. Although individual VAM estimates may be unstable, exploring systematic relationships across a large sample involves many different data points. Nevertheless, I would consider the limitations presented here in using results from this study.

4.2.3 Assessments of Student Learning

This study also introduces the question of which assessments are most appropriate for use in creating VAM estimates, or in measuring student learning in general. While some of the results were similar across models that differed by assessments used, some were different. For instance, instructional quality was the only significant predictor of urban teacher effectiveness in the math model using math state assessment VAM values. Instructional quality, pedagogical content knowledge, and race with pedagogical content knowledge as a mediator were all significant predictors of urban teacher effectiveness under the math BAM supplemental assessment model. These differences present the question of which are the most appropriate measures of student achievement to be used in value-added modeling. It also suggests that we question which assessments truly measure students' knowledge of ELA and math.

If we are to better understand what contributes to or influences student learning, we must identify the best measures. State assessments, on one hand, may mirror the content that teachers teach in the classrooms. In these cases student assessment scores do not necessarily reflect students' knowledge of content but their knowledge of how to take state standardized assessments. On the other hand, supplemental assessments may not be directly aligned to current content curriculum making the results of these assessments invalid and unreliable in measuring students' knowledge of the content taught. Therefore, in studies on teacher quality, researchers must identify the most valid and reliable assessments from which we can truly measure what students are expected to know from their teachers' instruction. These assessments can then be used in future VAM models estimating teacher effectiveness.

4.2.4 Measures of Teacher Practices and Knowledge

I also find from this inquiry, analysis, and results that the field must do a better job of creating systems to measure instructional quality and knowledge for teaching specific content in classrooms. While some seemingly strong measures have been created and are widely used in evaluating teacher practices (e.g., CLASS, Framework for Teaching, CKT), there is still great variation in the data resulting from the use of these evaluation systems in which small to moderate levels of reliability are established (e.g., the measurement reliability for FFT ranges from 0.40 to 0.68 (Mihaly, et al., 2013)). Researchers' examination of observational instruments in the MET study found it challenging to establish high levels of reliability (Ho & Kane, 2013; Kane & Staiger, 2012). For any given teacher, scores varied from lesson to lesson, and for any given lesson scores varied from observer to observer (Ho & Kane, 2013; Kane & Staiger, 2012). This suggests the need for the development of stronger measures of instructional quality. Hill and colleagues (2012) go even further to argue that it is not enough to focus on the reliability of observational instruments alone and instead suggest that it we must ensure the development of reliable observation systems. These systems would require reliable and valid observation instruments along with scoring criteria capable of producing reliable evaluation scores and processes for rater recruitment, training, and certification (Hill, Charalambous & Kraft, 2012).

In the development of stronger measures, researchers must both know what practices or knowledge are most critical in developing and identifying quality teachers. These systems, often used in high-stakes evaluations of teachers, should be reexamined to ensure the focus of these evaluations is truly reflective of what it takes to be a quality teacher. Allen and colleagues, for instance, examined specific teacher practices using the CLASS evaluation system and concluded that classrooms with positive emotional climate sensitive to student needs and perspectives, that used diverse and engaging learning formats, and focused on analysis and problem solving were positively associated with higher levels of student achievement (Allen, et al., 2013). Examining specific attributes or points of content knowledge for teaching is also important as some teachers may be more effective at some forms of instruction and less effective in others, all influencing scores and knowledge of instruction (Darling-Hammond, et al., 2012). Researchers should continue to question and understand specific teacher practices or knowledge reflective of teacher quality and develop sound evaluative instruments based on these findings.

Once specific strategies are identified that reflect quality instruction and pedagogical content knowledge in the classroom, we must be sure that we can strongly identify and measure them in practice. This is of particular importance in urban classrooms, which may not resemble non-urban classrooms as they are often comprised of students from different backgrounds and students who may face difficult educational or social challenges. In these cases, it may be that commonly used measures that identify effective instructional practices or pedagogical content knowledge are not as applicable. The reliability of evaluation systems or content knowledge for teaching is likely to vary and be influenced by the context in which schools exist (Borman & Kimball, 2005). Therefore, it is possible that other dimensions of teachers' practices or pedagogical content knowledge not examined by these measures may be used to identify effective urban teachers. Thus, it is important to consider the context in which instruction takes place the development of observational measures and assessments of pedagogical content knowledge (Borman & Kimball, 2005).

4.2.5 Mediation

The models estimated in this study all consider mediators. SEM, as a method for estimating mediating and multi-level relationships, is useful for its handling of non-normal data. As it specifically relates to mediation, researchers argue for the use of bootstrapping as a resampling method to improve the precision of indirect effect estimates (Preacher, et al., 2010). Yet, the models I employ not only include mediation, but also multi-level modeling, simultaneously. The Mplus software used for the study's analysis does not allow for bootstrapping with multilevel models and instead uses a linearized sandwich estimator which only considers the variation among clusters which is favored among multilevel models (Asparouhov, 2004, 2006; Kovacevic, et al., 2006; Pierre & Saidi, 2008). Nevertheless, it should be noted that these estimates could be biased due to the modeling employed and software used (Kovacevic, et al., 2006).

4.2.6 A Limiting Database

While the MET database is the most comprehensive database on teachers and teaching to date, a more robust database with additional teacher characteristics, qualifications, or unobservable characteristics that may potentially contribute to teachers' effectiveness would allow for further, more extensive examination of teachers, particularly urban teachers. This database could include teachers' age, certification route, teacher preparation program, certification exam scores as well as measures of teachers' efficacy, grit (Duckworth, 2009), or sense of responsibility to their community or profession. Having these additional variables would allow researchers to examine more attributes of teachers in studies on teacher quality, as this study was limited by the data collected in MET. Furthermore, while the MET data were collected among six large districts, a more robust database would include not only more large districts but also more districts of all kinds in order for greater comparisons to be made. This would allow for larger sample sizes and greater power in estimating teacher effects. It would also allow for comparisons of urban districts with non-urban districts, which can be useful when comparative studies are warranted. It is equally as important that these more robust databases include stronger measures of teacher practices and pedagogical content knowledge from which stronger conclusions can be drawn. Conclusions drawn from this analysis were restricted by the data and the quality of the data made available on teachers from the MET study.

Chapter 5

CONCLUSIONS

At the heart of this study is the desire to find ways to improve student learning in urban communities, which continue to face a host of problems that impact their schools and achievement among their youth. Research has shown that teachers have a major impact on student learning and are considered by some researchers to be the single most important influence on student achievement (Sanders & Horn, 1998; Wenglinsky, 2000), thus, stressing the importance of having high quality teachers in the workforce (Darling-Hammond, 2003; Oakes, 2002). Yet, it is understood that the challenges faced in urban communities influence the practice of teaching in urban schools, often making it more difficult for urban teachers to contribute to student learning (Adamson & Darling-Hammond, 2012; Hanushek et al., 2004).

By using a framework for teacher quality that examines urban teacher attributes in relation to their value-added to student learning, I examined whether urban teacher characteristics and qualifications were predictive of their effectiveness, while also examining their instructional quality and pedagogical content knowledge as potential mediators. The sample of teachers drawn from the MET study and used in this analysis, taught in urban schools serving at least both 60% minority and 60% low-income students. This sample of urban teachers was established in order to focus on urban teachers serving our most academically at-risk students.

This study revealed there are several ways are able to improve student learning. First, results consistently indicate that increasing the levels of instructional quality among urban teachers may be a productive lever for improving student learning, in both ELA and math. Other strong indicators of urban teacher effectiveness include teachers' pedagogical content knowledge in math and teachers' race in ELA. Second, in terms of mediating effects, results demonstrated that mediators may help explain relationships between teacher attributes and their value-added to student learning. In one model, teachers' race, as mediated by their pedagogical content knowledge, was found to be predictive of their effectiveness in improving student learning in math. Lastly, contextual effects were also evident in ELA, where schools with more teachers with advanced degrees and higher levels of instructional quality are associated with greater impacts on students' learning of ELA.

