Title

Investigating the Relationships Among Teacher Social Capital, Teaching Practice, and Student Achievement Across Measures and Models

Author
Schaaf, Kevin James

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Investigating the Relationships

Among Teacher Social Capital, Teaching Practice, and Student Achievement

Across Measures and Models

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Kevin J. Schaaf

2017

ABSTRACT OF THE DISSERTATION

Investigating the Relationships

Among Teacher Social Capital, Teaching Practice, and Student Achievement

Across Measures and Models

by

Kevin J. Schaaf

Doctor of Philosophy in Education

University of California, Los Angeles, 2017

Professor José-Felipe Martínez-Fernandez, Chair

The quality of teachers is now a central focus of policy and research. In spite of this focus, two key gaps exist, and this dissertation helps to fill them: First, mediation models involving the study of the effects of teaching are rare; instead, most current investigations of teachers focus primarily on the effects of teach*ers*, with less attention to the contexts that may impact the quality of teach*ing* practice. This dissertation considers how teaching practice mediates the effects of key contextual factors in the school environment, specifically the social capital available to teachers. By considering the impacts of teacher social capital as well as those of teaching practice, this dissertation broadens the scope of the investigation into how teachers might impact student learning. Second, research on teachers has been conducted using widely different instruments and models, and there is a lack of knowledge about the consequences of particular choices of measures and modeling assumptions. This dissertation compares the relationships among teacher social capital, teaching practice, and student achievement across single- and multilevel models and models employing latent and manifest factors/indicators. In addition, this dissertation offers conceptual and empirical analyses of the survey items included in the instruments and a comparison of inferences between expert observation ratings and student survey responses to better understand what aspects of teaching are adequately reflected in the measures employed and how this may influence the estimated relationships.

The dissertation of Kevin J. Schaaf is approved.

Richard Desjardins

Mark P. Hansen

Meredith Phillips

Michael H. Seltzer

José-Felipe Martínez-Fernandez, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

ACKNOWLEDGEMENTS

I wish to acknowledge the financial support from the AERA-MET Dissertation Fellowship and the UCLA Department of Education. The views and findings here are my own and do not reflect the views or policies of these funding agencies.

My work owes much to the support and guidance of my advisor and committee chair, Dr. José-Felipe Martínez-Fernandez. His patience and insight were invaluable throughout. In addition, I am grateful for the mentoring I received from my committee: Dr. Michael Seltzer, Dr. Meredith Phillips, Dr. Mark Hansen, and Dr. Richard Desjardins, each of whom contributed unique and valuable perspective and added immeasurably to my thinking.   I wish to especially thank Dr. Meredith Phillips, whose inspiring teaching launched me toward my doctoral studies.

I am grateful to all my past colleagues and friends at UCLA, whose support, intellect, and helpfulness are models for me as a scholar and a person. In particular, I want to thank Dr. Megan Kuhfeld, who jumpstarted my work countless times with perceptive feedback, Dr. Jennifer Ho, Dr. Jane Li, Danny Dockterman, Liz Perez, Jason Tsui, Dr. Melissa Goodnight, Dr. Glory Tobiason, Dr. Alejandra Priede, Dr. Kat Schenke, Dr. Kyo Yamshiro, Dr. Jon Schweig, Dr. Jordan Rickles, Dr. Scott Monroe, and Dr. Kate Riedell.

I thank all my teachers who have inspired me to keep teaching and keep learning, especially the late, great Doug Fix, Amy Carbonel, Sister Carol Seidl, Stewart Ouwinga, M. Jean Jokippi, and Dr. Rhoda Flaxman.

Most of all, I am grateful for the support of my wonderful family, especially my parents, whose love of learning and unwavering faith, hope, and love have put me where I am, Phuong, my life partner who has stuck by me through every challenge, and our children, who are the source of my strength and joy.

VITA

EDUCATION

2009 **MPP,** *University of California, Los Angeles*, Los Angeles, CA.

2001 **M.Ed.,** *University of California, Los Angeles.* Multiple Subject CLAD credential.

1995 **A.B.,** *Brown University*, Providence, R.I. English and American Literature.

RESEARCH AND PUBLIC POLICY EXPERIENCE

2012 – 2017 *National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA.* **Graduate Student Researcher.**

2011 *Los Angeles Unified School District and UCLA: Classroom and School Environment Survey*. **Project Coordinator and Lead Research Assistant.**

2009 – 2012 *United Teachers Los Angeles*. **Research Consultant**.

2010 – 2011 *Los Angeles Educational Research Consortium*. **Research Consultant.**

2009 – 2010 *Los Angeles School Development Institute.* **Research Consultant**.

SELECTED PUBLICATIONS

Martinez, J. F., Taut, S., & Schaaf, K. (2016). Classroom Observation for Evaluating and Improving Teaching: An International Perspective. *Studies in Educational Evaluation, 49*, 15-29.

Schaaf, K., Leon, S., and Wang, J. (2016). Evaluation of the Impact of the Magnet Schools Assistance Program on Student Achievement in New Haven Public Schools: 2014-15. Report to District. CRESST.

Schaaf, K., & Dockterman, D. (2014). VAM in Greek, English, and Implication: Explanations of Different Models and their Effects on Aggregate and Individual Teacher Outcomes. *InterActions: UCLA Journal of Education and Information Studies, 10(1).*

Teacher Effectiveness Workgroup, United Teachers Los Angeles. (2012). *UTLA's Teacher Development and Evaluation Framework.* March 6, 2012.

Phillips, M., Yamashiro, K., Schaaf, K., and Schweig, J. (2011). The Los Angeles Unified School District Pilot of Classroom and School Environment Surveys: Technical Report Exploring Reliability and Validity in Nine School Improvement Grant Schools. Dec. 18, 2011.

TEACHING EXPERIENCE

2016 – 2017 *University of California, Los Angeles – Teacher Education.* ***Field Supervisor.***

2016 *California State University, Long Beach – Ed Leadership.* ***Adjunct Faculty.***

2008 *Teach For America – Los Angeles Institute.* **Corps Member Advisor**.

2000 – 2007 *Lawndale Elementary School District***.** **Teacher**.

1997 – 1999 *Nah-Tah-Wahsh Public School Academy*, Hannahville, MI. **Tech Assistant.**

1995 – 1997 *East Moorhead Elementary, Teach For America.* Moorhead, MS. **Teacher**.

ORGANIZING and COMMUNITY EXPERIENCE

2015 *Committee to Recall Centinela Valley School Board Members.* **Founder.**

2013 *The Teaching Diablogue.* *http://teachingdiablogue.wordpress.com*. **Founder.**

2009 – 2011 *St. Catherine Laboure Advisory School Board*, Torrance, CA. **Co-Chair.**

2009 – 2010 *Franklin D. Roosevelt Elementary PTA.* Lawndale, CA. **President.**

1998 – 2000 *Sunflower County Freedom Project, Sunflower, MS.* **School Organizer.**

AWARDS and HONORS

AERA-MET Dissertation Fellowship, 2015; Eugene V. Cota-Robles Fellowship, 2011-2015

Graduate Summer Research Mentorship, 2013; Marion Wells Fellowship, 2012-13;

Tyler Fellowship, 2011-12; Martin Fellowship: Outstanding Academic Achievement, 2009

PTA Outstanding Service Award, 2005; Lawndale Rotary Grant, 2004

Chapter I. Introduction

This dissertation is motivated by two perceived gaps in the research literature: First, although the importance of teaching quality is increasingly recognized, less attention has been paid to the broad range of factors that substantially influence the quality of teaching. One of these factors is social capital: how teachers interact with key adults and the implicit norms that govern the school community. A substantial body of literature suggests that strengthening teacher social capital holds promise as a means to improving teaching quality (e.g., Bryk & Schneider, 1996; Fullan & Hargreaves, 2013; Leana & Pil, 2006). Second, to properly scrutinize the relationship between teacher social capital, teaching practices, and student outcomes requires valid measurement and modeling of these constructs. Current measures have a great deal of room for improvement in terms of reliability, validity, and practicality, and the majority of research to date has been conducted using models which ignore two critical aspects of the data: nesting structure and measurement error. In response to these gaps, this dissertation investigates how the relationships among teacher social capital, teaching practices, and student achievement differ depending on the measures and models.

Research Questions

This dissertation investigates the following research questions:

1. What factors of teacher social capital can be identified using the Teacher Working Conditions Survey?

2. What factors of teaching practice can be identified using the Tripod student perception survey items? How do these factors compare to those identified in the Framework for Teaching observation ratings?
3. What are the relationships among teacher social capital, teaching practice and student achievement?
    a. How do the relationships among these constructs differ when estimated with a multilevel structural equation model versus a single-level structural equation model?
    b. How do the relationships among these constructs differ when estimated with a multilevel structural equation model that uses latent factors to represent the social capital and teaching practice constructs versus a model that uses manifest (observed) indicators?
    c. How do the relationships among the constructs differ when teaching practice is measured by expert observation ratings rather than students' survey responses?

To address these questions, I use the Measures of Effective Teaching (MET) database (Kane & Staiger, 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013). The MET dataset is to-date the single largest and most comprehensive dataset designed to examine questions of teaching quality/teacher effectiveness. It therefore provides a rich source of data for considering important questions related to the factors that influence how well teachers teach. It includes data on over 150,000 students and nearly 3,000 teachers, with ratings of teachers on six different observational rubrics and more than fifty student perception items, two surveys of teachers to measure school conditions, and several measures of student achievement outcomes.

I focus on middle school math classrooms because of evidence of larger teacher effects and greater variation in teachers' effectiveness in math than in reading (Hanushek, Kain, & Rivkin, 1998; Harris & Sass, 2011; Kraft & Papay, 2014) and prior MET research suggesting greater precision and reliability for math estimates (Kane, McCaffrey, Miller, and Staiger, 2013; Mihaly et al, 2013); evidence also suggests greater reliability for older students in responding to surveys (Peterson et al, 2000) and these findings are echoed in MET research that indicated that middle school indicators were more reliable than elementary school indicators and there was much greater year-to-year stability in middle school classroom estimates (Mihaly et al, 2013).

I conceptualize *how* teachers work with one another in terms of the literature on social capital, describing the extent of each teacher's network of relationships with key adults, the nature of the implicit norms that govern adult-adult interactions, the depth of these interactions, and the degree of expertise that each teacher is able to draw upon from others within her social network. In addition to research on social capital, similar issues have been investigated and described in reference to the impact of teachers' working conditions on teacher and student outcomes. Within the MET dataset, I use the Teacher Working Conditions Survey (Hirsch & Emerick, 2007) to measure the dimensions of teacher social capital just mentioned. I approximate teaching practice by using student perception survey items, focusing especially on the Tripod survey items (Ferguson, 2008), and ratings on the Framework For Teaching observation instrument (Danielson, 1996; 2007).

In addition to considering estimates of teaching practice from different instruments, this dissertation examines the substantive relationships between teacher social capital, teaching practice, and student outcomes using different modeling frameworks that account for measurement, sampling and clustering error differently. Considering how different measures and models may lead us to different inferences about the relationships among teacher social capital, teaching practice and student outcomes is an important issue both substantively and methodologically.

Why Teaching Practice Matters

If we look broadly across the field of education today, we would be hard-pressed to identify an issue that is more prominent than teaching quality. The consensus appears clear: Teachers matter. Teaching matters. Today, perhaps more than ever, classroom instruction is front and center among education issues (for a review see e.g., Correnti & Martinez, 2012). Teaching quality varies a great deal (Hanushek & Rivkin, 2006; Sanders, Wright, & Horn, 1997) and has large impacts on student test outcomes (Nye, Konstantopolous & Hedges, 2004); economists point to dramatic impacts on long-term student outcomes (Hanushek 2011; Chetty, Friedman, & Rockoff, 2011; 2013); some researchers argue that the quality of individual teachers has a larger impact than prominently-touted reforms such as reducing class-size or raising per-pupil spending (e.g., Darling-Hammond, 2000; Rivkin, Hanushek, & Kain, 2005); methodologists remind us that understanding classroom instruction is crucial for understanding how any and all other education reforms work or do not work (Raudenbush & Sadoff, 2008); and educators suggest that clear measurement of quality teaching may help other teachers improve their practice (Elmore, 1996).

As attention has swung toward the perceived need to improve the quality of our nation's teachers, research has also called attention to weaknesses in teacher quality. Part of the problem with education, it is argued, is that our teachers' human capital is weak: for instance, those who enter and stay in teaching have lower SAT/ACT test scores and graduate from less prestigious universities than the average college graduate (Murnane et al, 2009). The key policy question is: How can we improve the quality of the teachers in our classrooms?

The most prominent policies have focused on increased teacher evaluation and accountability. Federal attempts to improve the caliber of teachers have included the provisions in No Child Left Behind that attempted to mandate a “highly-qualified teacher” in every classroom (U.S. Department of Education, 2002), the Race to the Top emphasis on teacher evaluation noted above, and discussion of increased evaluation for university teacher preparation programs (U.S. Department of Education, 2014). The federal government has also promoted alternative routes to teacher certification such as Teach For America. Partly in response to Race to the Top, new and more rigorous evaluation systems have been implemented in many states and local districts.

Policymakers at the state and local levels have also attempted to promote quality teaching with certification requirements, licensure exams, incentives to encourage advanced degrees and longevity, alternative certification routes, and merit pay. However, most of these factors have not proven to be consistently related to improved student outcomes (see Goldhaber, 2002; Darling-Hammond et al, 2005; Glazerman et al, 2006; Springer et al, 2011, for reviews of the evidence). Among all these efforts, three approaches appear to hold the potential of recruiting people more likely to become effective teachers (and/or to train those people better or differently). First, a large number of studies have investigated the effectiveness of Teach for America teachers in comparison to their traditionally-certified peers, and, on the whole these studies appear to point toward small positive effects for TFA teachers in math (Clark, Chiang, Silva, McConnell, Sonnenfeld, Erbe, & Puma, 2013; Decker, Mayer, & Glazerman, 2004; Xu, Hannaway, &

Taylor, 2007). A second approach is career ladder programs. In the Los Angeles Unified School District, a career ladder program that recruited teacher aides to become teachers produced teachers who were slightly more effective than their peers *and* were more likely to continue teaching (Center for Education Policy Research, 2012). Third, apprenticeship/residency programs have proven effective in other fields and appear to hold promise as a means of providing new teachers with more authentic training experiences. However, each of these promising approaches also faces obstacles. To date, Teach For America faces difficulties in convincing new teachers to remain in the classroom for the long haul (only 16 percent of Teach For America teachers in Los Angeles remained in the classroom beyond three years: Center for Education Policy Research, 2012; in New York City, only 15 percent remained in public school classrooms beyond four years: Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006). Career ladder and apprenticeship/residency programs are currently mostly small and relatively few in number; they may be difficult to scale up to the size necessary to meet the full hiring needs of our teacher workforce because they ask for a greater commitment from both the prospective teachers in terms of time and the district or university in terms of funding.

While the spotlight on teach*ers* has remained steadfast and bright for some years, somewhat less attention has been paid to teach*ing*, to the quality of the practices teachers employ and the interactions in which they engage every day (Hiebert, 2013). Those who have studied the practices that make for quality teaching have built on a variety of learning theories. For instance, sociocultural theory and constructivism have led to a focus on building relationships with students, focusing instruction within each student's zone of proximal development, providing scaffolding as students learn, and engaging students in problem-based and inquiry learning (see e.g., Carpenter, Fennema, & Franke, 1996; Piaget, 1952; Pianta, LaParo, & Hamre, 2008;

Vygotsky, 1980). Theories of formative assessment and feedback have led to pedagogy focused on communicating clear learning goals and criteria for success, aligning assessments to learning goals, providing timely and meaningful feedback, and adjusting instruction based on assessments (see e.g., Black & Wiliam, 1998; Hattie & Timperley, 2007). The research base in these areas, and other areas of pedagogical theory, is extensive and deep, but policy efforts have paid far more attention to the research on teach*ers* than to these suggested prescriptions for teach*ing.* For instance, No Child Left Behind (US Department of Education, 2002) focused prominently on teachers by mandating that every teacher be "highly qualified" and authorizing the Troops to Teachers program, but efforts to change teach*ing* practice were more limited, coming in small grants to programs like the National Writing Project and professional development efforts like the Teaching American History grant. Even a grant intended to improve the use of technology illustrates this point because it is referred to not as an effort to improve teach*ing*, but as the "Teach*er* Quality Enhancement Grant." Similarly, Race to the Top (US Department of Education, 2009) focused on specific criteria for how states must evaluate teach*ers*, but did not spell out particular teach*ing* practices.

Meanwhile, some research has focused on understanding the role of teachers as mediating influences on student outcomes rather than as initiators or input factors (Raudenbush & Sadoff, 2008). This perspective views instructional practice as mediating the impact of a wide range of policies, from professional development to curricular innovations. Typical education policies deliver curriculum, strategies, schedules, and more from state, district, and school administrators down to teachers; therefore, if we are to understand what makes for quality teaching practice, it makes sense to carefully investigate the contexts within which teaching takes place. In short, looking at teaching as a mediating factor means looking not only at the impacts *of* *teaching* on students' outcomes but at the impacts of other factors *on teaching* practice. Research evidence provides support for this conceptualization, suggesting that teachers' and/or schools' effectiveness can vary based on the physical facilities in which teaching occurs (Schneider, 2003; Uline & Tschannen-Moran, 2008); the available resources in terms of per-pupil expenditures, teacher salaries, teacher-pupil ratio, and school size (Greenwald, Hedges, & Laine, 1996); colleagues' performance (Little, 1990; Jackson and Bruegmann, 2009), and the school climate (Haynes, Emmons, & Ben-Avie, 1997; Uline & Tschannen-Moran, 2008; MacNeil, Prater, & Busch, 2009). Futernick (2010) argues that teachers' ineffectiveness is driven in large part by factors such as inadequate professional development and training, high rates of turnover in school leadership, and being assigned to teach courses out of their field of expertise. For instance, in 2003-04, 32 percent of core academic subjects in New York City high schools were taught by out-of-field teachers and Ingersoll (2003) found that 38 percent of secondary teachers teaching math courses did not have a major or minor in math or a related discpline, so Futernick's argument implies that those teach*ers* would likely conduct more effective teach*ing* if they were assigned to teach classes within their field of expertise. Taken together, these findings and arguments point to the conclusion that the contexts within which teachers work have significant impacts on the quality of their work.

Why Teacher Social Capital Matters

If the quality of teaching is partly determined by the surrounding context, then understanding and improving the quality of the teach*ers*, the main human capital available within schools, is only part of the puzzle; we must also attend to the factors that will affect the teach*ing*. A crucial aspect of the context that affects teachers is the social capital available to teachers – the relationships among teachers and the ways in which those relationships do or do not function to benefit teaching and learning. In detail, teacher social capital includes the extent of a teacher's personal networks, from which she can, for instance, learn new teaching strategies (Leana & Pil, 2006); the presence or absence of positive norms of trust and caring that shape teachers' relationships, leading, for instance, to increased or decreased likelihood of accepting constructive criticism and to efficient or inefficient group decision-making (Bryk and Schneider, 1996); the level of expertise that is available to a teacher through her networks, providing her with access to relevant knowledge and resources (Jackson & Bruegmann, 2009; Coburn et al, 2012; Taylor & Tyler, 2012); and the depth of the interactions that take place within collaboration opportunities, bringing teachers to a deeper understanding of student work, or simply enabling them to schedule intervention groups (Horn & Little, 2010; Coburn, Mata, & Choi, 2013).

Some have suggested that a powerful way to improve human capital in schools is to focus first on strengthening the social capital within those institutions (Hargreaves & Fullan 2013; Leana, 2011). Logic and research suggest that improved social capital will lead to improved teacher learning (Frank, Zhao, and Borman, 2004; Kraft & Papay, 2014) and motivation (Kraft, Papay, Charner-Laird, Johnson, & Reinhorn, 2012) as well as improved teacher retention (Ladd, 2009; 2011), all of which should contribute to improved teaching. This dissertation will provide further insight into the strength of the relationships among teacher social capital, teaching practices, and student achievement.

Some research attention has focused on teachers' social capital and has found that effective schools are defined by a trusting environment in which teachers consistently collaborate and rely on one another for instructional advice (Bryk & Schneider, 2003). To illustrate the importance of social capital, think of the age-old critique of educational reform: reform makes no difference because each teacher “just closes her door and teaches the way she *knows* works best” (i.e., people keep doing what they’re accustomed to doing). This aphorism points to the problems caused by widespread lack of social capital: when teachers don’t trust one another or collaborate, they can’t learn from one another, they are likely to resist change of any sort, and school improvement becomes less likely. Researchers studying effective schools in Chicago have consistently identified high levels of trust in the most successful schools (Bryk and Schneider, 2003). Leana & Pil, (2006) found that when teachers frequently asked one another for instructional advice and indicated a feeling of mutual trust, student performance in math was higher, and Frank, Zhao, and Borman (2004) saw that teachers with high levels of access and response to social capital were more likely to adopt technological innovations.

Alongside the research on social capital within schools, a parallel thread has defined many of these same aspects in terms of positive working conditions. Specifically, the quality of teachers’ working conditions is defined by the quality of the school leadership and the extent to which teachers are trusted with decision-making authority, supported in efforts to improve student learning, and provided with opportunities to improve themselves (Ladd, 2009). The evidence suggests that poor working conditions make teachers more likely to leave and that these poor conditions are more common in schools with large populations of disadvantaged students (Ladd, 2009; 2011). Johnson, Kraft, and Papay (2012) found that “good” working conditions related positively to teacher satisfaction and student achievement, and Kraft and Papay (2014) found that good working conditions were related to improved teacher growth measured by how teachers’ value-added effects changed over time. The definitions of “good” working conditions and “greater” social capital overlap substantially. For instance, the Teacher Working Conditions Survey measures a dimension termed “time,” which Ladd (2009) aligns with the category described by Johnson (2006) as a “manageable work load.” This dimension includes a measure of whether adequate time is provided for teacher collaboration, and thus is clearly tapping into some aspects of teacher social capital. Similarly, what is called “high quality professional development” includes measures of the extent to which teachers are working together and learning from one another (Ladd, 2009). The Measures of Effective Teaching project also looked into some aspects of this question and found that characteristics of positive working conditions such as manageable time demands and quality of professional development were positively related to improved teacher practice in the domains of academic press and academic support (Ferguson and Hirsch, 2014).

Why Measures Matter

To better understand how and in what ways teacher social capital impacts teaching practice and student achievement, we need to measure the constructs thoroughly and accurately. Both teaching practice and teacher social capital can be defined and operationalized in many different ways and in different contexts, which poses a number of complex measurement problems. First among these, the number and nature of dimensions of teaching practice is far from a settled question. Conceptually speaking, the job of teaching clearly comprises a number of separate roles. Teachers are expected to be caring yet have firm control of the classroom, to provide students cognitive challenge along with the support they need to make steady progress. Research literature, teaching standards, and observation frameworks identify a range of distinguishable dimensions of teacher practice, from one general factor of teaching practice (Greenwald & Gilmore, 1997; Peterson, Wahlquist, & Bone, 2000) up to as many as nine distinguishable dimensions (Follman, 1995; Kyriakides, 2005; Martinez, Borko, & Stecher, 2012). When and if consensus can be reached on the number and nature of dimensions of teaching practice, considerable debate remains concerning the relative importance of one dimension versus another. In particular, scholars have debated the importance of a learning climate that focuses on cognitive challenge, or "academic press" (Phillips, 1997) versus one that centers on social support (Anson et al, 1991; Noddings, 1988). Others have argued that these different dimensions are not mutually exclusive, and may even be mutually reinforcing (Lee & Smith, 1999). These interrelationships among dimensions further complicate the measurement of teaching practice. Second, the available measures of teaching practice have long been critiqued on the basis of reliability, validity, and the extent to which separate measures are measuring the same construct of teaching (Kane & Staiger, 2012; Rothstein & Mathis, 2013; Thomas, 1929). This dissertation addresses these measurement challenges through careful conceptual and empirical analysis of the measures employed, and through a comparison of how inferences shift when observation ratings rather than student survey ratings are used to measure teaching practice.

As for social capital, it also consists of a combination of interrelated dimensions. Originally, social capital was conceptualized as social obligations / connections (Bourdieu, 1986) or as a resource for action that takes the form of obligations and expectations, information channels, and social norms (Coleman, 1988). In school contexts, teacher social capital has been viewed as taking the form of schoolwide norms of trust, respect, and caring (Bryk & Schneider, 1996), the role that teachers' working conditions play in developing a stable and effective teaching staff (Ladd, 2009; Kraft & Papay, 2014), the impact of collaborative structures on teaching quality (Hargreaves & Fullan, 2013), the effects of social networks on the spread of innovations (Frank, Zhao, & Borman, 2004), and the importance of teachers having regular interactions with colleagues that delve below the surface and involve teachers in analysis of planning, instruction, and assessment (Horn & Little, 2010). This dissertation draws from all of these strands of research to develop a conceptual model of the dimensions of teacher social capital and operationalizes this model using items from the Teacher Working Conditions Survey, refining the model based on empirical analysis.

The ultimate goal of teachers working together and improving their practices is student learning. Student learning gains are estimated in this dissertation through state standardized mathematics test outcomes, with prior student achievement on state math tests as a covariate. One of the strengths of the MET dataset that is used here is that it includes an alternative measure of student math achievement, the Balanced Assessment of Mathematics (BAM). Therefore, future studies can compare how results found here using state test outcomes may differ from results that might be obtained using the BAM, and such comparisons might provide insight into the extent to which test outcomes can be viewed as a suitable proxy for student learning.

Close attention to the measurement of teaching practice, teacher social capital, and student achievement will provide a clearer picture of each of these constructs and enable us to develop a better understanding of the relationships among them.

Why Models Matter

After the constructs of teaching practice and teacher social capital are adequately defined, their importance understood, and their measurement determined, one question remains: how will the relationships among these constructs be estimated? Much available research focuses on the teacher as the exogenous causal variable of the relevant student outcome, thus ignoring the extent to which teaching is itself caused by and constrained by the surrounding context. From the other side, much of the research on teacher social capital estimates the effects of teacher social capital on student outcomes without illuminating the intermediate mechanism. If teachers are collaborating and learning from one another and these collaborations are benefitting students, it makes sense to hypothesize that the benefits to students are being delivered through improved classroom instruction. Figure 1.1 shows a conceptual framework depicting these relationships: teacher social capital influences teaching practices, which in turn impact student learning.

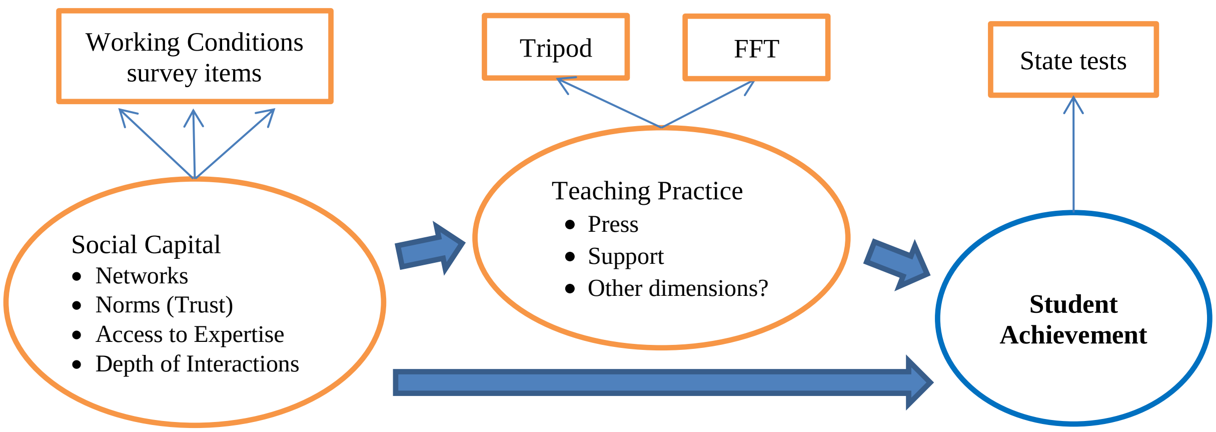

Figure 1.1. Conceptual Framework: Latent Constructs in ovals. Observed indicators in rectangles. Thick arrows indicate predicted relationships between latent traits. Thin arrows show relationships between underlying latent constructs and observed indicators by which they are measured.

This conceptual framework can be estimated through a number of different mathematical models. A multilevel structural equation model (MSEM) is one candidate that has not been applied to this area of research but that promises to properly account for the modeling and data complexities (Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthén, 2008; Preacher, Zyphur, & Zhang, 2010). MSEM can address the complications of 1) accounting for hierarchically structured (*multilevel*, *nested*, or *clustered*) data in order to correctly estimate standard errors and distinguish between what constructs mean at the individual versus the group level; 2) measuring latent constructs, assessing their dimensionality, and estimating the relationships among them while accounting for measurement error; and 3) estimating a mediation model within multilevel data.

First, the MSEM framework accounts for the clustered data structure that is often seen in educational data, in which students are clustered, or nested, within classrooms, and classrooms are nested within schools. Teaching practice is inherently a classroom-level construct, and thus the classroom level should be explicitly reflected in the statistical models (Morin, Marsh, Nagengast, & Salas, 2014). If, on the other hand, a multilevel model is not used, researchers can either analyze classroom effects at the student level (i.e., apply a single level model to the nested data), or aggregate student responses to the classroom level. Analyzing classroom effects at the student level confounds the effects on students and classrooms, and implicitly assumes both effects are the same (Morin & Marsh, et al, 2014). This can lead to biased estimates (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011) and to identifying the wrong number of factors or associating items with the wrong factors (Schweig, 2014). Alternatively, teaching practice may be represented by aggregating student level responses to the classroom level, and then estimating a single-level model relating these aggregates to other classroom-level variables. Using aggregated scores does not control for the variability in the individual responses (Morin & Marsh, et al, 2014), assumes that the aggregates represent the same constructs as the individual responses, and assumes that each construct has the same dimensions and the same reliability across all levels of clustering (Schweig, 2014). If there is substantial variability among individuals, or if these assumptions do not hold, then the parameter estimates will be biased.

Ignoring the variability among individuals may cause the relationships among the aggregated variables to appear stronger than they actually are, resulting in a greater risk of falsely finding significant relationships where none exist (Kaplan & Elliot, 1997b).

Second, teacher social capital, teaching quality, and student achievement are inherently latent constructs because they cannot be directly observed. But these constructs are often represented in models that implicitly assume observed variables are measured without error (e.g., Hirsch & Emerick, 2007; Loeb, Darling-Hammond, & Luczak, 2005). These are also known as *manifest* models because the variables are taken as given (as *observed*), whereas in a latent variable framework, observed variables are conceived of as indicators of unobservable (*latent*) factors. Employing manifest variables with unknown reliability means that estimates of the relationships among these constructs are attenuated by error, but the extent of that attenuation is unknowable. Ignoring measurement error can ultimately lead to erroneous conclusions on critical questions about the relationships among constructs (Lüdtke et al, 2011). For instance, Lüdtke and colleagues provide an example of how in a student survey, measurement error in student ratings led to an underestimate of the classroom-level effect on achievement, in comparison to a latent model that accounted for the measurement error in the survey ratings.

Finally, the MSEM framework, as demonstrated by Preacher, Zyphur, and Zhang (2010), can serve as a general, unified modeling paradigm for investigating hypotheses involving mediation within hierarchically structured data. Mediation hypotheses in certain types of nested data structures have been examined by others through multilevel modeling and through two-stage analyses that combine ordinary least squares (OLS) and multilevel models (MLM), but MSEM subsumes these methods as special cases and avoids several important limitations of mediation analysis with MLM. Specifically, the MSEM framework automatically separates the between and within effects and thus allows for unbiased estimates of the between group effects (Preacher, Zyphur, & Zhang, 2010; Schweig, 2014).

Notwithstanding the complications detailed above, much recent research on teaching quality and teacher social capital has employed models that do not account for clustered data structures, measurement error, or mediation. To build a deeper understanding of the relationships among these constructs, and to gain empirical evidence with regard to the theoretical advantages of the MSEM modeling approach, I compare the results from this MSEM model against 1) a single level latent model (i.e., an SEM model that is not multilevel, and 2) a manifest multilevel SEM model (i.e., a model that uses composite variables created by aggregating the items to their respective factors). In addition, I compare the results from models using different measures of teaching practice – one using the Tripod student perception survey and one using the Framework For Teaching observation rating – to gain insight into the extent to which our inferences about these critical relationships change depending on the measures used to estimate the constructs. Further, because the MET dataset used in this dissertation has itself been the source of a number of influential studies on teaching practice, the comparisons among models and measures in this dissertation may hold particular appeal for those interested in the implications of that prior work using these data.

The Context: Middle School Mathematics Classrooms

One final note regarding the application of the findings in this dissertation: these analyses focus on middle school math classrooms, a focus which has important implications. This focus is based on research showing larger teacher effects and greater variation in teachers' effectiveness in math than in reading (Hanushek, Kain, & Rivkin, 1998; Harris & Sass, 2011; Kraft & Papay, 2014). Also, prior research involving this MET dataset found that teacher effect estimates in math were larger and more precisely estimated than the effects in ELA classrooms (Kane, McCaffrey, Miller, and Staiger, 2013) and math value-added indicators were more reliable than ELA indicators because the stable component variability was larger (Mihaly, McCaffrey, Staiger, & Lockwood, 2013). The focus is on grades 6 – 8 because older students may be better able than elementary students to respond reliably and rate their teachers on the criteria most relevant to improving student learning (Peterson et al, 2000). In addition, prior MET research indicated that middle school indicators derived from the Tripod and FFT, among other measures, were more reliable than elementary school indicators and there was much greater year-to-year stability in middle school classroom estimates (Mihaly et al, 2013).

Middle school math classrooms might also be unique in ways that merit further attention in regard to teacher social capital and teaching practice. Middle school is the time when the climate of schooling tends to shift, when parental involvement lessens, and also the time of departmental specialization among teachers, when collaboration might be critical. Teacher social capital might matter more at this crucial age in child development, when children are vulnerable to peer influences and perhaps in need of a supportive, protective climate/environment. Math may be the subject in which it is easier to gauge the impact of teachers and teaching because ELA performance is likely impacted by History classes, electives, ELD instruction, even Science, as well as by English teachers. These are speculations, but they appear to signal an area for further study. At the same time, this focus requires tradeoffs and cautionary notes.

Math departments and math instruction in middle schools may vary in structure and substance from one school to the next, and are likely to differ markedly in comparison to elementary or high school math departments and math instruction. This dissertation does not examine structural differences among schools, but it may be important to keep the possibility of these differences in mind when interpreting the results. Also, while research has been consistent in finding larger effects and greater variation in math than in ELA outcomes, the reasons for these findings are not as clear or universally accepted. Perhaps the basis for these findings is that reading is an activity that is taught and takes place in the home and outside of schools as well as within classrooms whereas math is primarily taught within schools, perhaps the findings stem from math tests being more reliable as instruments than ELA assessments (e.g., 9 + 7 is more reliably exchangeable with 8 + 7 for assessing basic numeracy than any two sentences can be for assessing basic literacy), perhaps from truly greater variation in the quality of math instruction as opposed to reading instruction, perhaps from some combination of these and other factors.

Students in middle school are experiencing profound physical, biological, and social changes as they transition from childhood into adolescence. Educating students during this period has proven to be a consistent challenge: NAEP and TIMSS assessment results suggest weaker performance for students in the middle years (Meyer, 2011; Yecke, 2005) and grades five through eight are the least popular teaching assignment (Heller, Calderon, & Medrich, 2003). Dissatisfaction with the results for students of this age has been so persistent and prevalent that within this generation nearly all U.S. school districts have transformed their systems from junior highs serving 7th and 8th grade to middle schools serving 6th – 8th grades in the hopes of better meeting the needs of these students.

One possibility is that the challenges of the middle school years may dampen any effects of teacher social capital or teaching practices on student learning gains. Conversely, one large study (Williams, Kirst, Haertel, Rosin, Perry, Webman, & Woodward, 2010) found that some middle schools were producing strong student performance. Those schools featured clear and measurable goals, future-oriented missions, rigorous curricula, and the communication of responsibility to students, parents, and all parties in the educational system. It is possible that schools with positive ratings for teacher social capital and/or teaching practices would align closely to the characteristics identified by Williams, Kirst and colleagues and thus produce markedly better student performance.

The unique challenges of middle school, and the distinctiveness of middle school math classrooms, may impact the estimated effects of teacher social capital or teaching practices on student learning gains in unique ways and therefore readers should be cautious about applying the findings from this dissertation to other age groups or contexts.

Outline of Chapters

Chapter II details the research on measuring the quality of teacher practice. Chapter III explores the construct of social capital and the important role it plays within schools. Chapter IV explains the model framework, chapter V details the sample and methods, chapter VI reports the results related to the measurement of teacher social capital and teaching practice (Research Questions 1 and 2). Chapter VII reports the results related to question 3 from the MSEM model and the following chapters examine how those relationships differ when different models or measures are used. Chapter VIII describes how results differ when estimated using a single-level that ignores the nesting of students within classrooms versus the multilevel structural equation model. Chapter IX compares the results from a model employing manifest factors and thus ignoring the measurement error in the survey items to those in the model that employs latent factors and thereby accounts for that measurement error. Chapter X compares results from the model which estimates teaching practice using items from the Tripod student perception survey to a model using the Framework For Teaching observation ratings. Chapter XI concludes by summarizing the main findings and discussing the importance and limitations of the work.

Chapter II: Measuring Teaching Practice

Education is simply the soul of a society as it passes from one generation to another.

- Gilbert K. Chesterton[1]

An improved education system would lead to a dramatically different future for the U.S., because educational outcomes strongly affect economic growth and the distribution of income.

- George P. Shultz and Eric A. Hanushek[2]

Opinions on the purpose of education vary greatly, and they have changed over time. Following the publication of the Coleman report in 1966 (Coleman et al, 1966), many sought solutions to social problems through efforts that were largely outside the control of schools, such as the War on Poverty and the efforts to desegregate schools and society. The Coleman report was widely interpreted to suggest that differences between schools have little impact on student achievement because only about 10 percent of the variance in student test scores was associated with differences between schools (Carver, 1975).

Today, opinions have shifted. Education is now credited with playing a critical role in economic growth *and* having broad impacts on social outcomes such as physical and mental health, crime, participation in democracy and public institutions, and quality of life for individuals and communities (OECD, 2009). Alongside that credit, schools are also increasingly

[1] http://www.yusrablog.com/quotes/education-is-the-key-to-success-famous-education-quotes/

[2] http://www.wsj.com/articles/SB10001424052702303513404577356422025164482

being charged with accomplishing goals that range far beyond literacy and numeracy, including the development of "21st century skills," such as creativity, critical thinking, collaboration, persistence, and technological, civic, and financial literacy (see, e.g., Iowa core, 2010; Partnership for 21st Century Skills, 2007), creative and social skills (Frey & Osborne, 2013), resilience (Ginsburg & Jablow, 2005), relationships (Benson, 2007), self-control (Hofer, Kuhnle, Kilian, & Fries, 2012), grit (Duckworth, Peterson, Mathews, & Kelly, 2007; Strayhorn, 2014), higher-order skills, problem solving, dispositions, collaboration, and social-emotional intelligence (Darling-Hammond, Wilhoit, & Pittenger, 2014), health, safety, engagement, and sustainability, (ASCD, 2007). As the opening quotes suggest, visions for our education system may range from the spiritual to the economic; yet no matter where one falls on that spectrum, ambitious goals now appear in order for our schools.

In recent years, a consensus has developed that focuses on teachers and teaching as the key to solving social inequalities, raising economic output, and the like (see e.g., Correnti & Martinez, 2012, Gibbs, 2014). These efforts are sometimes discussed in terms of improving the human capital in schools, and that can be traced to decades of research findings that teachers have strongly varied impacts on student learning (see, e.g., Hanushek, 1970; Rivkin, Hanushek, & Kain, 2005). Teacher effects, typically defined as the impact of the teacher on student outcomes or the proportion of variance in student outcomes attributable to the teacher or classroom level, appear to vary a great deal based on as-yet unmeasured attributes (see e.g., Nye, Konstantopoulos, & Hedges, 2004; Kane & Staiger, 2008).This research base has established that variation in the quality of teachers (or classrooms) is responsible for the largest in-school portion of variation in student learning (Haertel, 2013; Hanushek, 2011), though it is important to note that the variation in student learning outcomes that can be attributed to teachers or classrooms is only about 10 percent of the total amount of variation (Schochet & Chiang, 2010).

Another approach has been to define teacher quality in terms of observable qualifications such as measures of teacher preparation and certification (e.g., Darling-Hammond, 2000), and years of experience and scores on credentialing tests (Clotfelter, Ladd, and Vigdor, 2007). Darling-Hammond argues that measures of teacher quality such as these are more strongly related to student achievement than other kinds of educational investments such as reduced class size, overall spending on education, and teacher salaries (2000).

Teachers with more years of experience have, on average, been shown to produce larger gains in student achievement, with the bulk of those gains coming from the first few years of teaching (Clotfelder, Ladd, & Vigdor, 2007; Harris & Sass, 2011), though some researchers find this evidence base less compelling and conclude that essentially no gains occur after the first year or two of teaching (Hanushek & Rivkin, 2006). For other measurable characteristics of teachers, the results are less clear. Attainment of advanced degrees typically shows no association with student outcomes (Buddin & Zamorro, 2009; Clotfelder, Ladd, & Vigdor, 2007; Rivkin, Hanushek, & Kain, 2005) and teachers' scores on licensing tests also are unrelated to student achievement (Buddin & Zamorro, 2011). Credential or licensure status (Clotfelder, Ladd, & Vigdor, 2007) and teacher scores on achievement tests (Hanushek & Rivkin, 2005) appear to have a clear impact only in certain secondary subjects.

Conceptualizing Teaching

The broad agreement that appears to exist regarding the importance of teachers, does not extend to agreement on how to define quality teaching, or on how to identify quality teachers.

The consensus breaks down for at least three reasons. First, different stakeholders hold dramatically differing perspectives on the issue. Students and teachers, parents and policymakers, researchers and reporters – all have points of view on the issue; those points of view differ in part because they arise from different people with different skins in the game.

Second, the consensus collapses in the face of the uncertain reliability and validity of available data on what goes on in classrooms. As long ago as 1929, Dorothy Thomas attempted to move the study of classroom interactions between teachers and students toward a more scientific basis, critiquing the "selective, inconsistent" descriptive accounts of classroom behavior that at their worst formed "such an intermixture of fact and interpretation as to be utterly worthless from the scientific point of view," and, even at their best, were "dependent on the recorder" (p. 3, in Chávez, 1984, p. 238). Echoes of these inconsistent intermixtures can be heard in today's debates in which opponents stridently attack the reliability and validity of the other side's evidence, whether that evidence stems from qualitative studies or value-added analyses of teachers' effectiveness.

Third, different stakeholders are often discussing entirely different questions. Some ask: What makes for a good teacher? Answers range from advanced degrees to years of experience, to sharing the same racial or ethnic background with students. Others discuss: How do we produce good teaching? Suggestions for policy solutions range from new curricula to improved professional development to increased accountability. Still others ask: Who can and should measure the differences in teaching quality? Disputants argue for principal evaluations, teacher observers, value-added models, student surveys, or some combination of multiple measures. The problem with this host of similar but not identical questions is that the specific questions often remain implicit, so that opponents end up speaking past one another without even being aware that they are debating separate questions.

Some questions about teacher quality are starkly different, but there are also subtle distinctions that turn out to be critical. In this category, there is the distinction between teach*er* quality and teach*ing* quality, and the distinction between good teaching and successful teaching. Teacher quality is a characteristic of individuals, either observable attributes like years of experience and degrees attained, or latent skills/abilities; teach*ing* quality denotes a focus on the quality of classroom practices and teacher - student interactions. The distinction matters for a number of reasons that go far beyond semantics. First, in terms of the observable attributes of teachers, with the notable exception of the first few years of experience, research has shown little correlation between most teacher qualifications and students' learning (see, e.g. Hiebert, 2013). There is little reason to believe that better teacher characteristics will lead to better teaching or learning. Second, if we turn our attention to the underlying effectiveness of teachers, our current abilities to measure the effectiveness of any individual teacher are highly imprecise and unstable from year-to-year (Baker et al, 2010). This suggests that a focus on teacher quality is likely to lead to debates about which measures are most appropriate and competition rather than collaboration among teachers. On the other hand, a focus on teach*ing* quality aligns with a focus on improving what teachers do. This would appear to be an appropriate and potentially beneficial focus because a review of research suggests that most factors that have medium to large effects on student outcomes are classroom-level factors that directly require teacher action, such as direct instruction or feedback practices (Hattie, 1999, in Muijs, 2006). Further, because teach*ing* quality is an attribute of classrooms rather than people, a focus on teaching can incorporate the contextual effects produced by long-studied classroom processes such as the effects of peers on other students and the dynamic interplay between students and teacher (Bell, Gitomer, McCaffrey, Hamre, Pianta, and Qi, 2012).

The second clarification that is needed is to distinguish between successful teaching and good teaching. While it is common to follow Hanushek's (2002) definition: "Good teachers are ones who get large gains in student achievement for their classes; bad teachers are just the opposite," this definition is more appropriately labeled as successful rather than good. The distinction is between the task sense of teaching, which requires that the teacher *try* to effect student learning and the achievement sense, which requires that the student actually learn (Ryle, 1949, cited in Fenstermacher and Richardson, 2000). Good teaching is not the same as successful teaching. Successful teaching might accurately describe how Fagin taught the Artful Dodger to successfully pick pockets without being caught, or the way that many U.S. schools successfully teach students to stand quietly in line. Successful teaching might also accurately describe unethical means such as cheating, threats, bribes, or beating children into rapt attention. Thoughtful observers may be quite reluctant to label such instances as constituting good teaching. Therefore, successful teaching may not be the same as good teaching (Fenstermacher and Richardson, 2005). Good teaching involves what is being taught (appropriate, challenging, and moral content), and how it is being taught (caring, supportive, and efficient guidance). But, good teaching will not always prove successful. Learning, according to the argument of Fenstermacher and Richardson, is determined not only by good teaching practices, but by three contextual factors outside the direct control of the teacher: the effort of the students, the social support for learning, and the opportunity to learn. Following this argument, teaching quality is defined as a measure of the extent to which teaching is both good and successful. Therefore, although this chapter discusses the effects of teaching practice on student learning, it is incumbent upon us to acknowledge that we are talking about the effects of the context + the quality of the teacher's practices.

From research literature, at least three main fields of inquiry have sought to define what constitutes quality teaching.

First, a variety of learning and pedagogical theories have been used to investigate the practices that make for quality teaching. Among the most prominent of the broad learning theories are 1) the behaviorist/positivist model that suggests that effective teachers *transmit* knowledge and skills to students and better teachers are clearer and more efficient in their transmission. This model forms the basis for process-product research as well as for many teachers' classroom management strategies; 2) a cognitive science based model of teachers as thinking professionals, which suggests that quality teaching is a progression through five stages of "expertise," and portrays the best teachers as those who are most expert in the elements of teaching (Fenstermacher & Richardson, 2005); 3) constructivist teaching, which theorizes that people construct new knowledge by assimilating new information into their existing mental frameworks or accommodating their mental representations of the world to fit new experiences (Piaget, 1952). Constructivism sees the role of the teacher as helping students build upon prior knowledge and question their assumptions, and describes good teaching not in terms of efficiency or expertise, but in terms of challenging and questioning; 4) and sociocultural theory, which posits that learning occurs first between people and only afterwards within the learner, and emphasizes collaboration and apprenticeship (Vygotsky, 1980).

The literature in each of these fields is vast, and in this brief summary, I seek only to call attention to the fact that these sometimes-contradictory theories of cognitive development lie at the base of learning standards (e.g., Next Generation Science Standards, National Council of

Teachers of Mathematics' Principles and Standards for School Mathematics); teaching standards (e.g., California Standards for the Teaching Profession, the National Board for Professional Teaching Standards' National Board Standards); and instructional reform efforts that emphasize teaching practices such as building relationships with students, focusing instruction within each student's zone of proximal development, providing scaffolding as students learn, and engaging students in problem-based and inquiry learning. For instance, Cognitively Guided Instruction (Carpenter, Fennema, & Franke, 1996) draws on theories of cognitive development and research on the development of students' mathematical thinking to prescribe a specific approach to math instruction in which teachers introduce problems that elicit children's mathematical thinking and then analyze this thinking and use it as the basis for further instruction. In many cases, these reform efforts have also produced efforts to measure the extent to which teachers adhere to these specific reform principles or standards in the form of rubrics, observation protocols, and survey items that emphasize specific teaching practices. For instance, the Reformed Teaching Observation Protocol focuses on measuring the extent to which teachers adhere to Arizona's reform principles based on the 2000 NCTM standards (Piburn, Sawada, Turley, Falconer, Benford, Bloom, & Judson, 2000).

Second, theories of formative assessment and feedback have led to pedagogy focused on communicating clear learning goals and criteria for success, aligning assessments to learning goals, providing timely and meaningful feedback, and adjusting instruction based on assessments (see e.g., Black & Wiliam, 1998; Hattie & Timperley, 2007; Sadler, 1989). Broadly, formative assessment research conceptualizes a close link between high-quality teaching and high-quality, ongoing, day-by-day and minute-by-minute formative assessment. In other words, good teaching consists of teachers assessing their students' understanding and adjusting their learning goals and instruction based on those assessments, as well as providing feedback to students to help them adjust their own learning. Good teaching also entails creating structures for students to monitor their own progress (self-assessment and peer assessment) and to adjust their learning strategies in response to the feedback they receive about their progress. Once again, formative assessment-based theories of teaching and learning have generated efforts, in the form of rubrics, observation protocols, and surveys, to measure the extent to which teaching practice aligns with these principles of reform (see e.g., FAST SCASS, 2012).