The findings from this study are limited to the data used in the analysis, thus I caution against generalizing these findings to all urban schools and contexts. Nevertheless, these findings are useful in guiding future research and in potential policy and practice implications focused on improving teaching and learning in urban schools of which I discuss below.

5.1 Directions for Future Research

Several directions for future research should be considered. This includes addressing alternative methodological approaches for examining urban teacher quality, the use of alternative student outcomes in studies on teacher quality, and more studies that consider the specific characteristics of urban schools and contexts.

5.1.1 Methodological Approaches and Considerations

The methodological approach used in this study was ambitious considering multilevel analyses and mediating effects, with results indicating that future research should consider both. Furthermore, while the modeling employed was complex, it was also limiting. I consider several different methodological approaches that can be used in futures studies that may examine teacher quality in urban schools.

Race is a significant predictor of teachers' effectiveness in improving students' learning of ELA, where minority teachers are more effective than non-White teachers. This finding is powerful in the context of urban schools, which tend to serve mostly minority students. Previous research has found strong, positive relationships between teachers and student learning when the race of the teacher and the student were the same. I did not examine this exact relationship between teachers and students in this study, although the finding of minority urban ELA teacher effectiveness leads me to believe that this relationship might be evident in urban schools. Therefore, it is recommended that researchers continue to investigate the role of race in urban schooling, among a more generalizable sample. This would require matched comparison studies, specifically among urban teachers and schools, to understand if teachers are more effective in improving student learning in when teacher and student race are shared. Researchers should also question whether race matters more in urban schools than non-urban school – comparing differences in teacher effects when teacher and students shared race is a factor, between these two different school contexts. If indeed these relationships exist, the research must go deeper to untangle what shared classroom, school, community, or cultural identities influence the effectiveness of urban teachers on their students when race is shared.

Instructional quality was identified as a significant predictor of teacher effectiveness across all models, therefore future research should examine which specific instructional practices of urban teachers lead to their effectiveness in the classroom. It is important to note that the measure used in this study was an overall measure of instructional quality, combining individual indicators from CLASS and FFT evaluation frameworks. This overall measure consisted of 20 dimensions of instructional quality – from classroom management to instructional learning formats. In order to really leverage instructional quality in reform efforts, we must further investigate which specific practices or combinations of practices are most effective in urban classrooms and build the capacity to engage in those specific practices (Darling-Hammond, et al., 2012). These specific practices may not even be those measured under the two evaluation frameworks examined in this study (i.e., CLASS, FFT) as practices outside of these frameworks may be just as or even more critical to our understanding of urban teacher quality. Gay (2002, 2010), for example, argues that culturally responsive instruction in schools serving students of ethnically diverse populations can play a meaningful role in their learning and development. Duckworth (2009) also identifies teacher grit and level of life satisfaction as significant predictors of teachers' effectiveness in under-resourced public schools. Teacher attributes such as these that are not included in the CLASS or FFT frameworks should be examined along with other specific instructional practices, which may be predictive of urban teacher effectiveness.

The finding that pedagogical content knowledge mediates the relationship between teachers' race and their value-added to student learning warrants close examination of mediators in future studies on teacher quality. Most studies of teacher quality fail to examine any mediating effects, although there may be other teacher or teaching attributes that may account for differences between teacher characteristics or qualifications and their effectiveness in improving student learning. Understanding mediating relationships might be meaningful to reformers as they can focus on these attributes in the hiring, preparation, or development of teachers.

Furthermore, a comparative study of urban versus non-urban teachers rather than an exploratory study such as this may be more beneficial in teasing out similarities and differences in what makes an effective urban teacher versus a non-urban teacher and in our understanding schooling and teaching in these different contexts. Results from a study of this kind can be used to identify which predictors of urban teacher effectiveness differ significantly from those of non-urban schools. It might also further extrapolate differences in the experiences of teachers and students in these different schooling contexts. This type of study would require a much larger database on teachers and teaching, which would need to expand to include rural and suburban schools and districts.

Future research must also continue to recognize the nested nature of educational data and examine teacher quality accordingly, using multilevel models. This study used multilevel modeling to identify if any contextual effects on student learning existed above and beyond the individual effect of teachers. Results showed two contextual effects among urban ELA teachers – schools having more teachers with advanced degrees and with higher levels of education have significant impacts on teachers' effectiveness. Contextual effects may be just as or even more important than the individual effects of teachers on student learning. Contextual effects speak to the dynamics of the school above and beyond the teacher effect, which may not only help explain teacher effectiveness but also help us understand how students' experience schools where contextual effects exist. A growing number of studies examine contextual effects in relation to teacher quality, but researchers should continue to examine these effects in our efforts to understand teacher quality and reform urban education.

5.1.2 Measuring Teacher Quality

There still remains the question of whether or not student achievement, based on standardized assessment scores, is the only outcome that should be measured in understanding teacher quality. The use of standardized assessments has become part of the increasing debate over their utility in measuring student learning and in evaluating teacher performance. The argument against the use of standardized assessments is fueled by the frequency in which the assessments are administered, the reliability and validity these tests, and their use in high-stakes decisions within schools (Kohn, 2000; Smith & Fey, 2000). Researchers believe that it is the quantification of assessment scores and the ability to determine their reliability and validity that makes it easier for them to be used as data points in education decision-making (Kohn, 2000). Nevertheless, there are other outcomes related to students' experiences in the classroom that should be considered in efforts to understand and define teacher quality further. These outcomes can include measures of student attitudes (e.g., motivation, responsibility or attachment), their behaviors (e.g., engagement, attendance or study habits), or other measures of academic performance (e.g., grades, subject mastery) (Zins, et al., 2007).

However, most teacher quality studies to date have focused on using standardized measures of student achievement, knowing that there is some value in the use of these assessments (Smith & Fey, 2000). This includes the present study, which uses a framework on teacher quality where I only use measures based on student assessment scores as outcomes. Thus, researchers may consider examining other related outcomes of teachers' instruction in the classroom – either individually or in conjunction with assessment-based outcomes in future studies of teacher quality.

5.1.3 Strengthen Contextual Focus in Teacher Quality Studies

Urban teacher quality cannot be examined without a true understanding and focus on the context in which these schools exist. Local urban cultures, the urban political economy, the bureaucratic structure of urban schools, and the community and social service support networks serving urban communities often influence teaching and learning in urban schools (Oakes, et al., 2002).

For studies on urban schooling, for example, this is not only important in sampling procedures – developing a sample that is representative of urban schools – but in the development of measures used in studying teacher quality. For example, researchers must consider the context in which teacher observations/evaluations take place as instructional practices or knowledge for teaching may differ accordingly. Measures of teachers' practices or knowledge cannot be developed under a "one-size fits all" mindset. While structurally, urban schools resemble those of non-urban schools, they may have different characteristics or contexts, where different approaches to teaching may be necessary. Researchers have suggested that the reliability of evaluation systems is likely to vary and be influenced by the context in which schools exist (Borman & Kimball, 2005). Considering the characteristics of urban communities, schools, and classrooms, particularly those that serve the most academically at-risk, we must create and validate measures within these environments and classrooms.

5.2 Policy and Practice Implications

It is important to understand the potential implications of these findings on policy and practice. Understanding attributes of urban teacher effectiveness can be meaningful in improving both the quality of teaching in urban schools and student learning in these contexts. Below, I highlight the implications of these findings on the hiring of urban teachers, how we might target the preparation of urban teachers, and in the evaluation of teacher practices and knowledge for teaching in the classroom.

5.2.1 Implications for Hiring

Urban teachers' quality of instruction in both ELA and math and their pedagogical content knowledge in math were found to be significant predictors of teachers' effectiveness. Knowing this, educators can institute policies or practices with which they can identify teachers with strong instructional quality or pedagogical content knowledge prior to selecting them for employment in urban schools.