Third, comparative education researchers have identified characteristics of high quality instruction, particularly in mathematics, describing high quality instruction as focusing explicitly on deep conceptual understanding and engaging students in productive struggle with challenging problems and content (Hiebert & Grouws, 2007). These researchers have also specifically characterized teaching in U.S. classrooms in general, and particularly in mathematics classrooms, as rarely if ever demonstrating these two key features. Specifically, the teaching in U.S. math classrooms, as compared to the teaching in Japan and Germany, has been described as almost uniformly weak in providing students with opportunities to struggle productively and in developing students' conceptual understanding (Hiebert & Grouws, 2007). These weaknesses, along with "particularly weak" attitudes towards mathematics (p. 106) and a gap in knowledge of school mathematical concepts are also shown in comparison to teachers in China (Ma 1999). Instead, U.S. instruction is described as persistently following a pattern of controlled recitation of low-level question-and-answer routines (Hoetker & Ahlbrand, 1969; Hiebert & Grouws, 2007). In addition, these comparative education scholars have consistently characterized U.S. mathematics instruction as essentially similar in practice from one classroom to the next. Yet, interestingly, teacher effects consistently show greater variation in mathematics achievement growth than in reading achievement (Hanushek, Kain, & Rivkin, 1998; Harris & Sass, 2011; Kraft & Papay, 2014). This apparent paradox may simply be a reflection of the greater reliability of our mathematics assessments (Phillips, M. personal communication, May 6, 2016), but it also may relate to the fact that much of the reading "work" done by children takes place (or does not!) outside of the classroom and out of the control of the teacher or the school (Harris & Sass, 2011). Regardless of the cause, it appears to paint a rather bleak picture: our math instruction is uniformly weak, and our ability to impact reading achievement is limited.

Each of these three broad areas of the literature on teaching practices provides another plank in the argument for the importance of measuring teaching practice carefully and for the particular importance of this dissertation. The breadth and depth of research in learning, development and pedagogy makes it clear there is a considerable amount yet to be learned about what teaching practices ought to look like; instructional reform efforts and formative assessment research suggest that particular pedagogical approaches hold substantial promise; and comparative education research suggests that our business-as-usual teaching practices leave a great deal of room for improvement. Therefore, it is critically important to measure carefully what actually *is* being done and to examine the relationships found between these teaching practices, aspects of the school context, and key student outcomes.

A fourth strand of research provides another reason. Some researchers have emphasized how teaching and teachers mediate the impact of countless other educational policies on student outcomes, and thus need to be measured accurately if the impact of other educational policies is to be known with any degree of certainty. Educational policies create the "intended instructional regime," the instruction students are intended to receive, but there is inevitable slippage between what the policy intends and what is in fact enacted (Raudenbush, 2008). Therefore, good measures of teaching are needed to accurately measure and meaningfully improve countless other educational policies such as curricula and professional development.

The above argument hopefully makes clear the importance of measuring instruction. Unfortunately, achieving this goal has proven elusive for a variety of reasons.

To start with, teaching is difficult work. It is emotionally taxing to be relied upon and confronted each day by dozens of children's unmet needs and emotional upheavals; it is intellectually challenging to prepare and deliver five or more lessons each day, assess the needs of many, many students and adjust learning goals and activities in response to those assessments; and it is physically demanding to walk and stand and lean and squat and talk and talk all day. The difficulty level implies that consistent performance may be unlikely and may partly explain why observational studies of teaching typically find a great deal of variance in ratings of quality from one occasion to the next.

In addition, the nature and purpose of teaching changes from day to day and lesson to lesson. These changes are dependent on purpose. As Hiebert & Grouws (2007) argue, good teaching is good teaching FOR a specific goal; effective teaching tactics for one goal may not work for another goal. This also makes it difficult and costly to measure teaching through observation because we are likely to need many occasions to get reliable estimates. Previous MET project research found that four observations were necessary to achieve reliabilities from 0.63 to 0.67 for four out of the five observation instruments they studied. The fifth instrument had a reliability of only 0.39 with four observations, and reliability for these instruments based on only one observation ranged from 0.30 – 0.37 for the first four instruments and was only 0.14 for the fifth instrument (Kane & Staiger, 2012).

Measuring teaching also requires measuring multiple roles, which teachers often engage in simultaneously and which overlap and depend on one another in inconsistent ways. This incredible complexity of teaching carries with it a host of complications when one attempts to evaluate the skill with which it is being done. First, arriving at a consistent operational model of teaching is difficult because the different dimensions of a teacher's job often rely heavily on one another. Second, the breadth of the job (teachers perform many distinct roles) makes it difficult for a single measurement instrument to adequately capture all important aspects. Even the best observation rubric by the most expert observer cannot tell us how the teacher interacts with parents or provides feedback on written work. Acknowledging this complexity, many advocate for evaluating teachers with multiple measures (Mihaly et al, 2013) and sometimes extend that advice to include multiple judges as well (United Teachers Los Angeles, 2012). Third, the depth of the job (a single teaching action might signify several distinct meanings) makes it difficult for even a trained observer to confidently draw the correct inferences from a particular teaching act. Even an expert observer may completely miss the reason a teacher calls on a particular student or elects to gather a small group for review of a concept. Because of the complexity of teachers' choices, some teacher advocates do not accept the validity of certain raters of teachers, in particular, students, principals, and outsiders to the particular grade level or subject area being evaluated (Greenwald, 1997; Murnane & Cohen, 1986). In addition, the choices that are most appropriate shift from day to day and moment to moment, such that what constitutes effective practice on Monday during a probability experiment will not be effective for Thursday's review of key vocabulary. As Murnane and Cohen describe it, there are "no clearly defined behaviors that consistently result in high performance," so it is difficult for an outside observer such as a principal to provide a consistent rationale for the rating of one teacher as better than another (p.

13). Perhaps it is not surprising that scholars and practitioners continue to have fundamental disagreements about what teaching is and should be.

To arrive at an agreed-upon conceptual model for teaching might seem to be the first step toward measuring teaching practice consistently. Yet, as alluded to above, behaviorism, cognitive science, constructivism, sociocultural learning theories and other approaches each arise from competing conceptions of even more fundamental constructs such as how people learn and what we ought to value as the outcomes of education (Fenstermacher and Richardson, 2005). Since different scholars "differ on what is critical to the doing of teaching" (Fenstermacher and Richardson, p. 34), it is not surprising that consistent measurement has also proven elusive.

Operationalizing Teaching

To operationalize teaching means taking a definition of teaching and making it measurable. It should be no surprise then, that scholars have operationalized the job quite differently, from the three main domains of the CLASS (Pianta, Hamre, Haynes, Mintz, & LaParo, 2006), to Danielson's four domains with a total of 22 components (Danielson, 2013) to the five subscales of the Reformed Teaching Observation Protocol (Piburn & Sawada, 2009) to the nine domains identified by Follman in a student survey (1995). Some researchers have found that an underlying global factor of teaching quality dominates (Greenwald & Gilmore, 1997; Peterson, Wahlquist, & Bone, 2000), but the bulk of the research supports the idea that teaching is multidimensional in concept and in its ability to be measured (Greenwald, 1997; Pianta & Hamre, 2009; B. Rowan, personal communication, October. 9, 2015). Identifying consistent and stable dimensions across studies has proven more difficult. Rowan contends that measures of teaching quality tend to consistently identify three dimensions of teaching: roughly described as classroom management, instructional support, and emotional support. However, Follman is joined by Kyriakides (2005) and Martinez, Borko, & Stecher (2012) in identifying as many as nine separate dimensions of quality teaching, and the nine dimensions identified are not the same across the three studies just mentioned. Some other dimensions of teaching quality that are frequently identified are holding high expectations, communicating and delivering instruction effectively, engaging students, planning instruction, assessing students, understanding subject matter and pedagogy, understanding child development, differentiating instruction for a diverse student population, and taking on professional responsibilities (often including those outside of the classroom). Dimensions such as these appear commonly in research, conceptual frameworks, teaching standards, and teacher observation rubrics, though they are described in somewhat different terms and sometimes overlap with one another. Interested readers may refer to the National Board for Professional Teaching Standards (2012), the Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards and Learning Progressions for Teachers 1.0 (2013), the California Standards for the Teaching Profession (2009), the Illinois Professional Teaching Standards (2013), Teach For America's Teaching As Leadership Framework (Farr, 2010), as well as the Framework For Teaching, the CLASS, and numerous other published articles and sources.

Part of the difficulty in measuring the different dimensions of teaching arises from the many ways that conceptually distinct dimensions of teaching nonetheless share critical aspects with one another. Delving a bit deeper into Pianta and Hamre and colleagues's conceptualization, we find ten dimensions within the three domains: *Classroom Organization, Instructional Supports,* and *Emotional Supports*. One dimension within *Classroom Organization* begins to illustrate the difficulty of defining the job of a teacher: *Instructional Learning Formats*, including indicators such as clarity and engaging approach, is seen by Pianta and Hamre as part of *Classroom Organization* but to my way of thinking, it overlaps considerably with the *Instructional Supports* offered by the teacher. To take another example, the CLASS Framework includes the indicator *Action to address problems* as a part of the *Emotional Supports* provided by a teacher, yet when a teacher takes such action, she may be simultaneously reinforcing the classroom norms that are part of *Behavior Management* and perhaps also providing the student with valuable feedback that serves as part of his *Instructional Supports*. This example makes clear how teaching involves different roles that overlap in complex and inconsistent ways, making it a considerable challenge to define the job in neat categories. These interactions and unresolved questions about how to group the CLASS dimensions could easily be raised for Danielson's *Framework for Teaching* or any other theoretical model that attempts to define teaching, illustrating how hard it is to pin down one accepted definition of teaching.

Thinking of this same issue from the perspective of classroom teachers, picture a teacher who struggles with the management aspects of the job. This teacher is likely to find himself unable to present any content clearly and too worn out to offer meaningful emotional support. One could easily imagine such a teacher receiving universally low marks on an observation rubric, just as we could imagine a teacher who develops strong management skills then finding it easier to present content clearly and offer emotional support, thus seeing concurrent rises in scores in all three domains. When such scenarios are common, ratings across different dimensions are said to suffer from a "halo effect." Scores in different domains will be so highly correlated that it will be impossible to distinguish among separate domains and teaching in all its complexity will become operationally defined as a unidimensional "good teaching" construct. At the logical end of this data reduction, Peterson, Wahlquist, and Bone (2000) find that a single global item, “This is a good teacher,” best represents students’ ratings of their teachers, with additional items not able to add substantially to our understanding of individual teachers’ quality.

Measuring Teaching

As the above discussion makes clear, the conceptual and operational challenges involved in defining teaching quality have substantial and obvious implications for our ability to measure teaching adequately. But even when those challenges have been dealt with, a third challenge remains: what instruments will most appropriately allow us to measure teaching? Multiple instruments have been developed using many different approaches, each approach with its pros and cons. I review several of the most prominent and promising measurement methods below, and then detail the two particular instruments that will be employed in this dissertation.

Classroom Observation Instruments

Classroom observation has a long history in efforts to measure teaching quality (see e.g., Medley & Mitzel, 1963) and classroom observation has long been the method of choice for gaining systematic insight into teaching practices (Stallings, 1977). Classroom observation instruments have recently enjoyed increased attention within the context of the renewed emphasis on teacher accountability policies. Indeed, all the states granted funding under the Race to the Top legislation in the United States included a new or redesigned classroom observation component for teacher evaluation (Reform Support Network, 2012). Within the context of these accountability policies, observation-based assessments of teachers in the classroom are viewed as critical for understanding the mechanisms linking classroom practices to improvements in student outcomes, for informing feedback to guide teacher improvement efforts, and for producing valid estimates of the quality of teaching (OECD, 2009; 2013; Kane and Staiger, 2012; Reform Support Network, 2012). Classroom observation also enjoys considerable face validity among educators, and is seen as the key source of information supporting diagnosis of strengths and areas for growth (Protheroe, 2002), formative evaluation and feedback (Danielson, 1996; Goe, Bell, & Little, 2008; Pianta et al., 2007), and teacher self-reflection (Richards, 1991).

Alongside the growing interest in teacher accountability, standards of effective teaching have become widespread in the past decade (e.g., Danielson, 1996; Campbell et. al., 2004) and modern observation instruments are typically based on such standards. The expectation is that clear and explicit teaching standards, and matching observation rubrics, will provide useful guidance to teachers and administrators for understanding and promoting high-quality instruction (Ingvarson & Rowe, 2008; Darling-Hammond, 2012).

Observation instruments vary in how they define teaching quality. The Framework For Teaching (FFT) defines four domains with a total of 22 components (Danielson, 2013), the Reformed Teaching Observation Protocol consists of five subscales (Piburn & Sawada, 2009), the Classroom Assessment Scoring System (CLASS) conceives of teaching as consisting of a teacher's *Classroom Organization*, *Emotional Supports for students, and Instructional Supports*, with ten dimensions within those three domains (Pianta et al, 2006; 2007). These differences are based in part on differences in the conceptions of even more fundamental constructs such as how people learn and what we ought to value as the outcomes of education (Fentsermacher & Richardson, 2005). The FFT comes from a constructivist background (Danielson, 1996; 2007; Kane & Staiger, 2012), and thus emphasizes intellectual engagement and teachers' questioning strategies. The CLASS protocol focuses on interactions between students and teachers and includes more emphasis on measuring teachers' emotional as well as instructional support for students (Kane & Staiger, 2012; Pianta et al, 2006). However, overlap can also be found. Several scholars have conceptually mapped the FFT to the CLASS and identified dimensions of teaching quality that both appear to be measuring (Kane & Staiger, 2012; Rowan, 2015) and prior MET research indicates a correlation of 0.88 between the overall ratings on the two instruments (Kane & Staiger, 2012).

Student Surveys

The ability of student surveys to elicit valid and reliable ratings of teachers and teaching depends on 1) the students themselves as raters, and 2) the soundness of the survey items. Lortie (1975) provides numerous examples demonstrating how teachers evaluate their success based upon what they see and hear from their students, suggesting that teachers implicitly view students as valid raters of teaching, and Follman characterizes students as "the criterial clientele," arguing that "no other individual or group has their breadth, depth, or length of experience with the teacher, and crucially, that teachers look to their students for indications of their teaching performance, rather than to outside sources" (Follman, 1995, p.58). In addition, students uniquely know how their peer pupils perceive teaching, provide information inexpensively, and student survey ratings provide 20 to 150 or more data points on an individual teacher, thus leading to high reliability (Peterson, Wahlquist, & Bone, 2000). Empirical evidence also points in favor of the reliability and validity of student ratings. Follman (1995) reviews research on elementary student ratings and concludes that elementary pupils as young as age four can rate reliably, and while they are vulnerable to biases such as leniency and halo effects, these vulnerabilities may not be much greater than those for older raters including high school students, college students and business executives. Greenwald (1997) reviews the literature on student ratings of teachers and finds a consensus emerging in the 1980s emphasizing the validity of student ratings. He finds that a majority of studies on the topic in the 1970s questioned the validity of student ratings, especially questioning whether teachers could get good ratings in return for giving good grades, but that research on student ratings after 1980 provided evidence of the construct validity of student ratings through correlations between ratings and test-based measures of achievement (correlations averaging 0.40), through path analysis studies showing associations among student ratings and measures of things such as student motivation and interest, and through multitrait-multimethod studies that found evidence for both convergent and discriminant validity of student ratings. Peterson et al (2000) also found evidence that students responded with reliability and validity and were able to discriminate between the teacher as a source of learning and as a person who shows respect and care. Kyriakides's 2001 review of the measurement of teaching in Cyprus cited evidence that student ratings of teachers are highly stable from one year to the next (correlations from 0.87 to 0.89) and that student ratings of the same instructors and courses were also highly correlated, from 0.70 to 0.87 (in Kyriakides, 2005). Kyriakides's 2005 study also showed evidence of two student surveys providing valid and reliable data, and found that these student ratings of teacher behavior were highly correlated with value-added measures of student cognitive and affective outcomes.

In considering the soundness of the survey items, Peterson et al argue that some items are more defensible than others. Desimone and Le Floch (2004) contend that poorly performing items can be refined through cognitive interviews. Worrell and Kuterbach (2001) argue that, for surveys to be reliable in recording teacher behavior, the behaviors rated must be low inference, operationalized as behaviors rather than as qualities that must be inferred. But Desimone and Le

Floch further claim that too often the items in use are not adequately piloted and refined and thus leave us with data from which we are unable to draw valid inferences.

An open question in the area of survey research is the conceptual model of teaching that can be rated through survey responses. Can surveys give us detailed information on the aspects of the job in which a particular teacher has excelled, or are these surveys measuring broader constructs such as overall effective teaching? Some research has found that using a single global item – “my teacher is a good teacher” – better represents the central views of students than a scale composite (Peterson et al, 2000) and that the factor structures for the entire battery of items are not consistent. Looking at this study suggests that different aspects of quality teaching tend to correlate highly, whether because student raters have difficulty differentiating these different aspects of the job (halo effects) or because teachers who are good at one aspect tend to also be good at others. On the other side of the spectrum, Kyriakides (2005) found consistent evidence for nine second-order factors and an underlying general dimension, suggesting that it may be possible to use student surveys to reliably and validly rate a complex model of teaching. Greenwald’s 1997 introduction to an issue focusing on student ratings of professors similarly found no consensus among the experts as to whether student ratings of teachers are conceptually and empirically dominated by a global factor, able to discriminate among lower order factors, or conceptually and empirically multidimensional.

Teacher Tests

In this category, at least three separate types of tests have been used. First, to directly tests teachers’ (or more typically, teacher candidates’) subject matter knowledge, skills, and pedagogical content knowledge, there are teacher tests such as the Praxis tests, the California

Subject Examinations for Teachers, and the Mathematics Knowledge for Teaching and other content knowledge surveys (Hill, Rowan, & Ball, 2005). Second, performance assessments such as the edTPA evaluate teachers (or teacher candidates) using rubric ratings of a portfolio of lesson plans, teaching videos, classroom assessments, and teachers' written reflections. These two types of tests are well-known and widely used for licensing teacher candidates. The third type of teacher test is a performance test which requires teachers to analyze video clips or vignettes of others' teaching. These teacher tests are designed to uncover how teachers are analyzing some of the difficult decisions they make on a daily basis. In what is termed the Classroom Video Analysis(CVA), teachers are asked to write about how the teacher and students in the video clip interact around the mathematical content, and the written responses are then rated on four dimensions related to content, student thinking, suggestions for improvement, and depth of interpretation (Kersting, Givvin, Thompson, Santagata, & Stigler, (2012). Kersting and colleagues consider the CVA to be a direct measure of "usable" teacher knowledge. They report strong correlations with teachers' scores on the Mathematics Knowledge for Teaching test. In addition, they find that this measure of teacher knowledge predicts instructional quality on video observations, which in turn predicts student learning, and also directly predicts student learning. Based on this evidence, they argue that the CVA measures the knowledge that teachers can actually use in improved instruction.

Teacher Logs

Teacher logs can be thought of as survey instruments administered daily or even more often (Rowan, Jacob, & Correnti, 2009). They have been utilized to obtain multiple first-hand accounts of teachers' actions at a low cost. In general, teacher self-reports have been found to suffer from a weak relationship between observed and reported behaviors, lack of reliability, and greater bias due to social desirability (Muijs, 200). However, the damning evidence against teacher self-reports is based mainly on end-of-year teacher surveys; the use of carefully constructed daily teacher logs that focus on recording instances of specific actions have the potential to ameliorate these issues and have been shown to achieve reliability and validity at levels similar to other measures of teaching (Rowan & Correnti, 2009). One of the main challenges of asking teachers to report on their own teaching is teacher memory: teachers may be highly inaccurate in reporting actions conducted some time ago. Teacher logs address this challenge by asking teachers to report at the end of a day or lesson on the instruction they just undertook, thus reducing the demands on memory and increasing the chances that they can report accurately. A second problem that is addressed by teacher logs is generalizability. It is both costly and difficult to obtain a sample of teaching observations that is generalizable to all of the teaching that takes place over the year. Because logs require lower cost than classroom observations (or portfolio or artifact analysis), data can be gathered on much larger samples of days, and these data are thus more likely to generalize to adequately measure the teaching that takes place over the entire year (Rowan & Correnti, 2009). Teacher logs have been utilized less generally than other measures and fewer thoroughly tested measures are available.

Portfolios/Artifact Analysis

Collecting and analyzing classroom assignments can provide a broad sense of the cognitive demand that a teacher expects in students' work, and the quality of a teacher's questions and feedback. By assessing these and other aspects of an assignment using an established rubric, each classroom assignment can be rated and compared to assignments from other classrooms. A portfolio can be assembled for each teacher, consisting of a number of assignments and perhaps also including other artifacts of classroom practice such as lesson plans and teachers' reflections. These portfolios thus provide a measure of teachers' written interactions with students and allow researchers to directly measure the strength of teachers' assessment practices, a critical part of teaching. Instruments using a portfolio of classroom artifacts have been found to measure instruction with reliability similar to classroom observations (Martinez, Borko, Stecher, Luskin, & Kloser, 2012). Further, these instruments correlate above .50 with observation ratings and at the same time may help to shed light on aspects of instruction not readily measurable when using observations. For instance, portfolios demonstrated substantially higher reliability than observation ratings in measuring the extent to which science teachers expected students to communicate using precise scientific thinking and the language of science (Martinez, Borko, Stecher, Luskin, & Kloser, 2012). Therefore, artifact-based portfolios also hold promise as tools for professional development.

Value-Added Methods (VAM)

Notwithstanding the strengths of each of the methods outlined above, currently the most prominent method of measuring teaching is through value-added estimation of the effect of each individual teacher on standardized student test outcomes. Value-added methods (VAM) is a term applied to a host of different modeling approaches that share one key aspect: they attempt to isolate the effects of the group level, with the groups typically being schools or classrooms/teachers. Value-added methods produce an estimate of the group deflection from the overall predicted value, and this group deflection is conceived of as the value-added "effect" of the school or teacher. It is important to note that teacher "effects" are not readily distinguishable from classroom “effects,” but this complication is often ignored in both research and policy approaches. In a multilevel modeling context, the value-added effects are the cluster-level residuals or deviations from the estimated mean effect. These effects can be estimated using models ranging from simple arithmetic calculation of gain scores and covariate-adjusted regression models to complex cross-classified and growth curve models. Models can also vary based on the outcome measure used, the covariates included, whether teacher effects are treated as fixed or random effects, and whether teacher effects are assumed to persist undiminished into future years or to decay over time (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). Some view these value-added estimates as *the* criterion of teaching quality: “Good teachers are ones who get large gains in student achievement for their classes; bad teachers are just the opposite” (Hanushek, 2002, p. 3, cited in Pianta & Hamre, 2009, p. 110; see also Mihaly, McCaffrey, Staiger, & Lockwood, 2013), but this view is not universally accepted. Instead, this is more appropriately labeled as successful teaching rather than good (Fenstermacher and Richardson, 2005). In addition, Kennedy (1999) views standardized test scores not as equivalent to student learning but rather as an approximate measure of complex student learning because of the widespread criticism that standardized tests measure an “overly narrow” range of intellectual work (p. 346). In this dissertation, student test outcomes and VAM estimates based upon them, are viewed as measures of student achievement. These student outcomes are distinct from teaching practice, though it is hypothesized that good teaching practice will predict student achievement.

Instruments for Measuring Teaching

Indicators of teaching quality from a variety of approaches correlate with one another at what are generally considered moderate levels. For instance, overall portfolio ratings of science teachers correlated 0.305 with teachers' overall self-assessments on a survey and five of nine portfolio dimensions of quality teaching practice showed a marginally significant relationship with student achievement (Martinez et al, 2012); the correlation between teachers' scores on a test asking them to respond to videos of practice and a more traditional test of pedagogical content knowledge was 0.406 and teachers' test scores on mathematical content, understanding of student thinking, and depth of interactions were indirectly related to student achievement through significant impacts on teaching quality as measured by videotaped observations, while teachers' test scores on suggestions for improvement had a significant direct impact on student achievement (Kersting et al, 2012); the disattenuated correlation between the Framework for Teaching (FFT) general observation instrument and the Mathematical Quality Instruction (MQI) observation instrument was 0.69 and the correlation of the overall Tripod student survey rating of each teacher with the underlying value-added measure of each teacher's effectiveness was 0.37 (Kane & Staiger, 2012). Correlations and relationships such as these provide some evidence of convergent validity - in some ways these different measures of teaching quality do indeed appear empirically to be measuring the same thing. The correlations are not, however, particularly strong, and clearly not so high as to create multicollinearity - these measures do not appear redundant. Instead, it appears that each of these indicators is measuring a somewhat separate and distinct aspect of this super-construct of teaching quality. It is as if each indicator is providing us with a handhold, but one on a leg, one on the trunk, a third on the tail. We are still blindly describing this elephantine construct.

Measuring teaching quality accurately is therefore clearly of vital importance to this study; and yet it also might be argued to be this study's Achilles' heel. In other words, teaching is clearly an incredibly complex endeavor, which researchers have sought for decades to measure, partly in vain. Currently, two of the leading approaches to measuring teaching practice are student surveys and observation protocols, and this study employs and compares prominent examples of both: Tripod student perception survey items (Ferguson, 2008; 2010; 2012; Tripod Project, 2016) and Framework For Teaching observation ratings (Danielson, 1996, 2007). The measures employed in this dissertation are, by necessity, less than ideal, but they are arguably among the best available and, also importantly, they are available within the same dataset and thus can be readily compared.

The Tripod Student Survey. The Tripod student survey was developed by Ronald Ferguson and colleagues through over a decade of work. In the MET database, Tripod items ask students to rate their teacher on 36 items. Theoretically, the Tripod was developed to measure seven dimensions of teacher practice, termed the Seven C's: Care, Control, Clarify, Challenge, Captivate, Confer and Consolidate. The conceptual framework of the Tripod is displayed in Figure 2.1. Each of these theoretical dimensions is measured using multiple items. For example, Care is measured by items like: *My teacher in this class makes me feel that s/he really cares about me*; Control is captured in items like: *Student behavior in this class is a problem*; Clarify is measured by items like: *My teacher explains difficult things clearly*; Challenge is approximated by items such as: *In this class, my teacher accepts nothing less than our full effort*; Captivate is measured by items such as: *My teacher makes lessons interesting*; Confer is measured by items such as: *My teacher gives us time to explain our ideas;* and Consolidate is measured by: *My* *teacher checks to make sure we understand what s/he is teaching us*, among other items A full list of Tripod items is available in Chapter V.

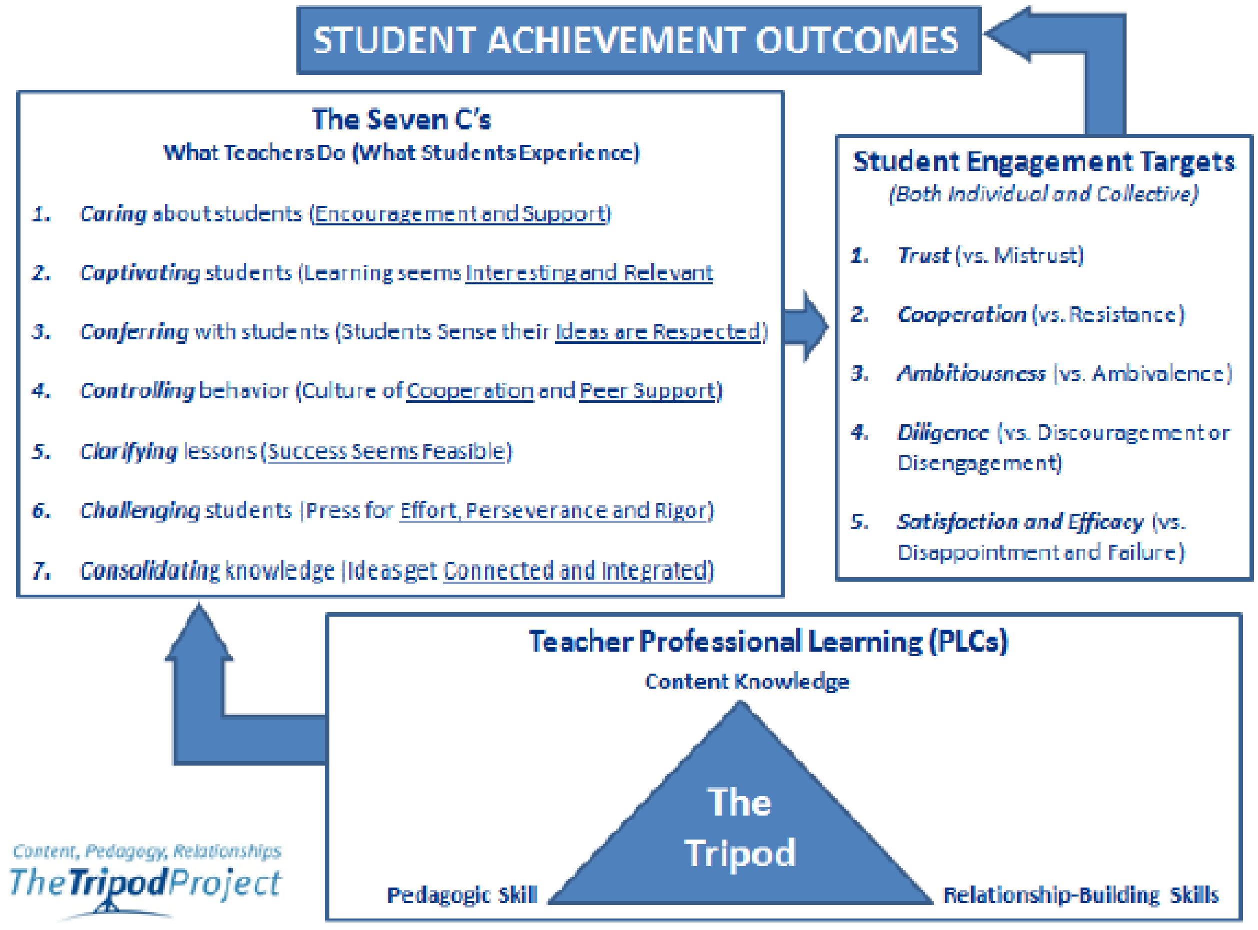

Figure 2.1. Tripod Student Survey Conceptual Framework

Tripod developers report "indices for the Seven C's constructed from MET survey data have proven highly reliable—in the range of 0.80 and above," and "remarkably stable during the school year" (Ferguson, 2010, p. 4). The analytic structure of the items was confirmed through factor analyses conducted by Mihaly et al (2013), with Cronbach's alpha measures of reliability in the range of 0.80 for each of the Seven C's. Further, a composite Tripod score equal to the average of all of the items was correlated 0.44 with the value-added measure derived by Mihaly and colleagues and 0.50 with the Framework For Teaching observation measure in middle school math classrooms. (Correlations similar in magnitude were found in elementary and ELA classrooms). These correlations among various measures of teaching quality, which were reported based on the correlations among stable components, suggest some evidence for the criterion validity of the Tripod. In other words, the correlations among these measures suggest they measure a common aspect of teaching quality.

Yet the Tripod also has weaknesses. First, in spite of the widespread and growing use of the Tripod since its creation in 2001, up until recently, relatively little research has been published on its psychometric properties (Camburn, 2012). In particular, the seven-dimension conceptualization of the construct of teaching quality has not been reproduced in empirical results. For instance, although the items are grouped conceptually as the Seven C's, prior MET work utilizes all items as a single scale, thereby creating a broad measure of teaching quality (Kane & Staiger, 2012; Mihaly et al, 2013). Raudenbush & Jean (2014) found the scores from the seven dimensions to be highly inter-correlated. In addition, Schweig (2014_EEPA) conducted multilevel confirmatory factor analysis on the Tripod survey and found evidence that the Tripod survey was measuring two rather than seven dimensions at the teacher/classroom level. In recent work, the survey author reframes the Tripod as measuring two dimensions- *Press* or "strictness", made up of Control and Challenge, and *Support*, composed of the other five dimensions (Ferguson & Danielson, 2014). A multilevel item factor analysis of this Ferguson & Danielson conceptualization demonstrated that control items and challenge items within the press dimension did not fit well with each other (Kuhfeld, under review). Kuhfeld's analysis resulted in a different two-factor structure for the Tripod, made up of one Control dimension and another dimension consisting of the other 6 Cs. Wallace, Kelcey, & Ruzek

(2016), meanwhile, arrived at a bifactor structure for the Tripod, consisting of one general factor representing teacher responsivity to students and one specific classroom management factor. Thus, the bulk of the evidence suggests that the Tripod is able to measure two dimensions of teaching practice rather than seven.

A second problem with the Tripod survey is that the item wording sometimes switches between individual referents such as “me” and “my” and group referents such as “us” and “our.” Conceptually, this could introduce problems for the respondents as well as for researchers interpreting survey results in terms of being able to confidently differentiate between whether students are reporting on a classroom level construct or on an individual level construct. For instance, responses to an item such as *My teacher respects my ideas and suggestions* might reflect true differences among students rather than simply measurement error if a teacher listens to boys’ ideas more than girls’, or high performing students’ more than low. Because the Tripod includes these types of items alongside items which reference the entire class, such as *Students speak up and share their ideas about class work*, it is possible that inferences about results will conflate different levels of measurement.

A third problem is that certain items on the Tripod might not be able to clearly distinguish between a more effective and less effective teacher. For instance, for the item *Students get to decide how activities are done in this class*, a strict teacher who “runs a tight ship” might score lower than a novice colleague who has little or no control over his classroom, even if the former teacher is more effective. This item is intended to measure an important aspect of teaching, known as autonomy support. In theory, autonomy support is a vital part of teachers helping students to develop into self-directed learners, but the particular item wording is potentially problematic. These potential problems with item wording appear to fall in the category of items that might perform poorly and therefore might be candidates for refinement through cognitive interviews and other procedures recommended to be conducted in piloting survey measures (Desimone and Le Floch, 2004). Unfortunately, I have been unable to locate any published sources with full accounts of the development and validation process used in creating the Tripod survey. Therefore, although the MET project reports that the Tripod survey was developed over a decade of work by an accomplished team of researchers, crucial validation evidence is still lacking for the Tripod.

The Framework For Teaching. Among the most prominent of all classroom observation instruments, the Framework For Teaching (FFT) was designed based upon research on constructivist learning theory, teacher effects, and job analysis studies, with reference to state licensure systems and with input from expert teachers and researchers (Danielson, 1996; Kimball, 2002). It has been adapted in as many as 200 school districts in the United States including New York, Los Angeles, and Cincinnati and in at least two states (Milanowski, 2011) and has also been adapted abroad —e.g., as "Marco para la Buena Enseñanza" in Chile (Ministry of Education, 2004). This prominence suggests a high degree of face validity for the FFT among educators. Kimball (2002) found evidence corroborating this suggestion: he studied teachers' reactions in three districts which had recently implemented a redesigned teacher evaluation system that relied on the FFT, and in each of the three districts, teachers indicated that the new FFT-based evaluation standards provided clearer expectations for performance than the prior standards, and Kimball concluded that, "by and large, teachers accepted the standards, procedures, and evaluation outcomes (p. 258). The Framework For Teaching (FFT) separates teaching into four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities, with each domain having specific performance "components" and then "elements" nested within them (see Table 2.1). Each element has a rubric or rating scale that describes the four performance levels in terms of observable teacher or student behavior, labeled unsatisfactory, beginning, proficient, and distinguished. The FFT is intended to be used across subject areas at the elementary and secondary levels.

Table 2.1
Framework For Teaching domains and components (adapted from Danielson, 1996)

Domain	Components
1. Planning and Preparation	1a: Demonstrating Knowledge of Content and Pedagogy 1b: Demonstrating Knowledge of Students 1c: Setting Instructional Outcomes 1d: Demonstrating Knowledge of Resources 1e: Designing Coherent Instruction 1f: Designing Student Assessments
2. The Classroom Environment	2a: Creating an Environment of Respect and Rapport 2b: Establishing a Culture for Learning 2c: Managing Classroom Procedures 2d: Managing Student Behavior 2e: Organizing Physical Space
3. Instruction	3a: Communicating with Students 3b: Using Questioning and Discussion Techniques 3c: Engaging Students in Learning 3d: Using Assessment in Instruction 3e: Demonstrating Flexibility and Responsiveness
4. Professional Responsibilities	4a: Reflecting on Teaching 4b: Maintaining Accurate Records 4c: Communicating with Families 4d: Participating in a Professional Community 4e: Growing and Developing Professionally 4f: Showing Professionalism

The FFT was the most reliable observation measure in the MET report. For one observation of one lesson, the variance in the overall rating attributable to the teacher was 0.37, and after aggregating scores across four occasions with four different observers, 0.67 of the variance reflected consistent differences among teachers, which the authors refer to as an "implied reliability" (Kane & Staiger, 2012). In Cincinnati, teacher ratings based on a rubric adapted from the FFT were significantly related to gains in student achievement: a two standard deviation increase in average teacher evaluation rating was associated with a student achievement gain of about one-sixth of a standard deviation in math and one-fifth in English Language Arts (Kane, Taylor, Tyler, & Wooten, 2010). In a study of one charter school in Los Angeles, even larger associations were found for reading: a one standard deviation increase in average teacher rating was associated with a statistically significant increase of about 0.16 standard deviations in student literacy test scores, though the associations for math and language arts were not significant (Gallagher, 2004). In addition to this evidence and the validation work undertaken by Danielson and colleagues in the development of the FFT, Milanowski (2011) reviewed published and unpublished studies, including those just referenced as well as others, on the relationship between FFT-based teacher evaluation ratings and value-added measures of teacher effectiveness, and concluded that evaluation systems based on the FFT can produce reliable ratings that correlate with value-added estimates of teachers' contributions to student achievement. The evidence summarized above comes from a broad range of studies and suggests we can have a degree of confidence in the validity of the FFT. However, the associations between FFT ratings and student achievement gains are not large, and in some contexts, such as the math and language arts teachers in Gallagher's (2004) study, no significant associations were found. Gallagher investigated these non-associations qualitatively and found that this lack of association could be attributed to a lack of alignment between the teacher rubrics, the instructional materials, and the state standards upon which the tests were based. On a somewhat similar note, Milanowski's (2011) review includes cautionary notes describing the evidence as "limited" and speculating that the different correlations found in different districts might be due to differences in rater training, frequency of observations, or the use of multiple raters. On the whole, then, the evidence suggests that FFT can provide valid ratings of teaching quality, but the extent to which these ratings associate with the typically defined criterion of student achievement are likely to vary based on a number of factors beyond the quality of the teaching.

Additional Factors that Impact Teaching and Learning

When measuring teaching practice, it is also essential to account for, or at least consider the impact of, a host of additional factors that impact teaching quality. Resources such as leaking roofs or high-speed internet connections make quality teaching more or less difficult (Roza, Hill, Sclafani, & Speakman, 2004; Uline & Tschannen-Moran, 2007). Curricula, textbooks, and assessments vary widely in quality and in turn impinge upon or support quality instruction (Hagarty & Pepin, 2002; McDonnell, 1995). Student background characteristics (Clotfelter, Ladd, & Vigdor, and the effects students have on their peers (Lavy, Silva, Weinhardt, 2012; Lefgren, 2004) also impact teaching practice; teaching is commonly viewed as back-and-forth interaction between teacher and students in which both affect one another (Skinner & Belmont, 1993), and it is of course easier to teach well-behaved and well-prepared students, and conversely harder to teach students suffering from emotional trauma, malnourished, or chronically tired or stressed (Rothstein, 2004). Parent and community support, or the lack thereof, also impacts the quality of teaching (Becker & Epstein, 1982; Peña, 2000). The quality of training received by teachers also affects their practice, and this includes their education from Kindergarten to college (Bacolod, 2007) and from pre-service through credentialing (Clotfelter, Ladd, & Vigdor, 2007) and induction, as well as ongoing professional development (Angrist & Lavy, 2001; Desminone, Porter, Garet, Yoon, & Birman, 2002; Jacob & Lefgren, 2004) and informal learning from peer teachers (Jackson & Bruegmann, 2009) coaches and specialists, support providers, and administrators. In addition, teachers learn by doing and by studying their craft on their own, and so experience itself is a factor that has been investigated in some depth to determine the extent to which and under what circumstances it relates to quality teaching practice (Clotfelter, Ladd, & Vigdor, 2007; Hanushek, Kain, O'Brien, & Rivkin, 2005). The circumstances are themselves often studied as factors that lead to greater teacher learning, higher likelihood of teacher retention, and/or better teaching practice, often termed "school climate" (Louis & Marks, 1998; Thapa, Cohen, Guffy, & Higgins-D'Alessandro, 2013; Uline & Tschannen-Moran, 2007) or "working conditions" (Ladd, 2009; Kraft & Papay, 2014). And another factor impacting teaching quality that has garnered substantial attention is the quality of the school leadership, in particular whether the school principal serves as an "instructional leader" (Dufour, 2002) and/or follows a model of "distributed leadership" (Spillane, 2005).

Many of the factors discussed, in particular leadership, school climate, working conditions, informal and professional development learning opportunities, and parent and community support can be conceptualized under the mantle of the social capital available to teachers. The next chapter examines social capital in detail and discusses how these factors impact the quality of teaching and learning.

Chapter III: The Impact and Importance of Teachers' Social Capital

As Chapter II detailed, recent education research and policy have focused attention on meeting broad and ambitious goals through improving teacher quality. Chapter II further argues in favor of conceptualizing this research area by focusing on teach*ing* practices, on what teachers are doing, rather than who they are. In chapter III, I extend the scope of my investigation to include some of the factors that impact teaching practice. With society's goals for education expanding to include social as well as academic outcomes, and with research and policy interest focusing on identifying effective teaching practices, I contend that the social contexts within schools ought to also be examined more closely for potentially meaningful influences on teaching practices. In this chapter, I examine an important and understudied aspect of the school context that may strongly influence what teachers do and how well they do it. This factor is the social capital available to teachers.

Social capital research has its origins within research and theory on capital in general and human capital in particular. The concept of capital comes from economic theory and essentially refers to any factor or product that can add value or enhance a person's productivity—a spear can function as capital for a hunter, a railroad can be capital for a merchant, and a factory can be capital for a business. Human capital is defined as the intrinsic productive capabilities of human beings, and similarly, investments in human capital are thought to lead to increased capacity to perform (Eide & Schowalter, 2010). Human capital is often conceptualized as the skills, knowledge, and qualifications acquired by humans through learning. Once acquired, the knowledge, skills, and qualifications can be invested in productive activities and produce a return on that investment (Schultz, 1961; Schuller, Bynner, & Feinstein, 2004).

Social capital is defined as the combined actual or potential resources associated with a social network of personal relationships and group memberships (Bourdieu, 1986). In Bourdieu's conceptualization, these social networks provide "credit" to their members. Potential resources might be thought of as the benefit of the doubt that gets a man hired because the boss knows his family or the assumption of competence that gets a Kennedy elected to political office; actual resources may take the form of insider knowledge gained through casual conversations that take place within an elite club. Social capital has also been defined as a property of the relational ties among individuals in a social system (Coleman, 1990) and refers to norms, networks, and relationships (Schuller, Bynner, & Feinstein, 2004). Social capital is typically operationalized in terms of teachers in the level of trust, the frequency of collaboration around instruction, and the quality of the collaborative environment (e.g. Leana & Pil, 2006). Coleman also discussed how social capital is promoted by norms of trustworthiness and by network closure, with closure defined as a social network in which all members have interactions with all others and thus are able to combine forces to impose constraints on any one member (Coleman, 1988).

Within schools, the focus on teacher quality lies directly within the broad focus on improving human capital that has been a staple of economists' policy thinking for decades (Schultz, 1961). But in spite of the promise and the long-standing appeal of this focus on human capital, improving human capital within schools has so far proven challenging, at least within the United States. Therefore, while we seek to improve the human capital in schools by selecting, training, and retaining better teachers and principals, some researchers have called for educational policy to shift emphasis to improving the social capital within schools (Fullan & Hargreaves, 2013).

Attention to social capital in education coincides with broader social prominence of this idea, especially since the publication of *Bowling Alone* (Putnam, 1995), which highlighted and quantified the decline in civic participation in institutions like the Elks, Kiwanis, Boy Scouts, voting booths, and famously, bowling leagues. Putnam alerted people to both the importance of social capital as a resource for collective action and the growing evidence that this important but largely invisible part of our social fabric was being dismantled. Putnam noted that the U.S. had long been exceptional in terms of the strength of its social capital, and sounded the alarm on the dire consequences of a future without that strength.

After Putnam, a great deal of research has suggested that social capital plays a critical role in the functioning of many organizations, including schools. While social capital is a resource that can be put to negative as well as positive ends (Putnam, 1995), social capital within schools is typically viewed as a resource that aids organizational functioning, group decision-making, and teacher and student learning. This chapter examines the evidence highlighting the importance of social capital in schools. I first explain my conceptualization of teacher social capital and provide an overview of the main methodological and conceptual approaches to the study of social capital. In addition, I summarize the evidence of the importance of each dimension of social capital, and the benefits of social capital as it relates to important school, teacher and student outcomes. I conclude exploring some of the main issues involved in measuring social capital.

Conceptualizing Social Capital

Social capital is conceptualized here based on the literature on social capital (e.g., Bourdieu, 1986; Coleman, 1988; Leana & Pil, 2006), social network analysis (e.g., Valente, 1995; Coburn, Russell, Kaufmann, & Stein, 2012; Horn & Little, 2010), trust in schools (e.g., Bryk & Schneider, 1996, 2003; Tschannen-Moran & Hoy, 2000), and teacher working conditions (e.g., Ladd, 2009; Kraft & Papay, 2014), with attention also to literature on distributed leadership and parental and community involvement, two related strands of literature which might be viewed as focusing on the social capital involving administrator-teacher and parent-teacher networks. Social capital is defined as a construct comprising four dimensions. The first dimension is the networks that describe the structure of the available social capital. The next three dimensions define the nature and content of those networks. The networks can be thought of as the machinery, the "what" of social capital, while the norms, available expertise, and depth of interactions comprise "how" that machinery operates. These dimensions are further described in Table 3.1 and defined as follows:

1) Networks (especially teacher-teacher, -administrator, and -parent). The strength of a teacher's network is defined by the frequency of interactions and the obligations and expectations shared by the people in relationship. In addition, closure of the social network promotes the growth of social capital. A school with high teacher turnover and/or high student transiency might be described as lacking in network closure, resulting in people being somewhat less inclined to collaborate because they can have no reasonable expectation of long-term reciprocity;
2) Norms of a school (especially in terms of trust, respect, and caring). The norms are defined by the levels of trust, respect, and caring that teachers experience with their colleagues and administration. This includes both the extent to which they feel trusted (and respected and valued), and the extent to which they trust other teachers and administrators to be competent and to take positive action on behalf of the students;

3) Expertise/resources/human capital available to teachers through their networks. Access to expertise is defined by the extent to which teachers have access to support from colleagues, administration, support personnel, and professional development opportunities; and

4) Depth of the interactions taking place within a teacher's network. The depth of interactions that take place within a network depends on the human capital that is available, the structure and frequency of interactions, and the norms that govern those interactions. In this study, depth of interactions is defined by the extent to which teachers collaborate on meaningful instructional activities such as assessment, lesson planning, and instructional alignment.

A comparison of Table 3.1 with Table 3.2 (discussed below) illustrates how the dimensions I outline in my conceptualization of social capital are rooted strongly in several veins of research on social capital.

Table 3.1
Dimensions of Social Capital

Dimension	**Types**	**Definition & Dependencies**
Networks	• Teacher-Teacher • Teacher-Administration • Teacher-Parents • Administration-Community	Defined by number of connections, frequency of interaction, & shared obligations/expectations. Benefits dependent on norms, expertise, & depth of interactions. Social capital promoted by network closure
Norms	• Trust • Respect • Caring	Defined by teachers' perceptions of level of trust, their trust for others, & feelings of being trusted. Interdependent with networks.
Expertise	• Human capital • Informational Resources	Defined by teachers' reports of access to support from colleagues, admin, & others

Depth of Interactions	• Substantive • Cognitive	Dependent on available expertise, structure and frequency of interactions, & norms governing interactions

To understand how social capital functions to improve teachers' effectiveness, it is important to understand how human capital and social capital interact. First, any improvement in human capital should also serve to improve the social capital of a teacher by increasing the expertise available through her social network. On the other side of the coin, improvements in social capital ought to provide benefit to teachers in the form of increased learning, which would by definition improve their human capital. From the perspective of situated learning (see e.g., Horn, 2010) and sociocultural learning theory (Vygotsky, 1979) comes another illustration of how deeply intertwined social capital is with human capital. This theory views teachers' knowledge (or human capital) as socially constructed and posits that individual mental functioning is dependent on social structures and interactions. From a sociocultural perspective, learning occurs (human capital is created) through interactions (within social networks) based on the norms of a particular community (in other words, social capital is the basis for the creation of human capital). This also helps explain why Coleman (1990) envisioned social capital as a prerequisite for the accumulation of human capital and why Fullan and Hargreaves advocate for strengthening social capital as the "lead strategy" for school improvement (2013, p. 37). The prominent role of social capital in theory has attracted research interest from a variety of perspectives.

Approaches to the Study of Social Capital

The literature on social capital in schools is extensive, and suggests the potential for substantial benefits from teacher social capital in general and from each of the dimensions outlined in Table 3.1. Alongside that tradition, ideas that are largely synonymous with teachers' social capital are conceptualized using different terminology in several different research traditions. Literature on teacher working conditions overlaps with the construct of social capital while also measuring teachers' empowerment, trust in administrators, perceptions of professional development opportunities, and the physical conditions (physical capital) of the schools in which teachers work (see, e.g., Ladd, 2009). Social network theory, which examines how the configuration of social relations matters for a range of important outcomes (Coburn, Russell, Kaufmann, & Stein, 2012), also overlaps considerably with the study of social capital. Researchers focus on a number of different dimensions of social networks, notably tie strength, access to expertise, and depth of interaction routines (Coburn et al, 2012). These dimensions in some ways parallel social capital researchers' emphasis on information channels and norms, while drawing attention to the resources that are available to people through their social networks. In other words, social network analysis attempts to measure not just the frequency of collaboration and people's perceptions of trust, but also 1) the relative level of the expertise and information that people gain access to through their social networks; and 2) the types and relative depth of the interactions that take place among members of a social network. Table 3.2 outlines how different research traditions have conceptualized and defined these ideas. This table is presented as a brief comparison of some of the most prominent research in this area, meant to provide the reader with a sense of how researchers from various traditions have approached these questions in ways both similar and diverse. A full typology is beyond the scope of this paper. Instead, I hope to clarify two critical themes: 1) the importance of social capital has been demonstrated across a wide swath of research spanning divergent traditions, and 2) the differences in conceptual and methodological approach are likely to reveal different aspects of the construct of social capital.