For example, as part of the hiring process, schools can require potential teacher candidates to teach several demonstration lessons or complete assessments of content knowledge for teaching prior to their hiring. These lessons or assessments can be scored based on the evaluation framework or scoring rubrics. Educators can decide how to determine if teachers are scoring high on these measures by using scores from teachers already in their workforce who are identified as highly effective or by other research-based criterion that identifies effective urban teachers based on scores of instructional quality or content knowledge for teaching. Researchers, for instance, conclude that teachers scoring in the top quartile on measures of the CLASS and FFT instructional quality frameworks are associated with the most effective teachers in ELA (Kane & Staiger, 2012). Although I argue once again that effective practices or knowledge for teaching are likely influenced by school context, thus high-levels of instruction or knowledge should be based on instructional practices or knowledge for teaching in specific contexts. Nevertheless, schools can use scores of instructional quality or pedagogical content knowledge in their hiring decisions in order to increase levels of instructional quality in urban schools, which is likely to contribute to up to 4.5 months of additional student learning annually.

Increasing levels of instructional quality or pedagogical content knowledge across the urban teacher workforce may also lead to improvements in staffing challenges often faced by urban schools. As previously stated, urban schools continue to face staffing challenges (Darling-Hammond, 2003; Hanushek et al., 2004; Ingersoll & Merrill, 2010, 2014). This is likely due to their lack of preparation to teach in urban contexts, which is often attributable to the curriculum and structure of teacher preparation programs (Haberman, 1987, 1995; Stotko, et al., 2007). Thus, understanding the specific practices or knowledge of effective teachers may alleviate some of those challenges in maintaining a stable teacher workforce in urban schools.

Teachers' race was also identified as a significant predictor of urban teacher effectiveness in ELA classrooms, with minority teachers shown to be more effective than White teachers. While I understand that teacher hiring decisions cannot be solely based on teachers' racial or ethnic identity, I think it is important to recognize this finding as it relates to previous research that finds positive effects on student learning when minority teachers teach students of their own race (Dee, 2004; Clotfelter, 2010; Hanushek, et al., 2005). Urban schools, particularly in this sample, are comprised mostly of minority students. Therefore, if it is indeed true that minority teachers are significantly more effective at improving student learning in ELA in urban schools, there is likely the case that when minority students' and teachers' race match, students' learning of ELA is improved. Research has shown that there remains a gap in the workforce between the percentage of minority students and the percentage of minority teachers in American schools, which persists mainly as a result of minority teacher turnover related to school culture and conditions (Ingersoll & May, 2011). Therefore, if teachers' race truly does have an influence on student's learning of ELA in these contexts, as demonstrated in this study and in previous research, we must address this gap by making stronger attempts to attract minority teachers to teach in urban schools, yet base hiring decisions on the quality of their instruction.

This implication should not be interpreted as an argument to segregate schools based on similar student-teacher characteristics. Instead, I am arguing that we must better understand this relationship and attempt to capitalize on how it may influence student learning. For example, it may be the shared cultural experiences that drive the relationship. It may be that minority teachers bring this knowledge into their classroom and instruction, developing curricula that are responsive to the cultures of the students who are being taught. If indeed this is the case, all teachers can build their understanding of culturally responsive instruction and try to incorporate any shared identities or experiences into the classrooms in order to improve student learning (Gay, 2002, 2010). For example, culturally congruent instruction may recognize students' culture in a lesson or classroom activity in order to connect with students' identities which may influence teachers' abilities to improve student learning, particularly in urban contexts where students may come from varied backgrounds and cultures (Au & Kawakami, 1994; Gay, 2002).

Based on contextual effects identified in this study, increasing the rates of teachers in urban schools with advanced degrees and with high levels of instructional quality is also likely to have impacts on student learning. This requires educators to focus their hiring on building cohorts of teachers in urban school who share these characteristics. This finding urges policymakers to not only understand teacher effectiveness from the individual teacher angle, but also from a contextual angle. Reformers must better understand and build upon the social dynamics between teachers with these significant attributes in urban schools and their likeliness to influence student learning. This can also include further investigation of communities of practice and teacher leadership in schools that may also contribute to these contextual effects. Understanding how teachers learn from each other and strengthening these mechanisms or efforts in schools is likely to can impact policy and practitioners decisions related to strengthening contextual effects in schools.

5.2.2 Implications for Teacher Preparation

Among the significant attributes, policymakers and practitioners are able to guide the preparation of teachers based on their instructional quality and pedagogical content knowledge. In order for schools to increase the number of teachers with significant attributes there must be an applicant pool with the content knowledge for teaching math and instructional quality necessary for highly effective instruction in urban schools. Teacher preparation, and even urban teacher recruitment programs such as Teach for America, must understand the need to prepare teachers for instruction in urban contexts and must address the current curriculum and structure of their programs (Haberman, 1987, 1995; Stotko, et al., 2007). This requires more knowledge around the key practices and the content knowledge for teaching math in urban contexts in order to train future urban teachers adequately. Teacher preparation programs are beginning to address this concern by focusing their curriculum on teaching or providing student teaching opportunities in urban schools, which have been shown to positively influence the experiences of urban teachers (Anderson & Stillman, 2012; Eckert, 2013).

Nevertheless, some argue that preparation to teach in urban contexts can only be done on the job, with a teacher or coach, a support network, or specific trainings (Haberman, 1987, 1995; Haberman & Post, 1998; Oakes, et al., 2002). Researchers suggest that teacher learning is a process whereby novice teachers and expert teachers learn from one another in a community of practice (Oakes, et al., 2002). Therefore, this implication can also be related to the further development of teachers once they are hired. Teacher preparation programs can provide a foundational knowledge of the instructional practices and math pedagogical content knowledge needed to instruct in urban schools, which can be built upon once teachers begin teaching in these contexts.

5.2.3 Implications for Evaluation

Policymakers and practitioners continue to find ways to identify effective teachers in their schools through evaluation processes. Findings from this study indicate that in urban schools, among multiple measures (Darling-Hammond, et al., 2003), they should focus on teachers' instructional quality and their pedagogical content knowledge in math. While most schools already use teacher evaluation frameworks focused on their instructional practices, urban schools should also consider using assessments of content knowledge for teaching among their math teachers to identify their most effective teachers. Considering and understanding the urban context in the overall system of evaluating teachers is highly important if we really desire a higher quality urban teacher workforce.

Policymakers and practitioners must also continue to question the quality of studies used to define and understand teacher quality – findings which are often used in teacher evaluation systems. Sound research that produces strong causal effects (i.e., randomization studies, strong VAM estimations) are more likely to be useful in efforts to reform teaching. This requires more quality data and measures on teachers and the quality of teaching, as well as more rigorous and appropriate methodological approaches to studying teacher quality. The MET database is a step in the right direction of improving databases on teachers and teaching. MET produced one of most robust databases developed on teachers to date, although the study had several limitations, which I discussed in previous sections. Addressing these limitations in the development of larger databases on teachers will be beneficial to future studies on teacher quality. Researchers must also consider that previous approaches used in studying teacher quality have been limiting. Many have only examined direct effects and a growing few have used VAM estimates of teacher effects in outcomes or have considered the multilevel nature of educational data. As it relates to urban teachers specifically, researchers must also develop measures and databases that consider the specific urban context. Results from research conducted with a sound contextual understanding can be most applicable and useful to policymakers in efforts to improve teaching in urban contexts.

5.2.4 A Focus on the Urban Context

Policymakers must begin to prioritize an agenda that focuses specifically on improving urban schools. Achievement gaps between urban and non-urban schools will continue to exist if we do not increase our focus on understanding how to improve teaching in these contexts (Darling-Hammond, 2004; Ladson-Billings, 2006; Lee, 2002). Teacher preparation programs must focus on understanding effective practices of teachers in urban schools, as they have traditionally taught from the perspective that their programs prepare teachers to teach in all settings (Haberman, 1987, 1995; Stotko, et al., 2007). This is not to dismiss other challenges urban schools face that influence teacher attrition (e.g., working conditions, lower wages), which should remain of great concern to policymakers and practitioners. Although the focus of this study is on recruiting more effective teachers in urban schools, if we continue to fail in the preparation of urban teachers urban schools are likely to continue to face staffing challenges (Abel & Sewell, 1999; Darling-Hammond, 2003) and the inequitable distribution of quality teachers across schools will remain (Eckert, 2013; Jacob, B., 2007; Ingersoll, 2004; Ingersoll & Merrill, 2010; Lankford, et al., 2002). Furthermore, urban students will continue to be served by teachers who are less-qualified or ill-prepared to effectively improve their learning (Adamson & Darling-Hammond, 2012; Eckert, 2013; Lankford, et al., 2002).