Table 3.2
Comparing various conceptualizations of social capital

Coleman: Social capital	**Leana & Pil: Social Capital**	**Bryk & Schneider: Trust**	**Coburn: Social network analysis**	**Putnam: social capital**	**Ladd: Working Conditions**
Information Channels	Frequency of collaboration	-	Tie strength	Networks	Time for collaboration
-	-	-	Access to expertise	-	-
Norms/ Trustworthiness	Level of trust Quality of environment	Relational trust Respect Care	-	Norms of generalized reciprocity	Trust in leaders & colleagues
Obligations & Expectations	-	-	Interaction routines/depth	-	-

Coburn and colleagues (2012) make a compelling case that these differences in methodological and conceptual approach are important because they are likely to lead to different inferences about the importance of social capital and the impact that policies might have in promoting or suppressing it. For instance, many researchers have spoken of the importance of social capital in broad generalities, using terms such as “professional communities” but without clear definition. Other research has focused on measuring social capital through surveys of teachers’ perceptions, most notably the Chicago Consortium on School Research Survey and North Carolina Teacher Working Conditions Survey. Much of this research has focused on established social structures such as the school, the grade level, or the department as the locus of community. Social network theory, on the other hand, uses structured interviews, surveys, and observations to map patterns of individuals’ social networks and identify ties between individuals, the features of those ties, and how the ties between dyads interconnect to form a system of relations as a whole (Valente 1995). This type of intensive data-gathering has revealed that teachers' social networks vary considerably in the depth of interaction (Coburn and Russell 2008; Horn and Little 2010). Further, researchers have found that the depth of interaction varies even in networks with similar structure or resources (Coburn et al. 2013). While some interactions focus on quick exchanges about how students are doing or a pending deadline, others center on in-depth and substantive conversations about mathematical content or the nature of student learning. Frequency tables and measures of teachers' perceived levels of trust are unlikely to clarify which types of interactions are occurring, providing a strong rationale for additional emphasis on social network analysis.

Another unique contribution of social network theorists has been to focus attention on the mechanism by which social networks have impacts. Hypotheses have ranged from peer pressure and social control (Centola 2010; Coleman 1988) to providing a structure for coordinating action (Obstfeld 2005), to creating conditions for individuals and groups to learn from and with one another (Hansen 1999; Reagans and McEvily 2003; Uzzi and Lancaster 2003). Importantly, recent research has shown support for the hypothesis that teachers' social capital is providing benefits by means of creating conditions for teachers to learn from one another. By measuring social network interaction directly and assessing the depth of different interactions, researchers found that substantive, focused, targeted talk during which teachers grappled with pedagogical problems and math concepts led to enactment of the curriculum at deeper cognitive levels and greater ability of teachers to sustain the reform-based approach to instruction (Coburn et al, 2012). These results suggest that teachers' social capital provides benefits to teachers by increasing their learning rather than via social pressure or collective action. If social capital increases teacher learning, then the benefits for students may come about through improved teaching practices, and this dissertation will help to illuminate these relationships.

Benefits of Teachers' Social Capital

On the surface, everyone knows the motto, "it's not what you know; it's who you know." Clearly, popular sentiment acknowledges the importance of social capital. In the political arena, research related to social capital has entered the debate about educational reform, most notably in the slogan: "teacher working conditions are student learning conditions."

Beyond the popular imagination, a substantial body of research stretching back across five decades and spanning many different research traditions and approaches illustrates the importance of social capital for improving key outcomes for children. For instance, Putnam (2001) found strong correlations at the state and community level between the level of social capital and that group's educational performance, child welfare, health, amount of TV watching by kids, crime, aggression, percent of tax evasion, tolerance, economic and civic inequality, and happiness. Fullan and Hargreaves (2013) called for schools to develop their human, social, and decisional capital (the latest term defined as wisdom and expertise to make sound judgments about learners), citing examples of educational systems in Finland, Singapore, Alberta, Ontario, and California that use social capital to improve student outcomes.

Below I summarize the research evidence related to the four key components of social capital detailed before: networks, norms, expertise, and interactions.

Networks

Teacher-Teacher. Evidence for the importance of strong teacher-teacher networks comes from a variety of research perspectives and focuses mainly on how stronger networks are linked to greater learning for teachers. First, teachers learn more when they report talking to colleagues more often (Leana, 2011). Second, teachers who report talking to many different colleagues frequently are more likely to sustain reform-related instructional activities (Coburn, Russell, Kaufmann, & Stein, 2012). And third, research indicates that teachers improve mainly from on-the-job, informal learning (Kraft & Papay, 2014). In other words, social networks appear to be a key foundation for the ongoing professional learning of teachers, meaning that stronger social capital leads to improvements in human capital. Research has also linked stronger social networks with increased learning gains for students (Leana & Pil, 2006).

Teacher – Administrator. Research on distributed leadership has made clear that evidence lies in favor of a model of leadership that is "fundamentally concerned with building positive relationships and empowering others to lead" (Harris, 2002, p.15). In the same vein, Spillane (2005) defines distributed leadership as the interactions between people and their situation, rather than as the actions of individuals. Thus, we can see how the work on distributed leadership at its core is about the social capital available among a school's leaders. The flip side of the research on distributed leadership suggests that strong networks among teachers and administrators are likely to lead to more effective schools.

Teacher – Parent. A brief review of the research on parental involvement suggests two important findings that suggest that the parental involvement literature may have important connections to the study of teacher social capital. First, the amount of parental involvement has long been listed as a critical factor in school improvement, and second, some research suggests that the strength of teachers' networks with parents may positively impact the building of strong parental involvement. In relation to the first point, higher levels of parental involvement are associated with increased student achievement. For instance, a meta-analysis by Fan and Chen (2001) found a meaningful relationship between parental involvement, especially in terms of parental aspirations, and student achievement. And a study of several dimensions of the parental involvement of 24,000 eighth graders found that parents' discussion of school-related activities at home was moderately associated with students' achievement in reading and math (Sui-Chu & Wilms, 1996). This study also found that parents' participation at school was moderately associated with reading achievement. The second finding from parental involvement literature is what connects this research to my study of teacher social capital: research suggests that schools can influence the levels of parental involvement. For example, Bermúdez & Márquez (1996) found that school policies and the efforts of school personnel could change the level of parental involvement. Epstein (1988, p. 59), in synthesizing several studies of variations in school practices, posited an explanation for how schools might change the level or quality of parental involvement by concluding that, because "parents in all types of schools and at all grade levels express the need for clear communication about their children's attendance, behavior, academic progress, the curriculum, and how to help their children at home…" teachers ought to be recognized as the "key advisors" sought out by parents to find out how time spent at home can best benefit their children's learning. In other words, parents express a need for clear communication from teachers, and presumably, when this information is provided, parents can and at least sometimes do act on that information to improve their children's learning. Reinterpreting the parental involvement research from the perspective of social capital, we are left with the suggestion that the relative strength of the networks between teachers and parents may impact parents' interactions with their children and thus produce measurable effects on student achievement.

Administrator – Community. Leana and Pil (2006) analyzed social capital in two forms that they term internal social capital, measured by the strength of social networks among teachers, and external social capital, which was measured by the amount of time that principals spent interacting with stakeholders outside the school. They found that when principals spent more time on external relations like meeting with parents and developing community relations, the school's quality of instruction was higher and students' scores on standardized tests in reading and math were higher. It is important to note that this relationship was not established causally – it could be that principals in high functioning schools have the luxury of spending more time on community relations, or that the community of a high-achieving school demands more time from the principal, rather than that the external social capital is helping to drive improved teaching and learning. More work is needed in this area.

When people interact with one another regularly and develop strong networks, they also develop or alter the collective sense of interdependence, the underlying and typically unwritten norms of the community. Healthy and positive norms, described variously as "mutual obligations and expectations" (Coleman) or "generalized reciprocity" (Putnam) can function as internalized controls that make all interactions more efficient and effective. The following section discusses the importance of three key norms in school communities: trust, respect, and caring, with an emphasis on trust.

Norms: trust, respect, caring

The idea of trust has long been of interest to researchers. The depth and breadth of this research is such that it bears a quick summary in order to help us appreciate the potential importance of trust in school communities. Thirty years ago, Silver noted that the idea of trust has had a "centuries-long intellectual career" (Silver, 1985, p. 52, in Sztompka, 1999, p. 16). At the macro level, evidence suggests that in recent years that interest has grown, as sociologists, economists, and others have turned toward a focus on "soft variables…cultural intangibles…imponderables of social life" in general, and toward the construct of trust in particular (Sztompka, 1999, p. 11). Sztompka lays out four reasons behind researchers' turn toward theories resting on cultural intangibles such as meanings, symbols, and norms as opposed to theories such as behaviorism, game theory, and rational choice theory and models involving "hard variables" such as class position, demographic trends, and organizational forms. In brief, the reasoning rests on the growing recognition of the defects of institutional frameworks, the ways that the same institutions operate very differently when situated within different societies, and the key role played by cultural factors in geopolitics in general and the experiences of the post-communist Eastern European societies in particular. He details eight features of contemporary society that make trust a particularly relevant issue, and posits five reasons for the "new wave of sociological interest in trust" (p. 16). Trust has benefits not just for the two or more partners who agree to trust but for the wider community in terms of fostering the spread of communication, enabling collective action, civilizing disputes, strengthening the bonds of identity, and leading to greater cooperation and even sacrifices for others (p. 105). For our purposes, one particular reason stands out: trust is an integral component of social capital. Trust increases social capital, which Sztompka (p. 105) equates with what others have termed "moral density" (Durkheim in Cladis, 1992, p. 196), "spontaneous sociability" (Fukuyama, 1995, p. 27-

29), or “civic engagement” (Almond & Verba, 1965, p. 228). Trust and social capital also reinforce one another: “the theory of social capital presumes that, generally speaking, the more we connect with other people, the more we trust them, and vice versa” (Putnam, 1995b, p. 665, in Sztompka, p. 15).

Along with pointing us toward an appreciation of the complexity of the concept of trust, Sztompka’s work helps us to distinguish between the interpersonal trust that exists between people and the institutional trust that describes the perceptions of people toward institutions. For instance, teacher-parent networks are conditioned by institutional trust very differently than teacher-teacher networks because parents likely view teachers as representatives of the institution. Taken together, the evidence Sztompka presents also adds weight to the overarching claim in this paper that the social capital of teachers merits additional study. Sztompka shows how researchers are increasingly recognizing on a broad level the importance of trust and social capital for a well-functioning society. When we narrow our focus to the micro-societies that exist within our schools, it is not surprising that the evidence points toward a similar conclusion.

Most prominently, research from the Chicago Consortium on School research finds a number of key positive outcomes associated with schools’ levels of relational trust. First, relational trust varies substantially across schools and schools with more trust are associated with teachers being more committed to their school, more open to innovation, more willing to involve parents, and more committed to a collective responsibility for student welfare (Bryk & Schneider, 1996). This work focuses on the importance of relational trust as an organizational property, and emphasizes that it is likely to be even more important in loosely coupled organizations such as schools, where the principal actors do much of their work outside of the direct observation of supervisors. Schools are also dependent on trust among members because of their frequent need for collective action and the interdependency of the key members. To be successful, second grade teachers rely on first grade teachers doing their jobs well, just as teachers and administrators must rely on one another. Bryk and colleagues (Bryk, Lee, & Holland, 1993) see trust as a moral resource leading to better decision-making processes, and as a catalyst toward greater efficiency. For instance, greater trust reduces the number of contested decisions and leads to more time being spent on instructional goals rather than on debates or arguments. The normative value of trust also creates implicit, internalized obligations, and with this internal social control there is less need for formal policing and free riders and shirking lessen (Olson, 1965). Schools with greater trust, then, catalyze their teachers toward greater openness and commitment, function more efficiently, and tend to make better decisions. An established norm of trust also leads to greater learning for teachers, administrators, and students. A wide variety of literature demonstrates that people learn more in an atmosphere of trust. For instance, research on language acquisition theorizes that in an atmosphere of trust the learners' affective filter is lowered and learning of language improves (Krashen, 1981; 1982). It only stands to reason that teachers and other key adults would also learn more in a positive, trusting atmosphere.

Horn and Little's (2010) review sheds further light on how positive norms that go beyond trust play a role in leading to teachers' engaging in their work at greater levels of cognitive depth. They suggest that teacher social networks (alternately labeled "strong teacher learning communities," "inquiry communities," and "professional learning communities") provide learning benefits to teachers when they are characterized by norms of collective innovation, inquiry, and a capacity to disagree, and contrast these communities with pseudo communities in which the norms do not allow for critical self-examination (p. 233). Similar to the rich literature on opportunity to learn, this review suggests that teachers' learning is dependent on being afforded opportunities to reflect on, examine, and re-envision their instructional decisions. This evidence also provides an example of how the benefits from social networks are dependent on the quality of the norms that govern those networks.

Not surprisingly, Bryk and Schneider (2003) found that schools with high trust were more likely to demonstrate large gains in student learning. Schools with low scores on relational trust throughout the seven years of their study had almost no chance of showing improved academic outcomes in math and reading, whereas half of the schools that scored high on trust recorded substantial gains (averaging 8 percent in reading and 20 percent in math over five years). Tshannen-Moran and Hoy's summary of five decades of multidisciplinary research on trust in schools makes the importance of trust hit home. Following Baier (1986), they define trust as the reliance on others' competence and their willingness to take care of rather than harm what is entrusted to them, and then relate that definition to the way in which our schools look after our children, our money, our norms, our democracy, and our future. We trust our schools with so much, so it stands to reason that a school will need to have a high internal level of relational trust in order to function at a high level.

Much research also emphasizes the connection between trust, respect, and caring. For instance, Gambetta (1988) defined trust as resting on the elements of respect, personal regard, competence and integrity. Vodicka (2006) viewed trust as being made up of compassion, consistency, communication, and competency. And Bryk and Schneider (1996) saw trust as leading toward caring. Folk knowledge among teachers also recognizes the importance of caring, and implies that the following aspect of social capital, the expertise that is available to teachers through their networks, is at least in part dependent on the presence of a norm of caring: "No one cares how much you know, until they know how much you care." Yet research finds that teachers quite consistently earn high marks in caring and interpersonal warmth, and suggests that the variance in student performance outcomes is tied to sharp distinctions in the extent to which they provide their students with cognitively demanding and conceptually rich learning opportunities (Hiebert & Grouws, 2007). It is reasonable to explore how students' learning opportunities may depend in part on the extent to which teachers themselves have access to rich learning opportunities.

School climate research

Before moving on to discuss the expertise available to teachers, allow me to make a brief connection that illustrates the manner in which the dimensions of social capital are interrelated as well as the breadth of research that points to the importance of social capital in schools. Research on school climate relates closely to the benefits of both strong social networks and healthy norms among all stakeholders in a school community. School climate research defines a healthy school climate as depending on five elements: Safety, Relationships, Teaching and Learning processes, the Institutional Environment, and the School Improvement Process. Within these elements, connections to various dimensions of social capital are visible, especially the emphasis on norms for social-emotional safety, and the importance of support for professional relationships. Professional, positive relationships among adults have, in fact, been suggested to play a "critical foundational role" for a positive school climate (Thapa et al, 2013, summarizing findings of Guo, 2012, p. 364). Tschannen-Moran and Hoy (2000) also review research showing associations, likely running in both directions, between trust and a positive school climate. If strong social capital leads to a more positive school climate, then the benefits associated with a positive school climate will accrue to a school with strong social capital. These benefits, in brief, range from higher self-esteem and well-being for students to fewer incidents of psychological problems and substance abuse, from decreased absenteeism, suspensions, aggression, and violence to increased motivation to learn and improved academic outcomes (Thapa et al, 2013).

Expertise

Expertise, the third dimension of social capital, refers to the resources available to a teacher through her network. It stands to reason that a teacher surrounded by brilliant, hard-working colleagues will have more opportunities to learn and improve than a teacher languishing in the midst of a department of dullards. Indeed, teachers have been shown to learn from their most able colleagues, and students gained from being located in close proximity to a highly capable teacher as well (Jackson and Bruegmann, 2009). Jackson and Bruegmann’s research identified a number of effects that suggested peer learning among teachers is an important factor in explaining students’ success. Students had larger achievement gains in math and reading when they were in a class of a teacher with more effective colleagues, and teachers who were the weakest of their peer group showed the best performance relative to other teachers of that level. Kraft & Papay (2014) found that teachers working in schools rated higher in professional environment improved their effectiveness more over time than teachers working in less supportive contexts, thereby suggesting that peer learning is an important factor in teacher improvement and that the norms of a school may also influence the extent to which teachers are able to take advantage of peer learning. Taylor and Tyler’s (2012) study of a teacher evaluation program in Cincinnati showed that teachers improved their practice more when evaluated regularly by an expert teacher and one plausible explanation for these findings is that these teachers benefitted from greater access to expertise. In addition to the effects of learning from more able peers, teachers can also access expertise through groups learning together. Communities that share information and learn together can provide teachers with access to resources that would not be available to any individual attempting to learn on her own.

This dimension is also related to the often-noted observation that human capital and social capital interact. For instance, improvement in human capital serves to improve the social capital of a teacher by increasing the expertise available through her social network. On the other side of the coin, improvements in social capital ought to provide their benefit to teachers in the form of increased learning, thus improving their human capital.

Depth of interactions

Research on the depth of interactions sheds light on the mechanism through which social capital provides teachers with resources for learning. This research describes in detail how the relative depth of teacher interactions plays a key role in how much teachers benefit from those interactions (Horn & Little, 2010). Horn and Little examine about 100 hours of collaboration among six Algebra teachers at a diverse working class urban high school that showed students outperforming their peers in more affluent high schools and enrolling in advanced math courses at a higher than average rate. They detail how these teachers act out with their colleagues the complex teaching choices they face. Using what she terms teaching replays and rehearsals, the teachers are able to substantiate claims about teaching, anticipate student responses, and elicit emotional support. They also coordinate expectations and strategies with their peers to re-imagine how a teacher might intervene in a similar situation in the future, to reconsider student (dis)-engagement in terms of how it reflects students' anxieties and (lack of) conceptual understanding, and to rehearse clearer methods for explaining content.

Coburn and colleagues found that the depth of interaction varies even in networks with similar structure or resources (Coburn, Mata, & Choi, 2013). While some interactions focus on quick exchanges about how students are doing or a pending deadline, others center on in-depth and substantive conversations about mathematical content or the nature of student learning. Substantive, focused, targeted talk during which teachers grappled with pedagogical problems and math concepts led to enactment of the curriculum at deeper cognitive levels and greater ability of teachers to sustain a reform-based approach to instruction (Coburn, Russell, Kaufman, & Stein, 2012). In other words, it is the quality and depth of interactions that is key, rather than simply their frequency.

Evidence for Benefits of Social Capital for Students

The association between social capital and student learning has been demonstrated across a wide variety of studies, using different methodological approaches, and looking at a wide range of data, from the state level to individual schools and classrooms. At the broadest level, Putnam (2001) showed an association between states and communities with strong social capital and those with stronger educational performance. This by itself cannot be taken as evidence of a causal impact, but others have drilled down to find compelling causal links. At the school level, Bryk and colleagues' (2003) found that schools with high trust are more likely to demonstrate large gains in student learning. At the classroom level, Leana & Pil (2006) reported improved student achievement in mathematics when teachers frequently asked one another for instructional advice and indicated a feeling of mutual trust). Overall, the literature suggests that social capital has been shown to provide benefits to students through three main kinds of mechanisms, namely through improved teacher learning, lower teacher turnover, and improved teacher and student interactions.

Improved teacher learning and practice. Kraft and Papay (2014) found that a positive professional environment related to improved teacher growth, and they measured teacher growth by how teachers' value-added effects changed over time. Researchers have also found evidence of specific improvements in instructional practices that are linked to the presence of strong social capital. For instance, Ferguson and Hirsch (2014) found that manageable time demands and quality of professional development positively related to improved teacher practice as measured by observation rubric scores in the domains of academic press and academic support. Jackson and Bruegmann (2009) showed that teachers do better when they are surrounded by more able colleagues, thus suggesting that stronger social capital leads to greater teacher learning. Social network analysts have also examined questions of when, how, and under what conditions teachers change their practice for the better and sustain promising reforms, and these researchers have found support for the idea that strong social capital is linked to teacher learning. On the question of teacher change, teachers with high levels of access to expertise and response to social norms were more likely to adopt technological innovations (Frank, Zhao, and Borman, 2004). Once teachers have innovated, they often struggle to sustain these new practices, but strong social networks have also been shown to have positive benefits in influencing teachers to sustain reform-related instructional approaches in math (Coburn, Russell, Kaufmann, Stein, 2012). In this study, district support for a new approach to math instruction was strong for two years and involved the use of math coaches and regular professional development sessions that served to strengthen teachers' social networks. In the third year, the district support was abandoned as a new superintendent shifted priorities to English language development. However, teachers who had developed social networks with combinations of strong ties (they interacted with others in their network frequently), high-depth interaction (they discussed issues of substance regarding math teaching and learning), and high expertise (the people in their network had substantial math training) in the first two years of the initiative were more able to adjust instruction to new conditions while maintaining the core instructional approach.

The theory that teachers will learn from enhanced social networks is also the implicit backdrop to districts' support for Professional Learning Communities (PLCs), in which teachers meet regularly to identify goals, ways to assess progress toward those goals, and how to respond when students fail to meet those goals and Peer Assistance and Review (PAR) programs, in which veteran teachers mentor new and struggling teachers. Looking internationally, the success of Finland's education system is often attributed in part to the great amount of time afforded to teachers to work together outside of their classroom responsibilities. South Korea, similarly high performing but from a very different culture, structures its teachers' workdays in such a way that they spend three or more hours a day working together in a common office and typically go out to lunch with their colleagues every day (C. Myers Asch, personal communication, Feb. 19, 2015). Another popular strategy for strengthening teachers' social capital in hopes of leading to ongoing informal learning is through the use of teacher coaches (Keller, 2007). Observational research supports the efficacy of this model, showing that student achievement was higher with teachers who had more contact with coaches (Ross, 1992) and this research reinforces the theory behind social capital in that the presence of coaches is likely to improve both the expertise that is available to teachers and the frequency of their collaboration.

Lower teacher turnover. Improved social capital also indirectly improves human capital by lessening teacher turnover since poor working conditions make teachers more likely to leave (Ladd, 2009; 2011; Kraft et al 2012). Less teacher turnover should improve human capital because teachers improve with experience (Clotfelter, Ladd, & Vigdor, 2007). Returns to experience among teachers have been studied extensively, with many researchers finding that teachers improve most dramatically during their first few years on the job, with the average improvements leveling off after three-five years (see e.g., Goe & Stickler, 2008). The consensus has appeared to be that, on average, teachers do not improve their practice after the first few years. However, recent research suggests that at least some teachers continue to improve in later years of their careers, and further that this improvement comes mainly from the informal learning that typically develops within social networks (Kraft & Papay, 2014). The same authors reported that working environments characterized by trusting environments, time to collaborate, and professional development opportunities related positively to both teacher satisfaction and student achievement (Johnson, Kraft, and Papay, 2012; Kraft and Papay, 2014). These new findings have also turned attention away from the average returns to experience and toward examining what factors are associated with those teachers who continue to learn and improve throughout their careers. So a school with a high level of social capital suggests a scenario in which a new teacher will get a bit better over time, and will be more likely to stay at that school. Because others at that school are also more likely to stay, and thus more likely to have improved, our new teacher will learn more from her more able colleagues. In this more stable teaching force, relationships will generally strengthen with time, and the greater closure of this stable staff will tend to promote ever greater social capital (Coleman, 1990), thus leading to a positive feedback cycle. This scenario also suggests that improving the social capital/professional environment/working conditions at schools may provide at least a partial solution to the oft-noted problem of high teacher turnover at schools serving large proportions of disadvantaged students. This high turnover exacerbates inequality by resulting in disadvantaged students being more likely to be taught by inexperienced and less effective teachers (Clotfelder, Ladd, Vigdor, & Wheeler, 2007; Roza, Hill, Sclafani, & Speakman, 2004). Fullan and Hargreaves also point out that high levels of social capital are a prerequisite for attracting new high quality teachers (2013).

Stronger teacher student relationships. Leana and Pil (2006) also show evidence of a direct effect of social capital on reading achievement. This is in contrast to the indirect effect they find for math achievement – the effect for math achievement is mediated by improved instruction. They speculate that the direct effect of social capital on reading achievement is the result of reading improvement coming about through factors such as greater teacher involvement and input on literacy rather than through specifically improved instructional practices. Applying these findings on a broader scale, we might imagine that if teachers work in an environment in which they feel tightly connected to one another, they might be more likely or better able to establish strong relationships with their students and their students' parents, either of which could explain improved student learning even in the absence of any changes in instruction.

Improved social outcomes for students. As noted in chapter II, social outcomes for children are key in and of themselves. Improving teachers' social capital means educating students within institutions that are trusting, caring, and connected, and thus it is likely to also be critical for improving students' levels of trust, caring, and psychological outcomes such as motivation and identity. Levels of trust at regional levels are linked to peoples' participation in democracy and other civic institutions (Rothstein & Stolle, 2008; Sztompka, 1998). In addition, if students experience higher levels of trust, caring and feelings of connection while in school, this may lead them to engage in greater communication, cooperation, and civility in the wider community (Sztompka, 1999). These non-achievement outcomes are not directly addressed within this dissertation; this limitation should not be interpreted as implying any less importance to them, but simply as a reflection of the fact that adequate measures of students' social outcomes were not identified.

Obstacles to Improving Social Capital

Schools face several obstacles in developing teachers' social capital. First, the teaching profession in the U.S. has a long and still-pertinent history of an occupational norm emphasizing privacy, as Lortie's (1975) seminal work on the sociology of teachers brought to notice. In common parlance, U.S. teachers have for generations tended to "shut their classroom doors and teach the way they know how" rather than to adjust to reforms or engage collaboratively with colleagues. This norm inhibits the growth of social capital. Second, researchers have pointed out a lack of consensus among teachers about what constitutes good teaching (Bryk & Schneider, 1996). This lack of consensus has been variously attributed to a lack of coherence among teacher training programs or to the relatively high degree of ambiguity in research literature. Regardless of the cause, even teachers acknowledged to be successful often express doubts about their effectiveness (Lampert, 1985). Taken together, the tendency toward privacy, the lack of consensus, and the prevalence of self-doubt are likely to weaken social capital among teachers.

The third main obstacle comes from outside schools: since the 1960s, social capital in the U.S. has eroded at an alarming rate (Putnam, 1995). And along with that, Putnam (2001) identified a declining trend in altruism and trust for public institutions. Concurrent with these trends, people in the United States today trust traditional public schools much less than a generation ago (Tschannen-Moran & Hoy, 2000). Today, parents of 1.7 million U.S. students choose to homeschool their children rather than trusting them to schools of any kind, a number that was closer to zero a generation ago.[3] Millions more children now opt for charter schools and the rhetoric about school choice often stresses the fear of the many bad choices that are lurking in schools labeled as underperforming or failing. From *A Nation at Risk* in 1983 to No Child Left Behind to Race to the Top, the greater emphasis on high standards and accountability has fostered blame. The growing diversity and transience of society have made trust more difficult, and the tendency for distrust to be self-perpetuating has accelerated the problem (Tschannen-Moran & Hoy, 2000). Tschannen-Moran and Hoy point out how the growing distrust of schools has been accompanied, paradoxically, by a growing need for trust. Alongside their traditional role as caretakers of youth and molders of citizens, Americans increasingly ask schools to achieve ambitious goals of equity and economic advancement. These increased demands on schools suggest a greater need to trust schools. In addition, many of the reforms of American education implicitly rely on a base of trust and strong social capital. For instance, new forms of governance focusing on site-based management depend upon trust among school leaders and teachers, parental involvement initiatives require trust among teachers and parents, and cooperative learning in the classroom requires teachers to trust students.

Measuring Teacher Social Capital

As discussed, teacher social capital is a large and complex construct that has been examined from many different research perspectives. A great deal of research has fleshed out the

[3] http://nces.ed.gov/pubs2013/2013028/tables/table_07.asp

implications of just one aspect of social capital – trust – across a wide variety of settings. Others have approached this general idea from the angle of examining teachers' working conditions, the school climate, or types of school leadership practices. And many have sought to investigate the entire construct of social capital, in schools, other organizations, and across communities.

The wide variety of research perspectives on these issues suggests two complications potentially important for measuring this construct: 1) some aspects of teacher social capital may be distinct while others interact strongly and prove to be better measured as a single construct; and 2) some dimensions of social capital may prove easier to measure than others.

The different dimensions of teacher social capital outline above appear in some cases to be clearly distinct. For example, the amount of expertise potentially available to teachers in terms of things like the collective years of experience or level of training does not depend on the extent to which teachers trust one another. Thus, it appears likely that these two aspects can be measured as distinct dimensions. But other aspects appear deeply intertwined, suggesting that these dimensions may prove difficult to distinguish as separate factors. For instance, deep interactions focused on collaborating to create unit plans seem much more likely in a climate of mutual respect and among a staff who are afforded regular time to collaborate. Thus, dimensions such as these are expected to correlate strongly. It is hard to imagine, for instance, a school with a great deal of trust in which teachers almost never talk to one another or in which all the interactions take place at a superficial level. Measuring these dimensions as distinct factors may prove difficult. However, the conceptual distinctions appear substantively meaningful. The distinct factors may hold importance from an academic perspective in generating insights and revealing relationships that might be overlooked, and from a policy perspective in understanding specific mechanisms of influence and potential intervention. To give an example, a school community characterized by healthy norms of trust, strong social networks, and cognitively deep interactions but inhabited entirely by novice teachers is different from and implies different policy prescriptions than a school made up of many veteran teachers and a few novices, all of whom subscribe to a healthy trust in one another and interact frequently, but do so at a generally low level of cognitive engagement. Therefore, adequately measuring the breadth of a complex, multidimensional conceptualization of teacher social capital can be expected to be challenging.

The next measurement complication is that some dimensions of social capital may prove easier to measure than others. For instance, social network analysts have contended that certain dimensions such as depth of interactions and strength of the ties within networks require resource-intensive efforts in order to measure them adequately. Social network analysts have used multiple cognitive interviews or extended observations of collaboration among staff members to measure these dimensions by mapping out each individual's network and categorizing the relative strength of each tie between individuals as well as categorizing the relative depth of each interaction (see, e.g., Coburn, Russell, Kaufmann, & Stein, 2012; Horn & Little, 2010). It is an open question as to whether these dimensions can be reliably measured using methods that require a smaller commitment of time and resources.

The cognitive interviews and on-site observations conducted by social network analysts are two specific methods of gathering information on teacher social capital that can be described as in-depth, intensive approaches. Another in-depth approach is the ethnographic approach that embeds the researcher within a community in an attempt to allow the researcher to become privy to what are otherwise private interactions, ties, and norms. On the other end of the spectrum, researchers have gathered information on social capital through surveys of the participants in the organization or group. Surveys, in our case of teachers and perhaps of other school staff, have the advantage of potentially providing a great deal of information at a relatively low cost in terms of time and resources. They also may suffer from self-report bias and may not allow for the gathering of the in-depth type of information described above.

Teacher Working Conditions Survey

This dissertation relies on items from the Teacher Working Conditions Survey (Hirsch & Emerick, 2007), which, as its name suggests, is based in a research tradition investigating teacher working conditions rather than teacher social capital. This research measures constructs not generally considered part of social capital such as *Time* available to teachers and the adequacy of the physical plant, or *Facilities,* but on the whole, the conceptualization of teachers' working conditions overlaps considerably with the above conceptualization of teachers' social capital. Therefore, I anticipated that these survey items would provide an adequate measure of teacher social capital. I scrutinize the extent of that adequacy in chapter V.

The Teacher Working Conditions Survey (TWCS), sometimes referred to as the Teaching, Empowering, Leading, and Learning (TELL) survey instrument, was developed through the 2002- 2009 Governor's Teacher Working Conditions Initiative in North Carolina by Eric Hirsch at the New Teacher Center (Hirsch & Emerick, 2007; New Teacher Center, 2014). Conceptually, the survey as it was implemented in 2009 was intended to measure eight constructs/dimensions of teacher working conditions: termed *time management, facilities and resources, community support, student conduct management,teacher leadership,school leadership, professional development,* and *instructional support.* Validation work by Swanlund (2011) using a Rasch model found support for a structure largely similar, with some changes in the way items were grouped. Internal validity analyses by the New Teacher Center using factor analysis found support for this basic eight factor structure, though some items in the community support and involvement dimension had cross loadings and were deemed in need of further scrutiny (New Teacher Center, 2014). Swanlund's analysis concluded that the TWCS can produce consistent results across different samples and the New Teacher Center analyses found high internal consistency (alpha ranging from .86 - .96 for each factor).

Ladd's 2009 research used a conceptualization of the TWCS as measuring six dimensions of working conditions (with some items recategorized based on factor analyses). Ladd provides evidence that some elements of working conditions measured by the TWCS (particularly leadership and the availability of common planning time) have significant relationships with student achievement and teachers' decisions to continue teaching. Kraft and Papay (2014) selected 24 items from the survey to measure the professional environment in terms of *order, peer collaboration, principal leadership, professional development, school culture, and teacher evaluation,* and, based on principal components analysis and internal consistency, combined these items into a single composite. Teachers who worked in schools that scored high on this composite measure of professional environment were found to improve their effectiveness more than teachers working in schools that scored lower on professional environment. Taken together, Ladd's and Kraft and Papay's findings suggest that the TWCS has predictive validity as a measure of teacher working conditions in relation to these criteria of teacher departures, student achievement, and improved teacher effectiveness. In prior research using the MET dataset, Ferguson and Hirsch (2014) used composite indices of the first five dimensions of the TWCS listed above, along with modified versions of the last three indices to separate teachers into four categories. They found significant relationships among some composites of teacher working conditions, particularly those representing *community support*, *professional development,* and *instructional support*, and teachers' levels of academic press, expectations for students, and value-added effectiveness, thus providing additional support for the validity of some TWCS items. Ferguson and Hirsch's conceptual model and findings are discussed in more detail in chapter IV.

Conclusion

In sum, we have evidence that stronger social capital in schools may lead to important benefits for important people (students and teachers). Strong social capital is linked to improved student achievement, teacher learning and practice, teacher retention, teacher-student relations, organizational efficiency, and student social outcomes. Further, we have the suggestion from Coburn and colleagues (2013) that we *can* improve the social capital available to teachers, who are arguably the most important factor impacting student outcomes. Finally, important outstanding questions remain about how to adequately measure teacher social capital, and the TWCS holds promise as an instrument that can validly represent some of the most important aspects related to teacher social capital. This dissertation will shed light on these measurement issues as well as on how teacher social capital impacts teaching practices and student achievement, providing evidence on which dimensions and which pathways appear to hold the most promise for improving student outcomes.

Chapter IV: Model Framework

The preceding chapters make the case for the importance of studying how teacher social capital influences teaching practice and student learning. The hypothesized conceptual framework relating these constructs extends the conceptual model in Kane & Staiger (2012) by adding the impacts of teacher social capital. In their model, teaching is hypothesized to cause student outcomes. Teaching is measured by classroom observations, student surveys, gains on state tests, and combinations of these indicators with one group of students, and student outcomes are measured by gains on state tests, gains on supplemental tests, and positive student feedback in regard to the same teacher working with another group of students. In the model employed in this dissertation (see Figure 4.1) teaching practice is incorporated as a mediating factor rather than an initiator, and social capital has direct effects on both teaching practice and student achievement. This model also differs from Kane and Staiger's model in that measures of student achievement are viewed solely as student outcomes rather than being used to indicate both teaching practice and student learning. In Figure 4.1, thin arrows show how observed indicators are hypothesized to be related to underlying latent traits. Thick arrows indicate hypothesized relationships between these latent traits.

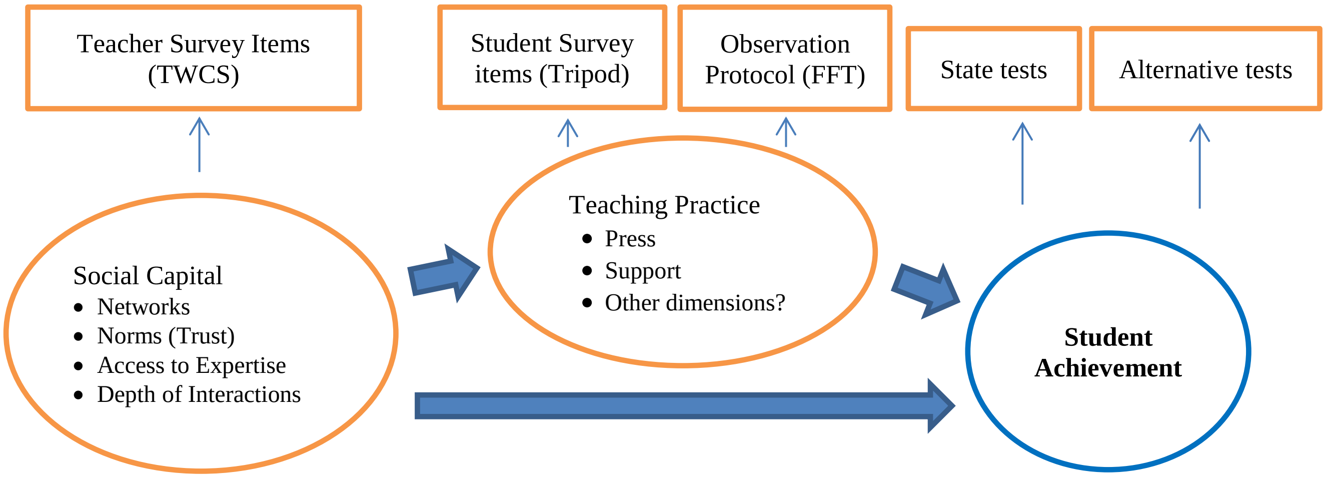

Figure 4.1. Conceptual Framework: Latent Constructs in ovals. Observed indicators in rectangles.

Teacher social capital and teaching practice may also be conceived as co-existing, related, and mutually reinforcing factors that all impact student outcomes. High-quality teaching practices may result in teachers placing increasing amounts of trust in colleagues and being more inclined to form strong networks with them. (see Figure 4.2). This model is also a reminder of the need to be cautious in drawing causal interpretations from observational data.

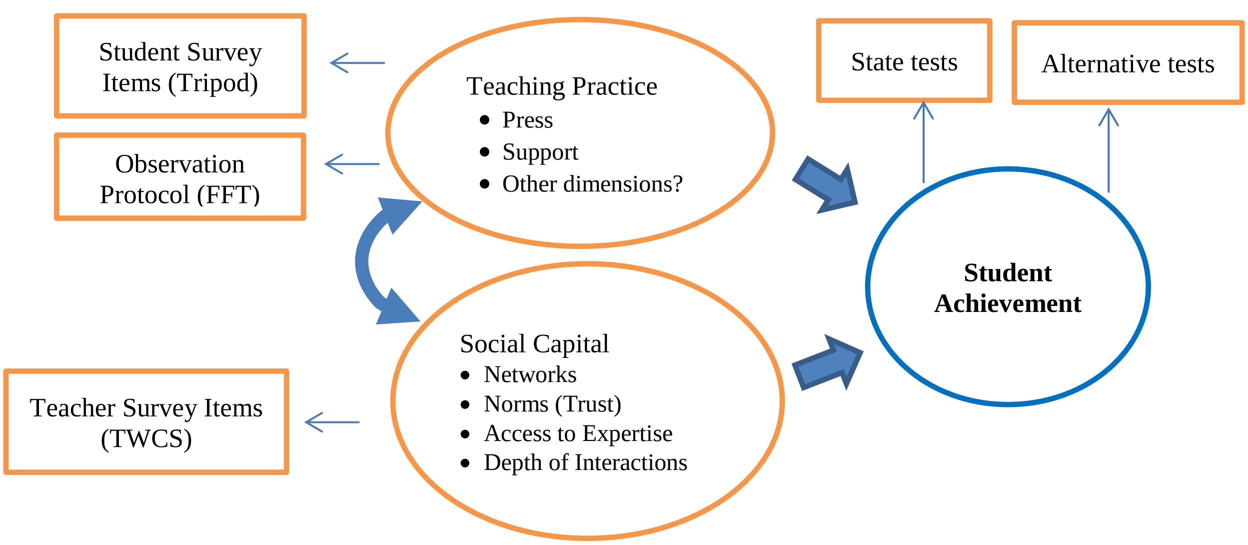

Figure 4.2. Alternative Conceptual Framework: Double-headed arrows indicate relationship with no hypothesized direction (not causal or causal impact in both direction).

Another alternative model is offered by Hirsch and Ferguson (2014). This model (see Figure 4.3 below) separates working conditions into "base" conditions and "teaching enablers", and separates teachers' underlying "beliefs and behaviors," from "teaching quality." The model aligns with what is known about our educational systems in that multiple factors are hypothesized to impact the quality of teaching and learning. It is also worth noting that their study, although resting on a quite different conceptual framework, also employs the MET dataset and takes advantage of many of the same measures as those used in this dissertation.

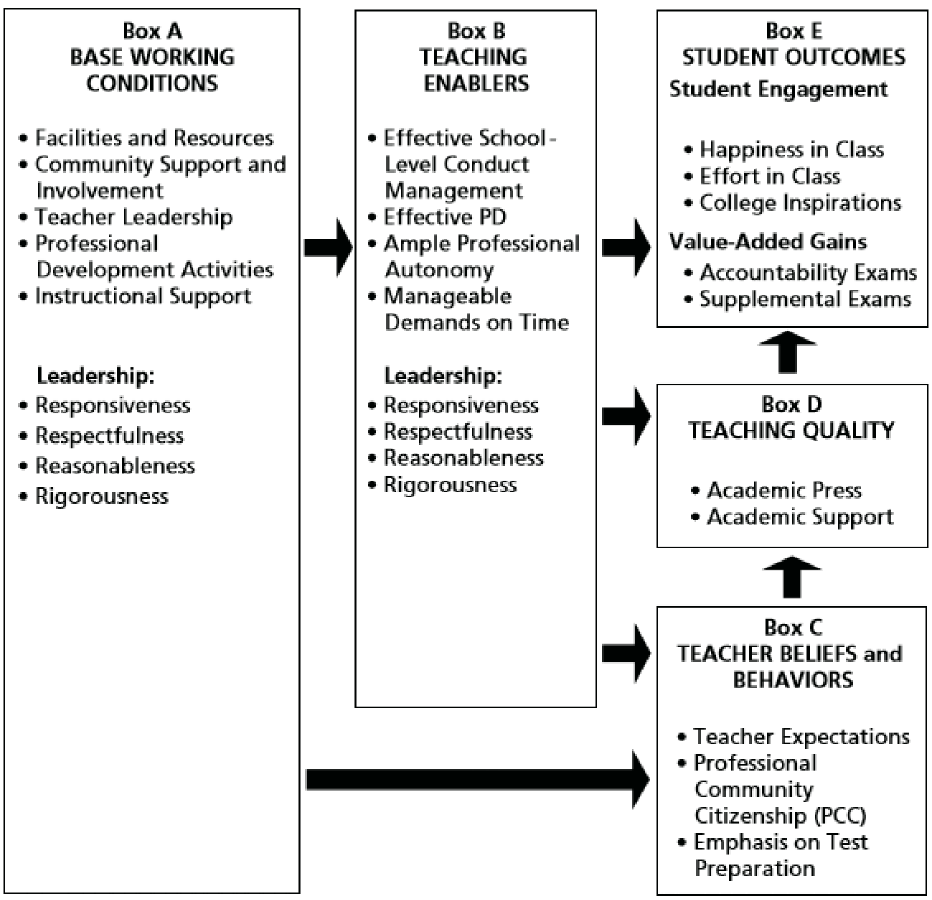

Figure 4.3. Ferguson & Hirsch's (2014) conceptual framework linking teacher working conditions to student outcomes. Note: From "How Working Conditions Predict Teaching Quality and Student Outcomes," by Ferguson, R. F & Hirsch, E. (2014) in Kane, Kerr, & Pianta, eds.,

Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project. 332 – 380. Bill and Melinda Gates Foundation. San Francisco, CA: Jossey Bass.

The conceptual framework in Figure 4.1 also has roots in organizational paradigms in educational research, illustrated in the comprehensive model of education shown in Figure 4.4 (Oakes, 1986). A comparison of the two models calls attention to potentially important factors such as fiscal resources and curriculum quality that are outside the scope of this study, but also reveals a number of parallels, such as the distinction drawn between teacher quality and teaching quality, and the impacts of school quality factors on the quality of teaching practices. Another point of similarity is that both models view the educational system as a multilevel structurally interconnected organization. In detail, these models present "potentially testable structural relationships among the variables comprising the inputs, processes, and outputs of schooling," and "the model is inherently multilevel in form, with a subset of the inputs and processes occurring at higher levels of the educational system" (Kaplan & Elliot, 1997sem, p.324).

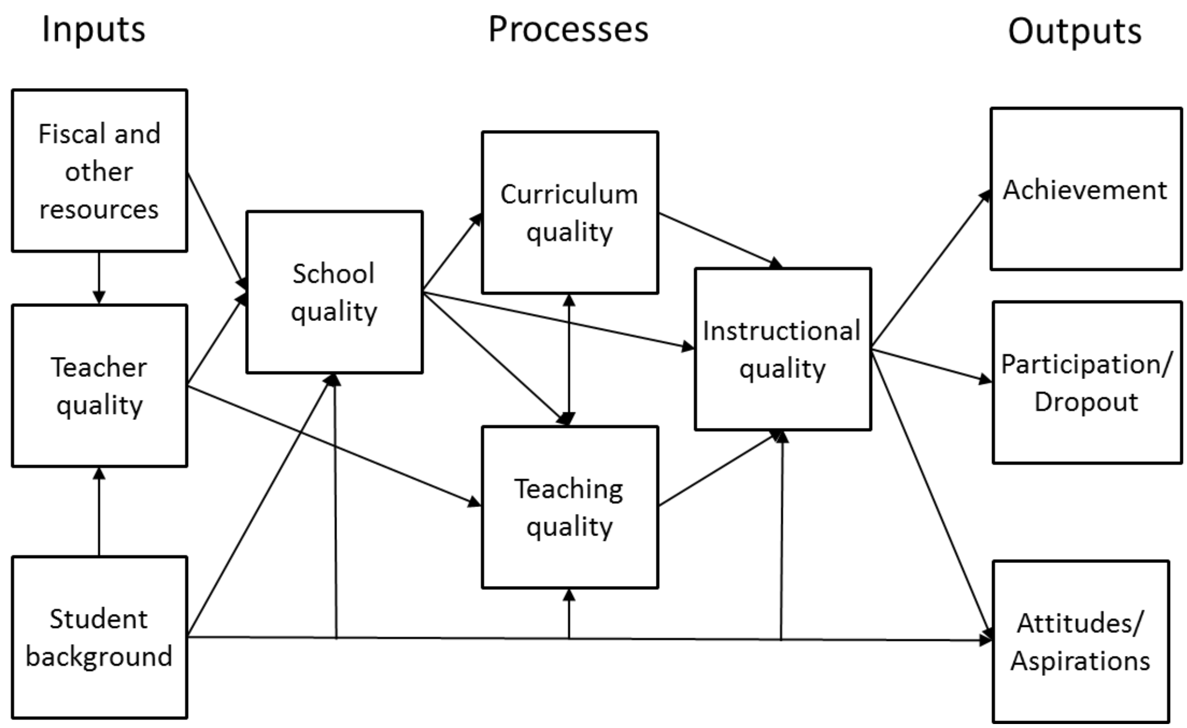

Figure 4.4. Oakes' (1986) comprehensive model of the educational system. *Note: From* "Educational indicators: A guide for policymakers," by J. Oakes (1986). New Brunswick, NJ: Rutgers University, Center for Policy Research in Education.

Modeling Approaches

The central research question examined in this study concerns the relationships among teacher social capital, teaching practice and student achievement. As the models in Figures 4.1 to 4.5 demonstrate, empirically addressing this question involves a number of data and analytic complexities. In particular, the hypotheses involve using clustered data to investigate relations among unobserved constructs that are imperfectly measured by multiple manifest variables, and connected through multilevel mediation links. The next section discusses a range of modeling approaches that can be used to investigate these questions. Below I discuss traditional approaches that include single-level manifest models, multilevel manifest models, and single-level latent models. Then I discuss the characteristics of the Multilevel Structural Equation modeling (MSEM) framework, and the rationale for its suitability for this particular area of application.

A number of different modeling approaches have historically been used to investigate conceptual models like those in Figures 4.1 to 4.4 that involve relations between teacher social capital, teaching practice, student achievement, and other outcomes and variables of interests at multiple levels. Within different approaches, the main options that can be identified are shown in Table 4.1. These include manifest models, typically with manifest composites formed after using factor analysis or principal component analysis to assess the dimensionality of the constructs. These manifest models can be single-level models that ignore any dependencies among individuals who are organized into groups such as classrooms or schools or the models may employ multilevel modeling techniques to account for those dependencies. The models also may be one-step models that investigate a single-step within the overall system such as the effect of school quality on teaching quality, or the effect of instructional quality on student achievement. Finally, researchers may employ latent factors at any or all points within the model, in order to include multiple indicators of a construct that cannot be perfectly measured and thereby account for the measurement error within the model. Researchers choose among these options to develop their overall modeling approach – for instance, a single-level model may include manifest or latent factors, and be a one-step model or a mediation model.