While there are current policies in place that consider teacher practices in the evaluations of teachers, particularly in high stakes cases, I argue that effective practices may be contextual (Borman & Kimball, 2005). Thus, researchers must continue to understand the context of urban schools in order to improve urban teaching and learning. The unique challenges of urban schools should be understood and considered when using attributes of teachers in defining and measuring teacher quality and in all aspects of reforming urban schools. Researchers and policymakers must increasingly see teaching and schooling in urban contexts as a unique field of study and begin to account for the specific challenges of urban communities in research on urban schools. If we can better identify and prepare high quality urban teachers, we can make a difference in student learning and can likely reduce the achievement gap between urban and non-urban schools (NAEP, 2014; NCES, 2013; Golding, et al., 2013).

The unique challenges of urban schools and communities has negatively influenced student learning in these contexts for many years. Increasing the effectiveness of urban teachers is likely to improve student achievement. I learned from this study that there are attributes of teachers that are predictive of their effectiveness in improving student learning in urban contexts. Policymakers can use these findings, along with future research on urban teachers, to better select and prepare teachers for urban schools. Results from this study can inform national conversations on urban teacher and school reform, drive future research on urban teachers and schooling, and ultimately help to improve the quality of education for our students who need it most.

REFERENCES

Abel, M. H., & Sewell, J. (1999). Stress and burnout in rural and urban secondary school teachers. *The Journal of Educational Research*, *92*(5), 287-293.

Adamson, F., & Darling-Hammond, L. (2012). Funding disparities and the inequitable distribution of teachers: Evaluating sources and solutions. *Education Policy Analysis Archives, 20*(37), 1-46.

Alexander, C., & Fuller, E. (2004, April). *Does teacher certification matter? Teacher certification and middle school mathematics achievement in Texas.* Paper presented at the annual meeting of the American Education Research Association, San Diego.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the Classroom Assessment Scoring System--Secondary. *School Psychology Review, 42*(1), 76-98.

Anderson, L., & Stillman, J. (2013). Student teaching's contribution to pre-service teacher development: A review of research focused on the preparation of teachers for urban and high-needs contexts. *Review of Educational Research, 83*(1), 3-69.

Asparouhov, T. (2004), "Weighting for unequal probability of selection in multilevel modeling." *Mplus Web Notes, 8*, 1-28.

Asparouhov, T. (2006), "General multi-level modelling with sampling weights. *Communications in Statistics – Theory and Methods, 35*(3), 439-460.

Asparouhov, T., & Muthén, B. (2010). Plausible values for latent variables using Mplus. Technical Report.

Au, K. H., & Kawakami, A. J. (1994). Cultural congruence in instruction. In E. R. Hollins, J. E. King, & W. C. Hayman (Eds.), *Teaching diverse populations: Formulating a knowledge base* (pp. 5-23). Albany: State University of New York Press.

Baker, B., & Corcoran, S. P. (2012). *The stealth inequities of school funding: How state and local school finance systems Perpetuate inequitable student spending.* Washington, DC: Center for American Progress.

Baker, B., & Duncombe, W. (2004). Balancing district needs and student needs: The role of economies of scale adjustments and pupil need weights in school finance formulas. *Journal of Education Finance*, 195-221.

Ballou, Dale. 2005. **"**Value-added assessment: Lessons from Tennessee.**"** In Value-Added Models in Education: Theory and Application, ed. Robert W. Lissitz. Maple Grove, MN: JAM.

Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6).

Bartlett, M. (1954). A further note on the multiplying factors for various c^2 approximations in factor analysis. *Journal of the Royal Statistical Society, 16*, 296-298.

Berger, L., Paxson, C., & Waldfogel, J. (2009). Income and child development. *Children and Youth Services Review*, *31*(9), 978-989.

Betts, J., Reuben, K., & Danenberg, A. (2000). *Equal resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California.* San Francisco, CA: Public Policy Institute of California.

Betts, J., Zau, A., & Rice, L. (2003). *Determinants of student achievement: New evidence from San Diego* (pp. 1-5821). San Francisco: Public Policy Institute of California.

Bill and Melinda Gates Foundation (2013a). *Measures of effective teaching longitudinal database.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Retrieved from *http://www.icpsr.umich.edu/icpsrweb/METLDB/index.jsp*

Bill and Melinda Gates Foundation (2013b). *Measures of effective teaching: 1-Study information.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Retrieved from http://www.icpsr.umich.edu/icpsrweb/METLDB/studies/34771.

Bliese, P. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco: Jossey-Bass.

Bondy, E., Ross, D., Hambacher, E., & Acosta, M. (2012). Becoming warm demanders: Perspectives and practices of first year teachers. *Urban Education, 48*(3), 420-450.

Borman, G., & Kimball, S. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106*(1), 3–20.

Bornstein, M., & Bradley, R. (Eds.). (2014). *Socioeconomic status, parenting, and child development*. New York, NY: Routledge.

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 Finance & Policy*, *1*(2), 176-216.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review, 95*(2), 166-171.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? Teacher attrition and student achievement*. NBER Working Paper 14022. Cambridge, MA: National Bureau of Economic Research.

Briggs, N., & MacCallum, R. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research, 38*, 25-56.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111-150.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions, 154*, 136-136.

Brunetti, G. (2006). Resilience under fire: Perspectives on the work of experienced, inner city high school teachers in the United States. *Teaching and Teacher Education, 22*(7), 812-825.

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

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27*, 509-540.

Cantrell, S., & Kane, T. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved June 8, 2014 from http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Cavalluzzo, L. (2004). *Is National Board Certification an Effective Signal of Teacher Quality?.* Alexandria, VA: *CNA Corporation*.

Centolanza, L. R. (2004, Fall). New Jersey teachers believe testing compromises sound practices. *ERS Spectrum, 22*(4).

Cheung, M. & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling, 12*, 598-619.

Cheung, M., Leung, K., & Au, K. (2006). Evaluating multilevel models in cross-cultural research: An Illustration with Social Axioms. *Journal of Cross-Cultural Psychology, 37*, 522-541.

Chou, V., & Tozer, S. (2008). What's urban got to do with it?: Meanings of "urban" in urban teacher preparation and development. In F. P. Peterman (Ed.), *Partnering to Prepare Urban Teachers: A Call to Activism* (pp. 1-20). New York: NY: Peter Lang Publishing.

Cohen, D. K., & Hill, H. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE Research Report No. RR-39). Philadelphia: Consortium for Policy Research in Education.

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

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

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

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F.D., York, R., (1966). *Equality of educational opportunity.* Washington, DC: Department of Health, Education and Welfare.

Coopersmith, J. (2009). *Characteristics of Public, Private, and Bureau of Indian Education Elementary and Secondary School Teachers in the United States: Results from the 2007-08 Schools and Staffing Survey. First Look. NCES 2009-324.* Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2009/2009324.pdf.

Crampton, F. E. (2003). "Unmet school infrastructure funding need as a critical educational capacity issue: setting the context", in Crampton, F.E. and Thompson, D.C. (Eds), *Saving America's School Infrastructure*, Greenwich, CT: Information Age Publishing, pp. 3-26.

Crampton, F. E. (2010). The economics and financing of urban schools: Toward a productive, solution-oriented discourse. *Educational Considerations*, *38*(1), 33-39.

Cudeck, R. (2000). Exploratory factor analysis. In H.E.A. Tinsley & S.D. Brown (Eds.), Handbook of applied multivariate statistics and mathematical modeling (pp. 265-296). New York: Academic Press.