Table 4.1
Modeling options for investigating multilevel relationships among latent variables

Modeling Option	**Weakness**	**Alternative**
Manifest model	Ignores measurement error (treats observed indicators as if they are measured without error)	Measure latent factors by multiple observed indicators

Unidimensionality assumed	Ignores possible multi-dimensionality of constructs	Assess dimensionality through factor analysis
Single-level model	Ignores dependencies in individuals organized into groups (clustered) / Conflates constructs and effects across levels (ecological fallacy)	Use Multi-level model
One-step model $X \rightarrow Y$	Ignores mediation, treats the question of *how* impacts happen as "black box" / Ignores contextual factors that may themselves be causing the causal factor	Use Mediation Model $X \rightarrow M \rightarrow Y$
Multilevel model for mediation	Conflates within and between components of effects between pairs of level 1 variables	Use multilevel structural equation modeling

Some examples of studies that investigated similar questions to those considered here include Ferguson and Hirsch (2014), Kraft and Papay (2014), and Ladd (2009). Ferguson and Hirsch use factor analysis to assess and amend the structure of their conceptualization of teacher working conditions, but then create composite indices of each factor of interest and use these composites in regression models to estimate how different dimensions of teacher working conditions predict teaching behaviors and student outcomes. Kraft and Papay focus their investigation on a single composite indicator of professional environment of the school, formed based on a factor analysis and used in cross-classified multilevel regression models to estimate the returns to experience for teachers and how those returns differ for teachers who work in schools with more or less supportive professional environments. Ladd uses factor analysis to group items from the TWCS into five factors and then creates school-level composites for each of these factors. These composites are used in linear regression and multinomial logistic regression models to estimate how teachers' planned and actual departures and students' test outcomes differ based on the working conditions composite factors. All of these approaches differ from the approach taken in this dissertation in two main respects: First, they separate the measurement analysis from the structural analysis of the relationships among the constructs, using the measurement analysis to arrive at composite indicators of the constructs of interest and thereby ignoring the measurement error when estimating the structural relationships whereas the approach in this dissertation estimates the measurement model within the structural model and thereby accounts for measurement error in the estimates. Second, Ferguson and Hirsch, Kraft and Papay, and Ladd estimate one-step models that do not estimate any intermediate effects from mediators.

Multilevel Manifest Model

Other studies have used manifest multilevel mediation models to investigate similar questions. For example, Lüdtke, Robitzsch, Trautwein, & Kunter (2009) aggregated student ratings to the classroom level to provide information about the learning environment and Anderman (2002), used student reports aggregated to the school level to examine the relations between school belonging and student depression and other psychological outcomes. When multilevel models involve mediated relationships, they are typically categorized in terms of the levels at which each construct is measured (Krull & MacKinnon, 1999). Within the context of multilevel modeling, the mediation model in this dissertation is referred to as a 2-1- 1 model because it involves a construct measured at level 2 impacting a mediating construct measured at level 1 and through that mediator as well as through direct effects having effects on the outcome measured at level 1 (see Figure 4.5). Teacher social capital is measured at level 2, via teachers' responses to survey items, and in the figure, the first estimate is of the direct effect of this level 2 variable on teaching quality, a variable measured at level 1. Second, the direct effect of teacher social capital on student achievement, another level 1 variable, is estimated. And third, the indirect effect of teacher social capital through teaching quality on student achievement is estimated. Teaching practice is measured at level 1 through student survey items collected from students who are nested within teachers (or, in the second model, through observer ratings collected from multiple raters for each teacher). In the manifest multilevel model, these student survey items (or observer ratings) are aggregated to form manifest composite indicators of teaching press and support, and these manifest composites ignore any measurement error coming from the items. Student achievement is a level 1 variable measured through individual student test scores.

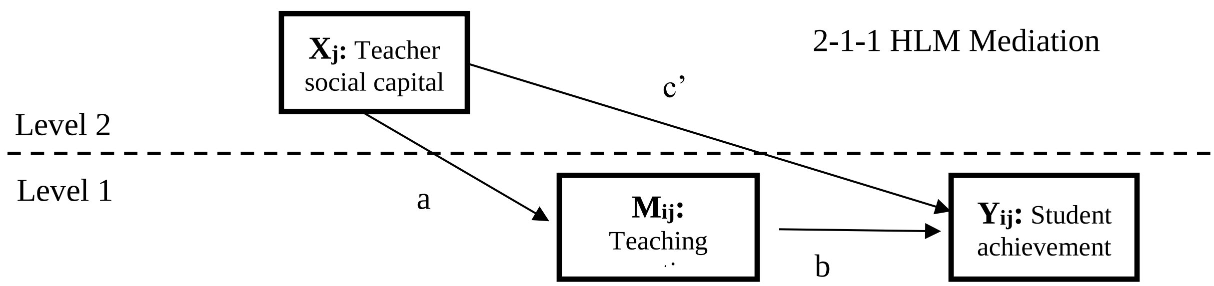

Figure 4.5 2-1-1 HLM Mediation Model. Note: adapted from “Testing Multilevel Mediation Using Hierarchical Linear Models: Problems and Solutions,” by Z. Zhang, M. J. Zyphur, & K. J. Preacher (2009) *Organizational Research Methods*, *12*(4), 695-719.

Single-level Latent Variable (Structural Equation Model: SEM)

Another plausible modeling approach to investigate these questions is a single-level latent structural equation model (SEM). In a single-level SEM mediation model, the responses to student survey items are aggregated to the classroom/teacher level and then these classroom average responses for each student perception survey item are included in the model. The estimate of teaching practice dimensions for each teacher will be based on a measurement model that is a confirmatory factor analysis of the classroom average scores for each item. This means that the model will provide estimates of the measurement error for each item, so that, for instance, an item that does not load strongly on a particular dimension could be a candidate for being dropped due to being an unreliable measure of these constructs. But the model will ignore the differences among students' responses within each classroom. Ignoring this source of variance may result in inflated estimates of the reliability of the items. Student test scores are similarly aggregated to the classroom level.

Multilevel Latent Structural Equation Model (MSEM)

The final model considered here is the multilevel structural equation model (MSEM). MSEM is a flexible and widely applicable modeling framework that combines strengths of latent variable modeling and multilevel modeling and encompasses multiple regression, path analysis, and multilevel modeling as special cases (Muthén & Asparahouv, 2007). Because the MSEM framework incorporates multiple indicators of latent constructs in multivariate data and directly models group dependencies in clustered data, it accounts for both measurement and sampling error (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011).

However, MSEM as a methodology is still in its relative infancy; though empirical studies have on the whole found support for the utility and accuracy of MSEM (Preacher, Zhang, & Zyphur, 2011; Lüdtke, Marsh, Robitzsch, & Trautwein, 2011), the results from Lüdtke and colleagues suggest that under certain conditions, particularly when factor loadings differ across the student level (Level 1 or L1), and the classroom group level (Level 2 or L2), and when the data provided less information about the L2 construct (e.g., low ICC, small number of L1 units), approaches with a simplified measurement model may outperform the MSEM model. As recently as 2010, Preacher and colleagues stated there was "no empirical evidence" that the theoretical advantages of the MSEM mediation model are borne out in actual data (Preacher, Zyphur, & Zhang, 2010, p. 163). Therefore, more empirical examples are needed to compare the results from MSEM with other commonly used modeling approaches. This dissertation provides one such example, examining the MSEM in comparison to other modeling approaches and considering how these approaches differ in terms of treatment of latent variables, multilevel data, and mediated relationships.

Latent variable framework. The first key strength of MSEM is that it incorporates latent variables, permitting us to estimate relationships involving unmeasurable constructs such as intelligence, quality, and learning. Constructs such as these are imperfectly measured by multiple observed variables, providing direct estimates of the extent of measurement error in the observed variables, and allowing the analyst to partial out that error from the estimated relationships among latent constructs. Data that contain multiple measurements of the same construct are often termed multivariate data and are common in educational settings. However, many researchers use manifest models employing a single fallible indicator for each construct (e.g., Hirsch & Emerick, 2007; Loeb, Darling-Hammond, & Luczak, 2005). A manifest model treats each observed variable as it is found and ignores any error that may exist in how each variable approximates the underlying construct it is meant to represent. When this is done, the estimates of the relationships among the underlying constructs will be attenuated by error and the misspecification of the model can lead to incorrect inferences about critical questions such as effect sizes.

MSEM's latent variable framework is sufficiently flexible as to allow for a number of options for modeling multidimensional constructs. For instance, multiple dimensions of teacher social capital can be allowed to correlate freely with one another. Alternatively, specific factors can be used to account for the unique elements in each dimension, with a general higher order factor loading on each of the specific factors and accounting for the covariation among these related dimensions. A third modeling option is a bifactor/testlet model, which is similar to the higher order model except that in this case the general order factor is specified to load on all of the observed indicators, with specific factors (sometimes called testlet factors) explaining the unique variance shared by specific groups of items. In addition, as will be detailed below, these alternative models can be assessed against one another for optimal fit to the data.

Multilevel framework. The next complexity that commonly confronts us in considering educational data is that it is not enough to look at students' behavior, development and achievement in isolation; we must also consider the impact of the ecosystem surrounding them, the environmental context (Bovaird, 2007). Children who share the same context tend to be more similar than children in other contexts, and what is more, the contexts themselves are often worthy of attention. This is the conceptual rationale for using multilevel modeling. Multilevel modeling (MLM), also termed hierarchical linear modeling, directly accounts for the nested structure that occurs when, for instance, students are clustered within classrooms or schools, or when repeated measures are nested within individuals (Raudenbush & Bryk, 1992). While single-level models rely on an assumption that each individual observation is independent of every other observation, multilevel modeling models the dependencies among observations that belong to the same group and assumes independence across group-level units. Multilevel modeling (MLM) represents a mixed combination of fixed and random effects, with models able to estimate a fixed average intercept along with random individual variability around the group intercept; a fixed average slope representing the average relationship between the predictor and the outcome along with random individual variability around this average relationship.

Morin, Marsh, Nagengast, and Salas (2014) insist that researchers interested in Level 2 constructs such as teaching practice must use statistical models that explicitly acknowledge the clustered nature of the data. Bovaird, among many others, also makes clear the practical/technical justifications for multilevel modeling. On this side, when students are, for instance, clustered within classrooms, then ignoring this clustering in our models results in underestimated standard errors and inflated Type I error rates. If classroom or teacher (Level 2) effects are analyzed at the student level (Level 1), the effects on students and classrooms are confounded and implicitly assumed to be equal, which is often not the case (Morin & Marsh, et al, 2014). On the other hand, when Level 1 data are aggregated to Level 2, important information is lost. Employing the group means results in reduced power, inaccurate representation of group-level relationships and risks the ecological fallacy, in which the analyst makes conclusions about individuals based only on analysis of group-level data. In addition, losing the information on within-group variation may make relations among group-level variables appear stronger than they actually are (Kaplan & Elliot, 1997b). In the context of measurement models, applying single-level analyses to hierarchically structured data can lead to identifying the wrong number of factors or associating items with the wrong factors (Schweig, 2014b). Muthén (1994) details the design effect, the extent to which single-level or simple random sampling assumptions underestimate the true variance of the estimator when the students are in truth sampled from within clusters; the larger the number of students sampled per group and the larger the intraclass correlation coefficient, the measure of the homogeneity of the observations within each cluster, the larger the underestimation of variance that will occur if multilevel techniques are not used.

The concern within the educational literature for using multilevel techniques also applies to the need for multilevel SEM; Muthén (1994) terms the most common perspective “varying parameters” and describes this approach as including any model for relationships among observed Level 1 individuals with parameters that can have different values for different Level 2 groups. He distinguishes this perspective from a “sampling perspective,” which draws on the idea of decomposing variation in student outcomes into a Level 2 school or classroom component and a Level 1 individual component. This sampling perspective also ought to remind us of another reason that multilevel modeling techniques are essential for correct inferences in the field of educational research: the popularity within the field of clustered sampling schemes. For instance, as pointed out by Kaplan and Elliot (1997b), large studies such as the National Educational Longitudinal Study (NELS) and Education Childhood Longitudinal Study (ECLS), and important tests such as the National Assessment of Educational Progress (NAEP) are sampled by selecting schools within districts and states, then classrooms or teachers within schools, and students within these classrooms. In longitudinal studies, researchers often conceive of multiple observations then being nested within each individual student.

Mediation framework. The MSEM can serve as a general, unified modeling paradigm for investigating hypotheses involving mediation in nested data (Preacher, Zyphur, and Zhang, 2010). MSEM subsumes multilevel modeling and two-stage analyses that combine ordinary least squares (OLS) and multilevel models (MLM) as special cases and avoids several important limitations of mediation analysis with MLM.

A basic mediation model hypothesizes that a predictor, X, has an impact on the outcome, Y, at least in part through an intermediate effect on M, a variable that is both affected by X and in turn affects Y.

$$X \rightarrow M \rightarrow Y$$

Because much of educational data is collected from data organized within clusters or hierarchical levels such as students within classrooms, a number of researchers have proposed multilevel modeling strategies for assessing mediation hypotheses within clustered data (see, e.g., Bauer, Preacher, & Gill, 2006; Krull & MacKinnon, 1999, 2001). In the situation considered in this study, I estimate the effect of Teacher Social Capital, X, through Teaching Quality, M, on

Student Math Achievement, Y. Teacher Social Capital is measured at Level 2, through survey responses from individual teachers, while Teaching Quality and Student Math Achievement are measured at Level 1, through student survey responses and student test scores respectively. Based on the level of the hierarchy from which data is collected for each of these constructs, this is considered a 2 – 1 – 1 design in the notation proposed by Krull and MacKinnon (1999). In this multilevel context, then, the mediation design is subscripted, with i indexing individuals and j indexing groups:

$$X_j \rightarrow M_{ij} \rightarrow Y_{ij}$$

Multilevel mediation models properly account for the dependencies among these student observations clustered within teachers and are thus preferable to methods based on assumptions of simple random sampling. However, these multilevel modeling approaches have a number of limitations for assessing mediation hypotheses, as outlined by Preacher, Zyhpur, and Zhang (2010). First, for 2 – 1 – 1 and 1 – 1 – 1 designs, the multilevel model (MLM) produces a conflated or biased estimate of the indirect effect (see Figure 4.6). When a traditional MLM approach is used to assess mediation in this design, the Within and Between components of the effect of M on Y are conflated into a weighted average. Therefore, these approaches assume the contextual effect is zero and produce a biased estimate of the Between indirect effect if the Within and Between effects differ in reality (Preacher Zhang, Zyphur, 2011). Second, MLM cannot be used for other designs that include a Level 2 variable as a dependent variable, i.e., at the second or third step, such as 1 – 1 – 2 or 2 – 2 – 1 designs (Preacher et al, 2010). MSEM, on the other hand, is flexible enough to accommodate variables assessed at Level 1 or Level 2 at any point in the mediation, thus allowing researchers to model 1 – 1 – 2, 2 – 2 – 1, 2 – 1 – 2, 1 – 2 – 2, or 1 – 2 – 1 designs. In a simulation study, MSEM dramatically reduced bias in estimating

Between indirect effects relative to MLM-based procedures; outperformed the two MLM approaches in confidence interval coverage under every condition of number of clusters, sample size within clusters, and ICC; had adequate power to detect a small (0.1) indirect effect for ICCs ≥ 0.2; showed excellent convergence; and underperformed only in terms of efficiency at low ICC (Preacher et al, 2011).

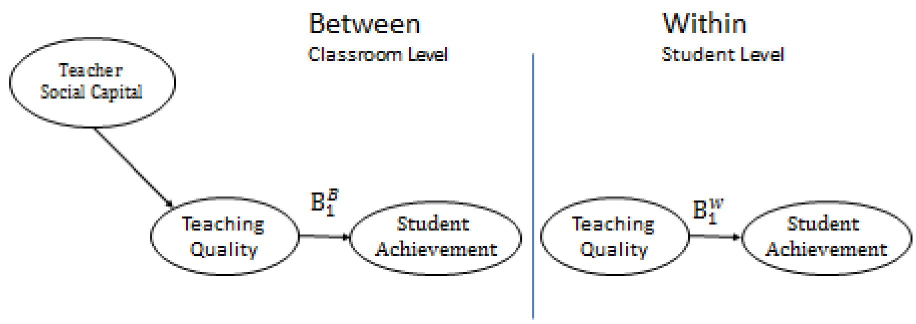

Figure 4.6. MSEM sets up different Between and Within models, allowing the Between and Within effects to automatically be unconflated. This means the Between cluster indirect effect is unbiased. In typical MLM mediation, the estimation of a 1 – 1 path conflates Within- and Between-cluster effects, leading to bias (under- or overestimation) in the indirect effect.

In sum, MSEM is a flexible modeling framework that encompasses more traditionally-utilized modeling frameworks such as multiple regression, path analysis, and multilevel modeling as special cases. Therefore, this framework can be applied to a wide variety of education research problems and the results found in these analyses can be readily compared to future investigations of related problems.

General Statistical Form of MSEM

Multilevel structural equation modeling (see e.g., Kaplan & Elliot, 1997; Muthén, 1989; 1994; Preacher, Zyphur, & Zhang, 2010) includes all the modeling capability of single-level SEM and adds the ability to properly account for nested data. MSEM allows researchers to 1) estimate latent constructs, thereby including estimates of measurement error as well as enabling direct modeling of multidimensional constructs, 2) incorporate consideration of the group context, thereby properly estimating the standard errors based on the dependencies among group members and separating between-group effects from within-group effects, and 3) directly model complex mediated causal pathways. To illustrate the MSEM framework, it helps to first examine the single-level structural equation modeling framework (Joreskog, 1977).

Single-Level SEM

Data Structure Model. Consider a situation such as measuring math skill with a test of 20 items, measuring teaching practice using a student survey with 30 questions, or measuring teaching practice using an observation framework which a rater uses to assign scores on 7 different rubrics. Each item is a unique measure of the underlying construct. Following the notation of Preacher et al (2010) and Muthén and Asparouhov (2008), the measurement model specifies *m* latent variables, *p* items (in our examples, *p* = 20, 30, and 7, respectively), *q* exogenous covariates, and N individuals or cases (*i* is the indicator for each individual case, *i* = 1…, N). The measurement model can be written as

$$\boldsymbol{Y}_i = \boldsymbol{v} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{K}\boldsymbol{X}_i + \boldsymbol{\epsilon}_i \qquad (1)$$

where $\boldsymbol{Y}_i$ is a vector of measured outcome variables; $\boldsymbol{v}$ is a vector of the item variable means/intercepts, of length *p;* $\boldsymbol{\eta}_i$ is a vector of latent variables, also called random effects, of size *m* x 1; $\boldsymbol{\Lambda}$ is a matrix of the factor loadings that indicate the strength of the association between each observed predictor and the underlying latent variable; $\mathbf{X}_i$ is a vector of exogenous covariates, i.e., there are no other variables predicting them, and $\boldsymbol{K}$ is a matrix of slope coefficients for the covariates.

The structural model can be written as

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i \qquad (2)$$

where $\boldsymbol{\alpha}$ is a vector of intercepts/means of the latent variables of dimensions *m* x 1; $\boldsymbol{B}$ is a matrix of structural regression coefficients of dimensions *m* x *m*; $\boldsymbol{\Gamma}$ is a matrix of slopes for the relationships of the exogenous covariates with the latent variables of dimensions *m* x *q*; $\boldsymbol{\zeta}$ is a vector of the latent variable residuals. The residuals $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$ are distributed multivariate normally with mean and covariance (0, $\boldsymbol{\Theta}$) and (0, $\boldsymbol{\Psi}$), respectively.

Covariance Structure Model. In SEM, the covariance structure model is our hypothesis of the covariance structure and this, rather than the data model, is the model that is estimated. This model accounts for the variances/covariances of the measurement model as well as those of the structural regression elements. $\boldsymbol{\Sigma}$ is the variance / covariance matrix of the vector of observed $\boldsymbol{Y}_i$. The model-implied covariance structure is written as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}\ [\boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}]' + \boldsymbol{\Theta} \quad (3)$$

Equation 3 tells us that the variances and covariances that we observe (in $\boldsymbol{\Sigma}$) are a function of the common factor loadings $\boldsymbol{\Lambda}$, the variances and covariances of the latent factors contained in the common factor covariance matrix $\boldsymbol{\Psi}$ (the latent factors, the $\boldsymbol{\eta}$'s, are not part of this model, only their variances and covariances), the structural regression coefficients in $\boldsymbol{B}$, and the variances and covariances of the residuals (unique variances) contained in $\boldsymbol{\Theta}$. $\mathbf{I}$ is an identity matrix. The errors $\boldsymbol{\zeta}_i$ and $\boldsymbol{\epsilon}_i$ are assumed uncorrelated, and the variance of $\boldsymbol{\zeta}_i = \boldsymbol{\Psi}$ and variance of $\boldsymbol{\epsilon}_i = \boldsymbol{\Theta}$:

Multilevel SEM

Data Structure Model. This SEM model can be expanded to account for nested data, i.e., become a multilevel SEM, by allowing the matrices of the pertinent coefficients to vary by group. This means that, for a two-level model, each variance is separated into two components, one representing the variance between groups, and the other representing how the individuals within each group deviate from that group. In the model equation, this is shown by adding a *j* subscript to the relevant coefficients to indicate the group or cluster. Equation 4 is equivalent to equation 1 except for the addition of *j* group-member subscripts for $\boldsymbol{Y}_i$, $\boldsymbol{v}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\eta}_i$, $\boldsymbol{K}$, $\boldsymbol{X}_i$, and $\boldsymbol{\epsilon}_i$. The *j* subscript allows some or all of the elements of these matrices to vary by group. The measurement model of the multilevel SEM is

$$\boldsymbol{Y}_{ij} = \boldsymbol{v}_j + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_{ij} + \boldsymbol{K}_j\boldsymbol{X}_{ij} + \boldsymbol{\epsilon}_{ij} \qquad (4)$$

If our outcome, $\boldsymbol{Y}_{ij}$, is measured on *p* indicators, say survey items, the matrices $\boldsymbol{Y}_{ij}$, $\boldsymbol{v}_j$ and $\boldsymbol{\epsilon}_i$ will be *p*-dimensional. $\boldsymbol{\Lambda}_j$ is still *p* x *m*, but in the multilevel case, *m*, the number of latent variables, includes latent variables at both the Within and Between levels, and may also include random slopes. $\boldsymbol{\eta}_{ij}$ remains an *m* x 1 vector of the latent variables and $\boldsymbol{K}_j$ is *p* x *q* matrix of the coefficients (slopes) representing the extent to which the outcome is associated with each of the exogenous covariates in $\boldsymbol{X}_{ij}$. $\boldsymbol{X}_{ij}$ is itself a vector of *q* length that contains the exogenous covariates. The Level-1 structural model of the multilevel SEM is

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\mu}_j + \boldsymbol{B}_j\boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j\mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij} \qquad (5)$$

with $\boldsymbol{\mu}_j$ consisting of the *m* intercepts/means of the latent variables arranged in a vector; $\boldsymbol{B}_j$ is a matrix of *m* x *m* dimensions containing the structural regression coefficients; $\boldsymbol{\Gamma}_j$ is of dimensions *m* x *q* to contain the slope parameters for the *q* exogenous covariates in the vector $\mathbf{X}_{ij}$. The *j* subscript indicates that these parameters are random and vary across groups. In the specific model for this study, the subscripts for $\boldsymbol{B}_j$ and $\boldsymbol{\Gamma}_j$ will be dropped and these parameters will be fixed across classrooms. The residuals $\boldsymbol{\epsilon}_{ij}$ and $\boldsymbol{\zeta}_{ij}$ are assumed distributed as in the single-level model in equations 1 and 2, multivariate normal (0, $\boldsymbol{\Theta}$) and (0, $\boldsymbol{\Psi}$), respectively. $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ do not vary across groups, so there is no *j* subscript for these matrices.

To represent the fact that the parameters vary at the Between level (i.e., from one classroom to the next) the Level-2 structural model can be written as

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\gamma}\mathbf{X}_j + \boldsymbol{\zeta}_j \tag{6}$$

The $\boldsymbol{\eta}_j$ vector is a column of all the random effects that potentially vary at Level-2 – these are all the elements of the parameter matrices $\boldsymbol{\alpha}_j$ and $\boldsymbol{B}_j$ that might vary at the group level. $\mathbf{X}_j$ is a vector of the group-level covariates. It is of length *s*, with *s* being the number of Level-2 exogenous covariates. $\boldsymbol{\mu}$ is a vector of dimensions *r* x 1 containing estimated fixed effects: the intercepts of the Between structural equations; $\boldsymbol{\beta}$ is an *r* x *r* matrix of fixed effects equal to the regression slopes of the latent variables regressed on one another. $\boldsymbol{\gamma}$ contains, in an *r* x *s* matrix, the regression slopes of the latent variables (random effects) regressed on the exogenous group-level covariates in $\mathbf{X}_j$. The residuals in $\boldsymbol{\zeta}_j$ are distributed multivariate normal with mean 0 and covariance matrix $\boldsymbol{\Psi}$.