Danielson, C. (2011). *The framework for teaching evaluation instrument: 2011 edition.* Princeton, NJ: The Danielson Group. Retrieved from https://www.teachscape.com/binaries/content/assets/teachscape-marketing-website/products/ffteval/2011-fft-evaluation-instrument.pdf.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1-44.

Darling-Hammond, L. & Youngs, P. (2002). Defining 'highly qualified teachers'; What does 'scientifically –based research' tell us? *Educational Researcher*, *31*(9), 13-25.

Darling-Hammond, L. (2003). Keeping good teachers: Why it matters, What leaders can do. *Educational Leadership, 60*(8), 6-13.

Darling-Hammond, L. (2004). The color line in American education: Race, resources, and student achievement. *Du Bois Review, 1*(02), 213-246.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race Ethnicity and Education, 10*(3), 245-260.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 8-15.

Darling-Hammond, L., Holtzman, D., Gatlin, S., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), 1-51.

Darling-Hammond, L., & Rustique-Forrester, E. (2005). The consequences of student testing for teaching and teacher quality. *Yearbook of the National Society for the Study of Education*, *104*(2), 289-319.

Decker, P., Mayer, D., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation*. Princeton, NJ: Mathematica Policy Research, Inc.

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

Desimone, L. (1999). Linking parent involvement with student achievement: Do race and income matter?. *The Journal of Educational Research*, *93*(1), 11-30.

Desimone, L., & Long, D. (2010). Teacher effects and the achievement gap: Do teacher and teaching quality influence the achievement gap between Black and White and high-and low-SES students in the early grades. *Teachers College Record, 112*(12), 3024-3073.

Dolton, P., & Van der Klaauw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics, 81*(3), 543-550.

Duncan, G., Brooks-Gunn, J., & Klebanov, P. (1994). Economic deprivation and early childhood development. *Child Development, 65*(2), 296-318.

Duncombe, W., & Yinger, J. (2005). How much more does a disadvantaged student cost?. *Economics of Education Review*, *24*(5), 513-532.

Durham-Barnes, J. (2011). The balancing act: The personal and professional challenges of urban teachers. *Perspectives on Urban Education*, *9*(1), 1-12.

Eckert, S. (2013). What do teaching qualifications mean in urban schools? A mixed-methods study of teacher preparation and qualification. *Journal of Teacher Education, 64*(1), 75-89.

Educational Testing Service (ETS) (2011) *Content Knowledge for Teaching: Innovations for the Next Generation of Teaching Assessments*. Princeton, NJ: Educational Testing Service. Retrieved from *http://www.ets.org/s/educator_licensure/ckt_handout.pdf*

Epstein, D. (2011). *Measuring Inequity in School Funding.* Washington, DC: Center for American Progress. Retrieved from http://cdn.americanprogress.org/wp-content/uploads/issues/2011/08/pdf/funding_equity.pdf

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review, 13*(1), 1-17.

Ehrenberg, R., Goldhaber, D., & Brewer, D. (1994). Do teachers' race, gender, and ethnicity matter? Evidence from NELS 88. *NBER Working Papers Series.* Cambridge, MA: National Bureau of Economic Research.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, *13*(1), 1-22.

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

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.): London: Sage.

Field, A. (2009). *Discovering statistics using SPSS, 3rd edition*. Thousand Oaks, CA: Sage Publications.

Frome, P., Lasater, B., & Cooney, S. (2005). *Well-qualified Teachers and High Quality Teaching: Are They the Same?*. Atlanta, GA: Southern Regional Education Board.

Fuller, E., & Alexander, C. (2004, April). *Does teacher certification matter? Teacher certification and middle school mathematics achievement in Texas.* Paper presented at the national meeting of the American Education Research Association, San Diego.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*(2), 106-116.

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice.* New York: Teachers College Press.

Gefen, D., Straub, D., & Boudreau, M. C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems*, *4*(1), 7.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75-96.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.

Goddard, R., Hoy, W., & Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479–507.

Goe, L. (2002). Legislating equity: The distribution of emergency permit teachers in California. *Education Policy Analysis Archives*, *10*(42), 1-36. Retrieved from http://epaa.asu.edu/ojs/article/viewFile/321/447.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.

Goldberg, L., & Velicer, W. (2006). Principles of exploratory factor analysis. In S. Strack (Ed.), *Differentiating normal and abnormal personality* (2nd ed.) (pp. 209-337). New York: Springer.

Goldhaber, D. (2002). "The Mystery of Good Teaching." *Education Next, 2*(1), 50-55.

Goldhaber, D. (2015). Exploring the Potential of Value-Added Performance Measures to Affect the Quality of the Teacher Workforce. *Educational Researcher*, *44*(2), 87-95.

Goldhaber, D., Gross, B., & Player, D. (2007). *"Are Public Schools Really Losing Their Best? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce."* CALDER Working Paper 12. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom & C. E. Finn, Jr. (Eds.), *Better teachers, better schools* (pp. 83–102). Washington, DC: The Thomas B. Fordham Foundation. Retrieved September 5, 2014, from http://www.edexcellencemedia.net/publications/1999/199907_betterteachersbetterschools/btrtchrs.pdf#page=91.

Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129-145.

Goldring, R., Gray, L., & Bitterman, A. (2013). *Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results from the 2011-12 Schools and Staffing Survey. First Look. NCES 2013-314.* Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2013/2013314.pdf.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gorsuch, R. L. (2003). Factor analysis. In J. A. Shinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 143-164). Hoboken, NJ: John Wiley.

Gregory, R. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston: Allyn and Bacon.

Gregory, A., Skiba, R., & Noguera, P. (2010). The achievement gap and the discipline gap two sides of the same coin?. *Educational Researcher, 39*(1), 59-68.

Guin, K. (2004). Chronic teacher turnover in urban elementary schools. *Education Policy Analysis Archives, 12*(42).

Haberman, M., & Post, L. (1998). Teachers for multicultural schools: The power of selection. *Theory into practice, 37*(2), 96-104.

Haberman, M., & Rickards, W. H. (1990). Urban teachers who quit: Why they leave and what they do. *Urban Education*, *25*(3), 297-303.

Hanushek, E., Kain, J., and Rivkin, S. (2004) Why public schools lose teachers. *The Journal of Human Resources, 39*(2), 326-354.

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005). *The market for teacher quality (Working Paper No. 11154).* Cambridge, MA: National Bureau for Economic Research.

Hanushek, E., Kain, J., & Rivkin, S. (1999). *Do higher salaries buy better teachers?* Working paper No. 7082. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, *100*(1), 84-117.

Harris, D., & Sass, T. (2006, March). *The effects of teacher training on teacher value-added.* In Florida State University), paper presented at the annual meeting of the American Education Finance Association. Tallahassee, FL: Florida State University.

Harris, D., & Sass, T. (2007). *Teacher training, teacher quality and student achievement* (Working Paper No. 3). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

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

Heck, R. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumaker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89-128). Mahwah, NJ: Erlbaum.

Heilig, J., & Jez, S. (2010). *Teach for America: A Review of the Evidence.* Boulder, CO, and Tempe, AZ: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from http://nepc.colorado.edu/files/PB-TeachAmerica-Heilig.pdf.

Heneman, H., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Brief No. RB-45). Philadelphia: Consortium for Policy Research in Education.

Herman, J.L., & Golan, S. (1993). Effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice, 12(*4), 20–25, 41–42.

Hill, C., Bloom, H., Black, A., & Lipsey, M. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Hill, H. (2010). *Mathematical Quality of Instruction (MQI). Coding Tool.* University of Michigan, Learning Mathematics for Teaching.

Hill, H., Rowan, B., & Ball, D. L. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 317-406.

Hill, H., Charalambous, C., & Kraft, M. (2012). When rater reliability is not enough teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*(2), 56-64.

Hirschfield, P. J. (2008). Preparing for prison? The criminalization of school discipline in the USA. *Theoretical Criminology*, *12*(1), 79-101.

Ho, A. D., & Kane, T. J. (2013). The reliability of classroom observations by school personnel. Seattle, WA: Bill and Melinda Gates Foundation.

Hofmann, D., & Gavin, M. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623-641.

Horn, J. (1965). A rational and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

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

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

Hunter, R., & Bartee, R. (2003). The achievement gap: Issues of competition, class, and race. *Education and Urban Society, 35*(2), 151-160.

Ingersoll, R. (2003a). The Teacher Shortage: Myth or Reality?. *Educational Horizons*, *81*(3), 146-52.

Ingersoll, R. (2003b). *Is there really a teacher shortage? A report co-sponsored by the Center for the Study of Teaching and Policy and the Center for Policy Research in Education.* Seattle, WA: University of Washington, Center for the Study of Teaching and Policy. Retrieved from https://depts.washington.edu/ctpmail/PDFs/Shortage-RI-09-2003.pdf.

Ingersoll, R. (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?* Washington, DC: Center for American Progress. Retrieved from http://cdn.americanprogress.org/wp-content/uploads/kf/Ingersoll-FINAL.pdf.

Ingersoll, R. & May, H. (2011). *Recruitment, retention, and the minority teacher shortage.* Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Ingersoll, R. & Merrill, L. (2010). Who's teaching our children? *Educational Leadership, 67*, 14-20.

Ingersoll, R. & Merrill, L. (2012). *Seven Trends: The Transformation of the Teaching Force.* Presented at the annual meeting of the American Educational Research Association. Retrieved from http://www.cpre.org/sites/default/files/workingpapers/1506_7trendsapril2014.pdf.

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. Retrieved from http://www.cpre.org/sites/default/files/workingpapers/1506_7trendsapril2014.pdf.

Jacob, A. (2012). Examining the relationship between student achievement and observable teacher characteristics: Implications for school leaders. *International Journal of Educational Leadership Preparation 7*(3), 1-13.

Jacob, B. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children, 17*(1), 129-153.

Jacob, B. A., & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education* (Faculty Research Working Paper Series RWP05-040). Cambridge, MA: Harvard University.

Jacob, B. A., Lefgren, L., & Sims, D. (2010). The persistence of teacher-induced learning. *Journal of Human Resources, 45*(4), 915-943.

Jeynes, W. H. (2005). A meta-analysis of the relation of parental involvement to urban elementary school student academic achievement. *Urban education*, *40*(3), 237-269.

Johnson, S., Kraft, M., & Papay, J. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1-39.

Jones, C. (2002). Teachers' perceptions of African American principals' leadership in urban schools. *Peabody Journal of Education*, *77*(1), 7-34.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

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

Kane, T., Rockoff, J., & Staiger, D. (2006, March). *What does certification tell us about teacher effectiveness? Evidence from New York City (NBER Working Paper No. 12155).* Cambridge, MA: National Bureau of Economic Research.

Kane, T., & Staiger, D. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). Cambridge, MA: National Bureau of Economic Research.

Kane, T., & Staiger, D. (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper.* MET Project Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf.

Kao, G., & Thompson, J. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology, 29*(1), 417-42.

Kaplan, D. (2009) *Structural equation modeling: Foundations and extensions, 2nd Edition.* Thousand Oaks, CA: SAGE Publications.

Keigher, A. (2010). *Teacher Attrition and Mobility: Results from the 2008-09 Teacher Follow-Up Survey. First Look. NCES 2010-353.* Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2010/2010353.pdf.

Kenny, D. (2014a). Measuring model fit. Retrieved from http://davidakenny.net/cm/fit.htm on January 15, 2015.

Kenny, D. (2014b). Mediation. Retrieved from http://davidakenny.net/cm/mediate.htm#IM on February 25, 2015.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools.* Portsmouth, NH: Heinemann.

Kovacevic, M., Huang, R., & You, Y. (2006). Bootstrapping for variance estimation in multi-level models fitted to survey data. *ASA Proceedings of the Survey Research Methods Section*, 3260-3269.

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

Kline, P. (1994). *An easy guide to factor analysis*. New York: Routledge.

Jonathan, K. (1991). Savage Inequalities: Children in America's Schools. *New York: Crown.*

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children.* San Francisco, CA: Jossey-Bass.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in the U.S. schools. *Educational Researcher, 35*(7), 3-12.

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

Laub, J. H. (2002). A century of delinquency research and delinquency theory. In M. K. Rosenheim, F. E. Zimring, D. S. Tanenhaus, & B. Dohrn (Eds.), *A century of juvenile justice* (pp. 179−205). Chicago: University of Chicago Press

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity?. *Educational Researcher, 31*(1), 3-12.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin, 126*(2), 309.

Lippman, L., Burns, S., & McArthur, E. (1996). *Urban schools: The challenge of location and poverty.* Washington, D.C.: National Center for Education Statistics, U.S. Department of Education, 1996.

Loeb, S., & Page, M. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics, 82*(3), 393-408.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44-70.

MacKinnon, D., Fairchild, A., & Fritz, M. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. Psychological bulletin, 97(3), 562.

Marsh, H., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*, *73*(1), 107.

Marvel, J., Lyter, D., Peltola, P., Strizek, G., Morton, B., & Rowland, R. (2007). *Teacher Attrition and Mobility: Results from the 2004-05 Teacher Follow-Up Survey. NCES 2007-307.* Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2007/2007307.pdf.

Matsko, K., & Hammerness, K. (2014). Unpacking the "urban" in urban teacher education: Making a case for context-specific preparation. *Journal of Teacher Education, 65*(2), 128-144.

McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND Corporation.

McCaffrey, D., Lockwood, J., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67-101.

McWhorter, J. (2000). *Losing the race.* New York, NY: Free Press.

Michigan Department of Education (2013). *Michigan Educational Assessment Program (MEAP).* Lansing, MI: Michigan Department of Education. Retrieved on August 9, 2013 from http://www.michigan.gov/mde/0,1607,7-140-22709_31168_40135---,00.html.

Mihaly, Kata, McCaffrey, D., Staiger, D., & Lockwood, J. (2013). A composite estimator of effective teaching. Seattle, WA: Bill and Melinda Gates Foundation.

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

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

Musil, C., Jones, S., & Warner, C. (1998). Structural equation modeling and its relationship to multiple regression and factor analysis. *Research in Nursing & Health*, *21*(3), 271-281.

Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox & I. Kreft (eds.), *Multilevel Modeling, A Special Issue of Sociological Methods & Research,* 22, 376-398.

Muthén, B., & Satorra, A. (1989) Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel Analysis of Educational Data.* San Diego, CA: Academic Press.

Muthén, L., & Muthén, B. (2008). Mplus (Version 5.1). *Los Angeles, CA: Muthén & Muthén*.

Muthén, L., & Muthén, B. (1998-2012). Mplus User's Guide. Seventh Edition. *Los Angeles, CA: Muthén & Muthén*.

National Center for Education Statistics (NCES) (2014). *Identification of rural locales.* Retrieved on February 22, 2015 from http://nces.ed.gov/ccd/rural_locales.asp.

National Center for Education Statistics (NCES) (2012). *Beginning K-12 Teacher Characteristics and Preparation by School Type, 2009.* Retrieved on June 8, 2014 from http://files.eric.ed.gov/fulltext/ED538983.pdf on March 16, 2014.

National Center for Education Statistics (NCES) (2013a). *Nations Report Card – National Assessment of Educational Progress (NAEP).* Retrieved on May 15, 2014 from http://nces.ed.gov/nationsreportcard/about/district.aspx.

National Center for Education Statistics (NCES) (2013b). *Schools and Staffing Survey (SASS).* Retrieved from http://nces.ed.gov/surveys/sass/index.asp.

National Center for Education Statistics (NCES) (2002) *Statistical Standards.* Retrieved from http://nces.ed.gov/statprog/2002/std2_6.asp.

Neiman, S. (2011). *Crime, Violence, Discipline, and Safety in U.S. Public Schools: Findings From the School Survey on Crime and Safety: 2009–10* (NCES 2011-320). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Newmann, F., Bryk, A., & Nagaoka, J. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago, IL: Consortium on Chicago School Research. Retrieved on June 1, 2014 from https://ccsr.uchicago.edu/sites/default/files/publications/p0a02.pdf.

Noguera, P. (2003). *City schools and the American dream: Reclaiming the promise of public education.* New York: Teachers College Press.

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

Oakes, J., Franke, M., Quartz, K., & Rogers, J. (2002). Research for high-quality urban teaching: Defining it, developing it, assessing it. *Journal of Teacher Education*, *53*(3), 228.

Ogbu, J. (1987). Variability in minority student performance: A problem in search of an explanation. *Anthropology and Education Quarterly, 18*(4), 312-334.

Osborne, J., & Costello, A. (2009). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, *12*(2), 131-146.

Palardy, G. & Rumberger, R. (2008). Teacher effectiveness in first grade: the importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis 30*(2), 111-140.

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore: Paul H. Brookes.

Pierre, F., & Saïdi, A. (2008). Implementing resampling methods for design-based variance estimation in multilevel models: Using HLM6 and SAS together. *JSM Proceedings of the Section on Survey Research Methods,* 788-795.

Ponitz, C., Rimm-Kaufman, S., Grimm, K., & Curby, T. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review*, *38*(1), 102-120.

Preacher, K., Zyphur, M., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209.

Price, J. H., & Everett, S. A. (1997). Teachers' perceptions of violence in the public schools: The Metlife survey. *American Journal of Health Behavior*, *21*(3), 178-86.

Quartz, K., Lyons, K., Masyn, K., Olsen, B., Anderson, L., Thomas, A., Goode, J., & Horng, E. (2004). *Urban Teacher Retention Policy: A Research Brief.* Los Angeles, CA: UCLA's Institute for Democracy, Education, and Access (IDEA).

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel Structural Equation Modeling. In Sik-Yum Lee (Ed.), *Handbook of Latent Variable and Related Models*. Amsterdam, The Netherlands: Elsevier Science.

Raudenbush, S. (2004). What are value-added models estimating and what does this imply for statistical practice?. *Journal of Educational and Behavioral Statistics, 29*(1), 121-129.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Raykov, T., & Marcoulides, G. (2006). *A first course in structural equation modeling, 2nd Edition.* New York: Psychology Press

Raymond, M., & Fletcher, S. (2002). The Teach for America Evaluation. *Education Next*, *2*(1), 62-68.

Reynolds, C., Livingston, R., & Willson, V. (2006). *Measurement and assessment in education*. Boston. Pearson.

Rice, J. (2003) *Teacher quality: Understanding the effectiveness of teacher attributes.* Economic Policy Institute, Washington, DC.

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

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

Rothstein, J. (2011). *Review of "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project."* Boulder, CO: National Education Policy Center. Retrieved July 6, 2014 from http://nepc.colorado.edu/thinktank/review-learning-about-teaching.

Rothstein, R. & Wilder, T. (2005, October 24). *The many dimensions of racial inequality.* Paper presented at the Social Costs of Inadequate Education Symposium. Teacher College, Columbia University, New York.

Rowan, B., Correnti, R., & Miller, R. (2002). What Large-Scale Survey Research Tells Us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools. *The Teachers College Record*, *104*(8), 1525-1567.

Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment in special and inclusive education* (10th ed.). Boston: Houghton Mifflin.

Sanders, W., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299–311.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, *12*(3), 247-256.

Scafidi, B., Sjoquist, D., & Stinebrickner, T. (2007). Race, poverty, and teacher mobility. *Economics of Education Review, 26*(2), 145-159.

Schochet, P. and Chiang H. (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains* (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Shann, M. H. (1998). Professional commitment and satisfaction among teachers in urban middle schools. *The Journal of Educational Research*, *92*(2), 67-73.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-23.

Shumacker, R., & Lomax, R. (2010). *A beginner's guide to structural equation modeling, 3rd edition.* New York: Routledge.

Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Instruction and achievement in Chicago elementary schools.* Chicago: Consortium on Chicago School Research.

Smith, D. L., & Smith, B. J. (2006). Perceptions of violence: The views of teachers who left urban schools. *The High School Journal*, *89*(3), 34-42.

Snijders, T. 'Power and Sample Size in Multilevel Linear Models'. In: B.S. Everitt and D.C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science.* Volume 3, 1570-1573. Chicester (etc.): Wiley, 2005.

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334-344.

Socias, M., Chambers, J., Esra, P., & Shambaugh, L. (2007). *The distribution of teaching and learning resources in California's middle and high schools* (Issues & Answers Report, REL 2007–No. 023). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs.

Stecher, B., & Barron, S. (1999). *Quadrennial milepost accountability testing in Kentucky.* CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

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

Teachstone Training (2014). *The CLASS Tool.* Charlottesville, VA: Teachstone Training. Retrieved on August 15, 2013 from http://teachstone.com/the-class-system/

Thomas, L. (1995). The use of second-order factor analysis in psychological research. Proceedings of the Annual Meeting of the Southwestern Psychological Association, San Antonio.

Thompson, F., & Levine, D. (1997). Examples of easily explainable suppressor variables in multiple regression research. *Multiple Linear Regression Viewpoints, 24*(1), 11.

U.S. Department of Education (2006). Schools and Staffing Survey, 2003–04, Public School, BIA School, and Private School Data Files. Washington, DC: National Center for Education Statistics.

Vandevoort, L., Amrein-Beardsley, A. & Berliner, D. (2004, September 8). National board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46), 1-117.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helms (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). New York: Guilford.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

Wacquant, Loïc (2001) 'Deadly Symbiosis: When Ghetto and Prison Meet and Mesh', in D. Garland (ed.) *Mass Imprisonment: Social Causes and Consequences*, pp. 82–120. London: Sage Publications.

Weglinsky, H. (2000). How Teaching Matters: Bringing the Classroom Back into Discussions of Teacher Quality (Policy Information Center Report). *Princeton, NJ: Educational Testing Service*.

Winters, M. A., & Cowen, J. M. (2013). Would a value-added system of retention improve the distribution of teacher quality? A Simulation of Alternative Policies. *Journal of Policy Analysis and Management, 32*(3), 634-654.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354-365.

Zhou, M. (2003). Urban education: Challenges in educating culturally diverse children. *Teachers College Record, 105*(2), 208-225.

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation, 17*(2-3), 191-210.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Appendix A

IRB APPROVAL LETTER

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	October 31, 2014
TO:	Akisha Jones
FROM:	University of Delaware IRB
STUDY TITLE:	[526119-2] Dissertation: Examining Teacher and Teaching Quality through Predictors of Urban Teacher Effectiveness
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	October 31, 2014
EXPIRATION DATE:	November 12, 2015
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Appendix B

ELA MSEM MODEL VARIABLE DESCRIPTIVES
(STATE ASSESSMENT VAMS)

Means and Standard Deviations of Variables in MSEM Model using ELA State Assessments

	M	S.D.
Within (Teacher) Level		
Male	0.13	0.3
White	0.48	0.5
Years Exp. In District	1.65	0.7
Master	0.45	0.5
Between (School) Level		
Male (sm)	0.14	0.2
White (sm)	0.48	0.4
Years Exp. In District (sm)	5.53	3.8
Master (sm)	0.42	0.4
Mediators/Dependent Variable		
Instructional Quality (Math)	0.00	0.1
Ped. Content Knowledge (Math)	63.29	1.8
VAM (ELA State Assessments)	0.00	0.1

Note: sm= school mean; All predictors and mediators were grand-mean centered in the model.

Variable Correlations from MSEM Model using ELA State Assessments

Within (Teacher) Level							
	Male	White	Years Exp. In Dist.	Master	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male	1.00	-	-	-	-	-	-
White	0.09	1.00	-	-	-	-	-
Years Exp. In Dist.	-0.02	-0.02	1.00	-	-	-	-
Master	-0.08	-0.24	-0.01	1.00	-	-	-
VAM (ELA State Assess)	-0.06	-0.04	-0.01	0.07	1.00	-	-
Instruc. Quality	0.07	0.34	0.01	-0.09	0.00	1.00	-
Ped. Content Knowledge	-0.06	-0.08	0.13	-0.17	0.07	0.01	1.00
Between (School) Level							
	Male (sm)	White (sm)	Years Exp. In Dist. (sm)	Master (sm)	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male (sm.)	1.00	-	-	-	-	-	-
White (sm.)	0.17	1.00	-	-	-	-	-
Years Exp. In Dist. (sm.)	-0.13	-0.08	1.00	-	-	-	-
Master (sm)	-0.20	-0.36	-0.13	1.00	-	-	-
VAM (ELA State Assess)	-0.44	-0.04	0.01	-0.03	1.00	-	-
Instructional Quality	-0.01	0.26	-0.57	-0.29	-0.04	1.00	-
Ped. Content Knowledge	-0.07	0.09	-0.11	0.60	0.36	-0.03	1.00

Note: sm= school mean

Appendix C

ELA MSEM MODEL VARIABLE DESCRIPTIVES
(SAT-9 SUPPLEMENTAL READING ASSESSMENT VAMS)

Means and Standard Deviations of Variables in MSEM Model using SAT-9 ELA Supplemental Assessments

	M	S.D.
Within (Teacher) Level		
Male	0.13	0.3
White	0.48	0.5
Years Exp. In District	1.65	0.7
Master	0.45	0.5
Between (School) Level		
Male (sm)	0.14	0.2
White (sm)	0.48	0.4
Years Exp. In District (sm)	2.21	0.8
Master (sm)	0.42	0.4
Mediators/Dependent Variable		
Instructional Quality (Math)	0.00	0.1
Ped. Content Knowledge (Math)	63.39	1.9
VAM (SAT-9 ELA Supplemental Assessments)	-0.01	0.1

Note: sm= school mean; All predictors and mediators were grand-mean centered in the model.

Variable Correlations from MSEM Model using SAT-9 ELA Supplemental Assessments

Within (Teacher) Level	Male	White	Years Exp. In Dist.	Master	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male	1.00	-	-	-	-	-	-
White	0.09	1.00	-	-	-	-	-
Years Exp. In Dist.	-0.02	-0.03	1.00	-	-	-	-
Master	-0.08	-0.24	-0.01	1.00	-	-	-
VAM (SAT-9 Assess)	-0.06	-0.05	-0.01	0.09	1.00	-	-
Instruc. Quality	0.07	0.34	0.02	-0.08	0.00	1.00	-
Ped. Content Knowledge	-0.11	0.03	0.06	0.01	0.17	0.02	1.00

Between (School) Level	Male (sm)	White (sm)	Years Exp. In Dist. (sm)	Master (sm)	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male (sm.)	1.00	-	-	-	-	-	-
White (sm.)	0.17	1.00	-	-	-	-	-
Years Exp. In Dist. (sm.)	-0.14	-0.11	1.00	-	-	-	-
Master (sm)	-0.19	-0.37	-0.10	1.00	-	-	-
VAM (SAT-9 Assess)	-0.44	-0.02	0.05	-0.14	1.00	-	-
Instructional Quality	0.00	0.23	-0.59	-0.41	-0.07	1.00	-
Ped. Content Knowledge	-0.09	-0.13	0.04	0.13	0.34	-0.40	1.00

Note: sm= school mean

Appendix D

MATH MSEM MODEL VARIABLE DESCRIPTIVES
(STATE ASSESSMENT VAMS)

Means and Standard Deviations of Variables in MSEM Model using Math State Assessments

	M	S.D.
Within (Teacher) Level		
Male	0.21	0.4
White	0.46	0.5
Years Exp. In District	1.60	0.8
Master	0.49	0.5
Between (School) Level		
Male (sm)	0.24	0.3
White (sm)	0.42	0.4
Years Exp. In District (sm)	5.50	3.9
Master (sm)	0.48	0.4
Mediators/Dependent Variable		
Instructional Quality (Math)	-0.01	0.1
Ped. Content Knowledge (Math)	54.64	4.0
VAM (Math State Assessments)	0.00	0.1

Note: sm= school mean; All predictors and mediators were grand-mean centered in the model.

Variable Correlations from MSEM Model using Math State Assessments

Within (Teacher) Level

	Male	White	Years Exp. In Dist.	Master	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male	1.00	-	-	-	-	-	-
White	0.04	1.00	-	-	-	-	-
Years Exp. In Dist.	-0.03	0.00	1.00	-	-	-	-
Master	-0.14	-0.30	0.05	1.00	-	-	-
VAM (State Assess)	0.03	-0.03	0.04	-0.03	1.00	-	-
Instruc. Quality	-0.23	-0.03	0.01	0.12	0.03	1.00	
Ped. Content Knowledge	0.11	0.34	-0.07	-0.21	0.13	-0.17	1.00

Between (School) Level

	Male (sm)	White (sm)	Years Exp. In Dist. (sm)	Master (sm)	VAM (State Assess)	Instruc. Quality	Ped. Content Knowledge
Male (sm.)	1.00	-	-	-	-	-	-
White (sm.)	0.14	1.00	-	-	-	-	-
Years Exp. In Dist. (sm.)	0.06	0.01	1.00	-	-	-	-
Master (sm)	-0.35	-0.48	-0.04	1.00	-	-	-
VAM (State Assess)	-0.17	0.13	0.03	0.25	1.00	-	-
Instructional Quality	0.05	0.16	0.18	-0.36	0.30	1.00	-
Ped. Content Knowledge	0.26	-0.05	0.13	-0.13	-0.06	0.00	1.00

Note: sm= school mean

Appendix E

MATH MSEM MODEL VARIABLE DESCRIPTIVES
(BAM SUPPLEMENTAL MATH ASSESSMENT VAMS)

Means and Standard Deviations of Variables in MSEM Model using BAM Math Supplemental Assessments

	M	S.D.
Within (Teacher) Level		
Male	0.21	0.5
White	0.46	0.8
Years Exp. In District	1.60	0.5
Master	0.49	0.4
Between (School) Level		
Male (sm)	0.24	0.3
White (sm)	0.42	0.4
Years Exp. In District (sm)	5.48	3.9
Master (sm)	0.47	0.4
Mediators/Dependent Variable		
Instructional Quality (Math)	-0.01	0.1
Ped. Content Knowledge (Math)	54.63	4.0
VAM (BAM Math Supplemental Assessments)	-0.01	0.1

Note: sm= school mean; All predictors and mediators were grand-mean centered in the model

Variable Correlations from MSEM Model using BAM Supplemental Assessments

Within (Teacher) Level	Male	White	Years Exp. In Dist.	Master	VAM (BAM Assess)	Instruc. Quality	Ped. Content Knowledge
Male	1.00	-	-	-	-	-	-
White	0.04	1.00	-	-	-	-	-
Years Exp. In Dist.	-0.03	0.00	1.00	-	-	-	-
Master	-0.14	-0.30	0.05	1.00	-	-	-
VAM (BAM Assess)	0.00	-0.01	0.00	-0.02	1.00	-	-
Instruc. Quality	-0.24	-0.03	0.01	0.12	0.07	1.00	-
Ped. Content Knowledge	0.11	0.34	-0.06	-0.21	0.13	-0.17	1.00

Between (School) Level	Male (sm)	White (sm)	Years Exp. In Dist. (sm)	Master (sm)	VAM (BAM Assess)	Instruc. Quality	Ped. Content Knowledge
Male (sm.)	1.00	-	-	-	-	-	-
White (sm.)	0.14	1.00	-	-	-	-	-
Years Exp. In Dist. (sm.)	0.07	0.01	1.00	-	-	-	-
Master (sm)	-0.36	-0.48	-0.05	1.00	-	-	-
VAM (BAM Assess)	-0.26	0.05	0.17	0.13	1.00	-	-
Instructional Quality	0.06	0.16	0.18	-0.38	0.30	1.00	-
Ped. Content Knowledge	0.26	-0.05	0.11	-0.18	-0.52	-0.02	1.00

Note: sm= school mean