Covariance Structure Model. The multilevel SEM separates the covariance structure into within and between components.

$$\boldsymbol{\Sigma}_{\mathrm{W}} = \boldsymbol{\Lambda}_{\mathrm{W}}(\mathbf{I} - \boldsymbol{B})_{\mathrm{W}}^{-1}\boldsymbol{\Psi}_{\mathrm{W}}\,[\boldsymbol{\Lambda}_{\mathrm{W}}(\mathbf{I} - \boldsymbol{B})_{\mathrm{W}}^{-1}]' + \boldsymbol{\Theta}_{\mathrm{W}} \tag{7}$$

$$\boldsymbol{\Sigma}_{\mathrm{B}} = \boldsymbol{\Lambda}_{\mathrm{B}}(\mathbf{I} - \boldsymbol{B})_{\mathrm{B}}^{-1}\boldsymbol{\Psi}_{\mathrm{B}}\,[\boldsymbol{\Lambda}_{\mathrm{B}}(\mathbf{I} - \boldsymbol{B})_{\mathrm{B}}^{-1}]' + \boldsymbol{\Theta}_{\mathrm{B}} \tag{8}$$

Once again, the variances and covariances in $\boldsymbol{\Sigma}$ are modeled as a function of the common factor loadings $\boldsymbol{\Lambda}$, the variances and covariances of the latent factors in the common factor covariance matrix $\boldsymbol{\Psi}$, the structural regression coefficients in $\boldsymbol{B}$, and the variances and covariances of the unique variances in $\boldsymbol{\Theta}$. The errors $\boldsymbol{\zeta}_i$ and $\boldsymbol{\epsilon}_i$ are assumed uncorrelated, and the variance of $\boldsymbol{\zeta}_i = \boldsymbol{\Psi}$ and variance of $\boldsymbol{\epsilon}_i = \boldsymbol{\Theta}$. This structure is identical to that of the single-level SEM except that it is separated into Within and Between levels.

Conclusion

Consistent findings have suggested the importance of teacher social capital in determining the quality of teaching and learning. It now makes sense to probe those relationships in depth, using a variety of sophisticated methodologies and conceptual approaches, to try to produce consistent results about the mechanisms by which these relationships operate. This dissertation examines these relationships using multilevel structural equation modeling, and compares the results to those found using alternative models – models that employ manifest composite indicators or that ignore the multilevel structure of the data.

Chapter V. Methods

This chapter details the methods used in seeking to answer the following research questions:

1. What factors of teacher social capital can be identified using the Teacher Working Conditions Survey?
2. What factors of teaching practice can be identified using the Tripod student perception survey items? How do these factors compare to those identified in the Framework for Teaching observation ratings?
3. What are the relationships among teacher social capital, teaching practice and student achievement?
   a. How do the relationships among these constructs differ when estimated with a multilevel structural equation model versus a single-level structural equation model?
   b. How do the relationships among these constructs differ when estimated with a multilevel structural equation model that uses latent factors to represent the social capital and teaching practice constructs versus a model that uses manifest (observed) indicators?
   c. How do the relationships among the constructs differ when teaching practice is measured by expert observation ratings rather than students' survey responses?

Sample and Data Preparation

The Measures of Effective Teaching (MET) Project, a research partnership funded by the Bill & Melinda Gates Foundation, collected a variety of indicators of teaching practice, measures of the school context, student achievement test scores in mathematics and English-Language Arts on both state-mandated standardized tests and an alternative standardized test, a test of teachers' mathematical content and instructional knowledge, and student and teacher background variables. The dataset includes more than 160,000 students, within the classrooms of 2,741 teachers in grades four through nine working in 317 schools, located in six large school districts in the United States during the 2009–10 and 2010-11 school years.

Because of the recruitment of volunteer teacher participants, along with other selection decisions, the MET User's Guide makes clear that the MET sample is not representative of any identified universe of school districts, nor of teachers in any district. Data from MET research reports suggest that, across a range of observable characteristics, including ethnicity, years of teaching experience, and value-added scores, teachers who participated in the MET Study were largely similar to teachers in the same districts who did not participate (Kane & Staiger, 2012). This study focuses on mathematics classrooms in grades 6 – 8 in year one of the MET Study. I focus on mathematics classrooms because research has consistently shown larger teacher effects and greater variation in teachers' effectiveness in math than in reading (Hanushek, Kain, & Rivkin, 1998; Harris & Sass, 2011; Kraft & Papay, 2014). Additionally, prior research using the MET data found that teacher effect estimates in ELA were smaller and less precisely estimated than the effects in math classrooms (Kane, McCaffrey, Miller, and Staiger, 2013). I focus on grades 6 – 8 in part because older students may be better able than elementary students to respond reliably and rate their teachers on the criteria most relevant to improving student learning (Peterson et al, 2000).

The analyzed sample with responses on the Tripod student perception survey includes 95 schools, 520 teachers, and 16,922 students; only 12,888 students and 382 teachers at 90 schools have responses to the Teacher Working Conditions Survey (see Tables 5.1 and 5.2). The measurement models for teaching practices are based on the larger analytic sample while the measurement models for teacher social capital are based on the reduced analytic sample with teacher responses to the Teacher Working Conditions Survey. The majority of teachers in the sample are White and female, though these proportions of White and female teachers are somewhat smaller than the nation's teacher workforce as a whole, which in 2011-12 was approximately 82 percent White and 76 percent female (US DOE, 2013). The structural effects are estimated using the full sample after cases with missing on all x-variables and cases with missing on the outcome are excluded, resulting in an analyzed sample of 15,644 students in 520 classrooms for the multilevel and single-level latent structural equation models, 15,628 students in 520 classrooms for the manifest model, and 15,614 students in 520 classrooms when the model is re-estimated using the Framework For Teaching observation ratings in place of the Tripod survey item responses. The sample characteristics do not differ appreciably whether considering the Tripod, FFT, or Teacher Working Conditions samples, nor when comparing across the latent and manifest model samples.

As shown in Table 5.2, the sample with responses to the Tripod includes close to one-third Black, one-third Hispanic, and one-third White students along about 7 percent Asian students, and about sixty percent of students receive free or reduced priced meals. The data includes state math test outcome scores for over 97 percent of the students, but only about 83 percent of students have scores on the Balanced Assessment of Mathematics, an alternative math outcome measure intended to measure problem-solving and conceptual understanding.

Table 5.1

School Sample Characteristics

School Characteristic	
Number of Schools	95
Number of Teachers	520
mean N of teachers per school	5.41
SD of teachers per school	2.91
minimum N of teachers in one school	1
maximum N of teachers in one school	17
N of schools w 3 or fewer Teachers	34
mean # of students per school	178.13
SD of students per school	107.23
minimum N of students in one school	11
maximum N of students in one school	479
N of schools w 100 or fewer Students	28

Table 5.2

Sample Characteristics: Teachers and Students

	Sample with Responses to TWCS			**Sample of all Middle School Math Teachers**		
	Obs	**Mean / %**	**Std. Dev**	**Obs**	**Mean / %**	**Std. Dev.**
Teacher Characteristics	382			520		
Male	372	0.27	0.45	507	0.29	0.45
Female	372	0.73	0.45	507	0.71	0.45
White	370	0.57	0.5	505	0.51	0.5
Black	370	0.33	0.47	505	0.38	0.48
Hispanic	370	0.06	0.23	505	0.06	0.24
Other	370	0.04	0.2	505	0.05	0.23
Master's Degree	303	0.28	0.45	407	0.28	0.45
Content Knowledge Test Score	282	-0.02	1	354	-0.03	1.02
Years Exp from WCS	381	3.78	1.4	382	3.78	1.4
Years in District from WCS	374	2.9	1.23	375	2.9	1.23
Student Characteristics	12,888			16,922		
Special Education	12,802	0.08	0.26	16,798	0.08	0.26
English Learners	12,825	0.13	0.34	16,840	0.14	0.35

Free or Red. Meals	10,428	0.59	0.49	13,656	0.61	0.49
White	12,825	0.30	0.46	16,840	0.27	0.44
Black	12,825	0.28	0.45	16,840	0.29	0.45
Hispanic	12,825	0.33	0.47	16,840	0.35	0.48
Asian	12,825	0.06	0.25	16,840	0.07	0.25
Other	12,825	0.03	0.16	16,840	0.02	0.15
ELA 2009	11,943	0.13	0.92	15,592	0.11	0.93
Math 2009	11,980	0.15	0.93	15,647	0.12	0.94
ELA 2010	12,523	0.14	0.92	16,353	0.12	0.93
Math 2010	12,553	0.16	0.93	16,394	0.15	0.94
BAM 2010	10,637	0.09	0.97	13,735	0.04	0.98

State math tests and BAM scores standardized by MET research team on the full MET sample for each district.

Table 5.3

Teachers' Years of Experience (Detail)*

Category	**Obs**	**%**
First Year (from WCS)	18	5%
2-3 Years (from WCS)	61	16%
4-6 years (from WCS)	83	22%
7-10 years (from WCS)	89	23%
11 + years (from WCS)	130	34%

**Sample with Responses to TWCS - no information available on full sample of Middle School Math Teachers*

Measures

Teacher Working Conditions Survey

Teacher social capital is measured by items on the Teacher Working Conditions Survey (TWCS), related to the five dimensions of social capital outlined in Chapter II, networks, norms, distributed leadership, access to expertise, and depth of interactions. The items are displayed in Table 5.4, grouped according to the conceptual map presented in Appendix A. Most items were four-point likert-type scales ranging from “Strongly Disagree” (1), to “Strongly Agree” (4). Four items employed the options “No Role”, “Some Role”, “A Moderate Role”, and “A Large Role”. For all items the “Don’t Know” option was recoded as missing.

Table 5.4

TWCS item means and standard deviations: Measuring social capital dimensions

	Obs	Mean	SD
Norms: Trust			
Teachers are recognized as educational experts	381	2.94	0.75
Teachers are trusted to make sound professional decisions about instruction	381	2.99	0.72
The faculty has an effective process for making group decisions to solve problems	381	2.78	0.91
In this school we take steps to solve problems	380	2.91	0.82
Teachers are effective leaders in this school	381	2.96	0.77
The faculty and leadership have a shared vision	381	2.95	0.78
There is an atmosphere of trust and mutual respect in this school	382	2.69	0.88
Teachers feel comfortable raising issues and concerns that are important to them	381	2.69	0.88
The school leadership consistently supports teachers	382	2.82	0.87
Norms: Caring:			
Leadership makes a sustained effort to address teacher concerns re:			
Leadership issues	380	2.95	1.03
The use of time in my school	380	2.89	0.92
Professional development	377	3.02	0.87
Teacher leadership	380	3.01	0.91
Instructional practices and support	381	3.02	0.80
T-P Networks (strength of teacher-parent networks)			
Parents/guardians are influential decision makers in this school	381	2.71	1.01
school maintains clear, two-way communication w/ parents/guardians & community	378	3.03	0.82
This school does a good job of encouraging parent/guardian	381	3.06	0.77
Teachers provide parents/guardians with useful information about	380	3.23	0.61
Parents/guardians know what is going on in this school	381	2.93	0.86
Parents/guardians support teachers, contributing to their success	379	2.61	0.80
Community members support teachers, contributing to their	380	2.86	0.98
The community we serve is supportive of this school	382	2.99	0.97
T-T Networks: Time For Collaboration (potential for t-t networks to arise)			
Teachers have time available to collaborate with colleagues	378	2.94	0.77
The non-instructional time provided for teachers in my school is sufficient	379	2.54	0.83
Collaborative planning time (in categories of number of hours)	380	2.61	0.84
An appropriate amount of time is provided for professional development	379	2.83	0.70

Distributed Leadership:			
Indicate the role teachers have in the following areas:			
Selecting instructional materials and resources	379	2.83	0.95
Devising teaching techniques	377	3.33	0.85
Setting grading and student assessment practices	379	3.21	0.87
School improvement planning	377	2.73	1.01
Access To Expertise (colleagues, support personnel, administration)			
Teachers have sufficient access to a broad range of professional support personnel	379	2.90	0.74
In this school, follow up is provided from professional development	380	2.79	0.94
Sufficient resources are available for professional development	381	2.93	0.75
Professional development deepens teachers' content knowledge	378	2.83	0.84
Provided supports (coaching, PLCs, etc.) translate to improvements in instruction	380	3.12	0.85
Teachers receive feedback that can help them improve teaching	379	2.92	0.83
Depth Of Interactions			
PD provides ongoing opportunities to work with colleagues to refine teaching	380	2.91	0.79
Teachers work in professional learning communities to develop and	378	3.14	0.77
I frequently plan and coordinate instruction with my students' other	379	2.77	0.85
I collaborate with other teachers to achieve consistency on how	381	3.08	0.78

Table 5.4 shows that the items collectively cover the construct of social capital appropriately, without apparent gaps. Each hypothesized dimension is represented by four or more items that characterize the main aspects of their respective dimension. Additional factor analyses investigating the extent to which the items in fact cover the hypothesized structure of social capital are described in the analytic section.

Tripod Classroom Environment Survey

The Tripod Student Survey was developed by Ron Ferguson at Harvard University, and is based upon classroom-level surveys developed by the Tripod Project for School Improvement (Ferguson, 2008). The Tripod measures student perceptions on seven theoretical dimensions (Seven C's) of instructional quality, namely Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate. *Care* is intended to measure students' perceptions of whether they feel encouraged and supported. *Control* measures student behavior in terms of respect and being consistently on-task. *Clarify* measures students' perceptions of teacher behaviors that help students' to better understand the content being taught. *Challenge* measures students' perceptions of classroom rigor and of teachers' push for persistence. *Captivate* measures students' perceptions of how well the teacher captures the attention and interest of students and how relevant the learning appears. *Confer* measures students' perceptions of how much a teacher respects and takes into account students' points of view when teaching. *Consolidate* measures students' perceptions of the extent to which teachers help students connect ideas to one another and integrate different curriculum topics. These seven dimensions are further grouped into two overarching factors of teaching practice: *Press,* which includes Control and Challenge, and *Support*, which includes the remaining five C's (see Chapter II for additional discussion of the conceptual framework underlying the Tripod Survey). *Press* is conceptualized as representing academic press – the extent to which students are on-task, behaving in strict accord with the rules, and being challenged with difficult and complex work. *Support* represents the extent to which the teacher supports student learning emotionally and academically with practices such as helpful feedback, engaging and clear content delivery, and caring interactions. In all, the student perception survey asked secondary students (grades 6-9) to rate 58 items on classroom environment and teacher's practices, 36 of which were considered as part of the Tripod's Seven C's. All items were rated on a five-point Likert-type response scale with response options *Totally Untrue; Mostly Untrue; Somewhat; Mostly True; and Totally True.* The items are displayed in Table 5.4, grouped by dimension, along with the means and standard deviations estimated for this analytic sample.

Table 5.5

Tripod Survey: Item means and standard deviations

Var.	Item Wording	Obs	Mean	SD
	Care			
A10	My teacher in this class makes me feel that he/she really cares about me	16,557	3.64	1.25
B146	My teacher seems to know if something is bothering me.	16,171	3.03	1.32
B34	My teacher really tries to understand how students feel about things.	16,308	3.45	1.23
	Control			
B112	Student behavior in this class is under control.	16,383	3.33	1.27
B113	I hate the way that students behave in this class.*	16,336	2.54	1.35
B114	Student behavior in this class makes the teacher angry.*	16,171	3.10	1.33
B138	Student behavior in this class is a problem.*	16,455	2.74	1.31
B46	My classmates behave the way my teacher wants them to	16,374	3.07	1.23
B49	Students in this class treat the teacher with respect.	16,647	3.53	1.18
B6	Our class stays busy and doesn't waste time	16,612	3.48	1.19
	Clarify			
B130	My teacher knows when the class understands, and when we do not.	16,586	3.84	1.11
B136	When s/he is teaching, my teacher thinks we understand even when we don't.*	16,380	2.47	1.25
B1	If you don't understand something, my teacher explains it another way.	16,600	4.04	1.08
B17	My teacher has several good ways to explain each topic that we cover in this class.	16,525	3.90	1.10
B80	My teacher explains difficult things clearly.	16,216	3.83	1.13
	Challenge			
B128	My teacher asks questions to be sure we are following along when s/he is teaching.	16,796	4.37	0.94
B133	My teacher asks students to explain more about the answers they give.	16,606	4.14	0.97
B21	In this class, my teacher accepts nothing less than our full effort.	16,077	3.98	1.08
B36	My teacher doesn't let people give up when the work gets hard.	16,567	4.02	1.09
B45	My teacher wants us to use our thinking skills, not just memorize things.	16,360	4.08	1.04
B59	My teacher wants me to explain my answers – why I think what I think	16,528	4.05	1.04
B70	In this class, we learn a lot almost every day	16,356	3.97	1.06
B90	In this class, we learn to correct our mistakes	16,188	4.05	1.04
	Captivate			
B141	This class does not keep my attention – I get bored.*	16,370	2.66	1.36
B29	My teacher makes learning enjoyable.	16,578	3.49	1.29
B44	My teacher makes lessons interesting.	16,375	3.48	1.26
B89	I like the ways we learn in this class.	16,893	3.82	1.03
	Confer			
B129	My teacher wants us to share our thoughts	16,705	3.63	1.21
B135	**Students get to decide how activities are done in this class.**	16,777	2.32	1.07
B154	My teacher gives us time to explain our ideas.	16,104	3.64	1.14
B155	Students speak up and share their ideas about class work.	16,166	3.52	1.19
A54	My teacher respects my ideas and suggestions.	16,127	3.67	1.15
	Consolidate			
B145	My teacher takes the time to summarize what we learn each day	16,259	3.50	1.23
B147	My teacher checks to make sure we understand what s/he is teaching us.	16,280	4.08	1.07
B58	We get helpful comments to let us know what we did wrong on assignments	16,401	3.67	1.20
B83	The comments that I get on my work in this class help me understand how to improve	16,419	3.68	1.17

* Indicates item is reverse coded; reported means are prior to reverse coding

Framework For Teaching

Teaching practice will also be measured through the FFT observation instrument. As implemented in the MET study, the FFT has four component measures used to assess each of two dimensions of teaching practice – *Classroom Managemen*t and *Instructional Support*, as shown in Table 5.6 (Mihaly et al, 2013). Raters watched 15 minutes at the beginning of the video, watched an additional ten minutes from the 25-35 minute mark, and then scored the video as a single segment. Each component was scored by raters on a four-point scale: unsatisfactory, basic, proficient, distinguished. Each of these component measures can be treated similarly to an item on a survey or a test. Each teacher was expected to turn in four videos of classroom practice (a small number of teachers in the analytic sample turned in fewer than four videos and thus have missing data), and these videos were then rated by one or two raters. Ratings on each FFT domain were averaged across raters to provide an average teacher-level rating on each of the eight domains.

Table 5.6

Components Rated on Framework For Teaching

Domain and Title	**Component**
Classroom Environment	
Respect	Creating an Environment of Respect and Rapport
Culture	Establishing a Culture for Learning
Manage Procedures	Managing Classroom Procedures
Manage Behavior	Managing Student Behavior
Physical Space (not rated in MET)	Organizing Physical Space
Instructional Support	
Communicate	Communicating with Students
Question	Using Questioning and Discussion Techniques
Engage	Engaging Students in Learning
Assess	Using Assessment in Instruction
Flexibility (not rated in MET)	Demonstrating Flexibility and Responsiveness

Student Achievement and Covariates

The models described in this dissertation are estimated using the state standards-based math tests as the proxy for student math achievement. Each district assesses students using a separate state standardized test. This means that estimates of student achievement using the state tests are not directly comparable. The MET research team created a standardized state math test variable, standardized on the full sample for each district, and I further address this issue through including district fixed effects to account for idiosyncratic differences across districts. All student test outcome variables are standardized (mean=0, SD = 1) to ease interpretation of results.

The model displayed in Figure 5.1 includes district fixed effects to account for idiosyncratic but systematic differences from one district to the next, and school racial composition (School percent Black and Hispanic) effects on teacher social capital factors; in addition, teaching practice factors are regressed on classroom average prior math achievement and classroom percent Black and Hispanic. Student math achievement is also regressed at Level 2 on these district fixed effects, the composition covariates (School percent Black and Hispanic, Classroom Average Prior Math Achievement, Classroom Percent Black and Hispanic), and at Level 1 on students' individual prior math scores, grade level, gender, race, and English learner status. See Table 5.2 above for descriptive information regarding these variables.

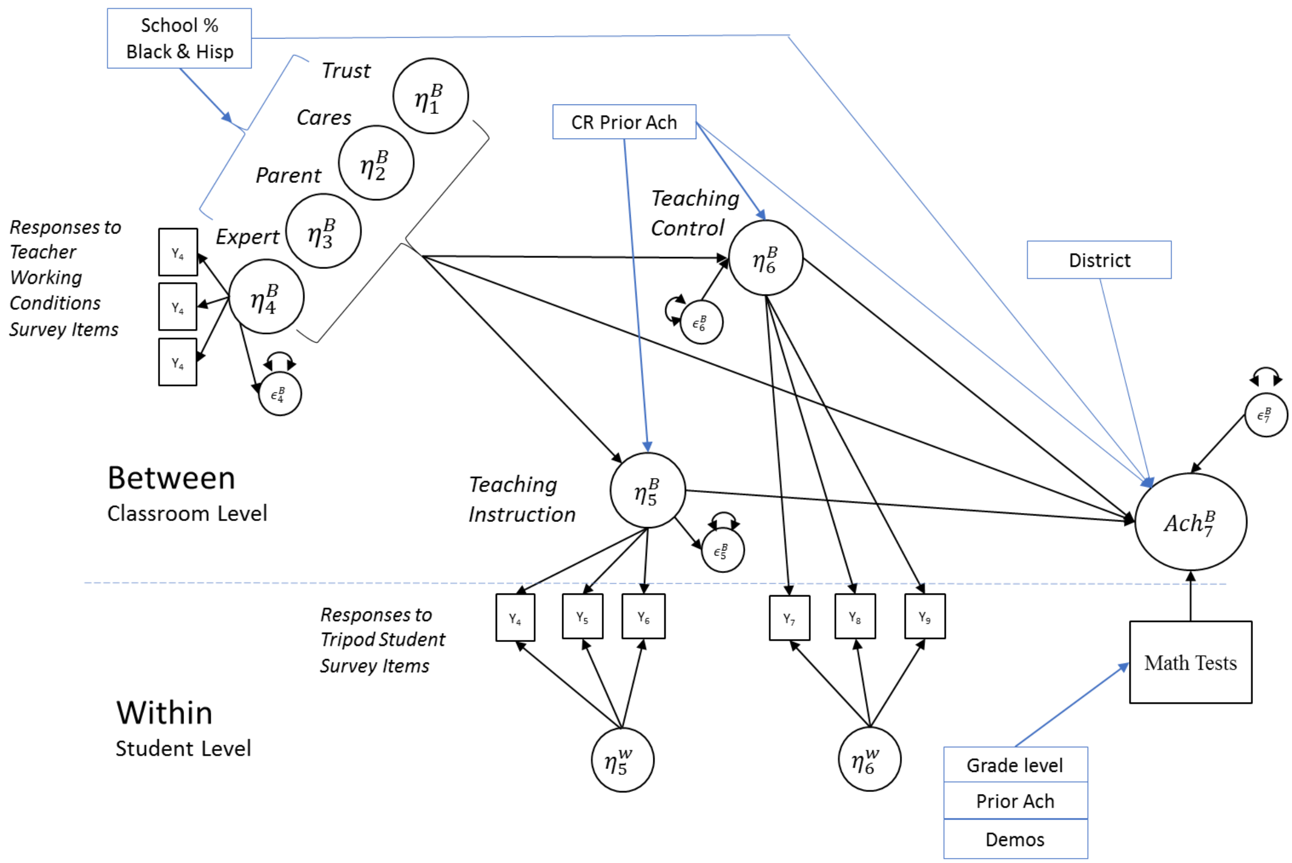

Figure 5.1. Multilevel Structural Equation Model. The 4 Teacher Social Capital factors are correlated, as are the 2 Teaching Practice factors. Model estimated using state math tests as the outcome.

Analytic Plan

Data exploration

I conducted data exploration, descriptive analyses, selection of the analytic sample, and merging of the data files in Stata 13. For all the variables theorized to represent the latent constructs of interest, this included exploration of means, standard deviations, frequency tables, Pearson's correlations, and Cronbach's alpha for the proposed factors. In addition, polychoric correlation matrices were examined for the categorical survey response variables. Polychoric correlations use the information from the observed ordinal variables to estimate the correlation between the theorized latent variables that underlie the observed responses. I estimate all measurement and structural models in Mplus 7.4. With the exception of the preliminary measurement analyses noted above which treated survey response variables as categorical variables, all analyses use the robust Huber-White standard error adjustment (i.e., the MLR estimator – maximum likelihood robust: Muthén & Muthén, 1988-2015). Model equations and Mplus syntax are provided in Appendix D.

Research question 1: the measurement model for teacher social capital. I conducted preliminary exploration of the correlations among items to refine the hypothesized factor structure, then estimated a series of single- and multi-level confirmatory factor analyses in Mplus 7.4, initially using a robust weighted least squares estimator, WLSMV, and analyzing a polychoric correlation matrix (Muthén & Muthén, 1988-2015) to account for the categorical data structure of the survey response variables but in later models treating the survey items as approximately continuous. Details on these analyses and results are reported in Chapter VI.

Research question 2: The measurement model for teaching practice. Initial analyses using the Tripod drew on the operationalization of teaching practice outlined by Ferguson and Danielson (2014), which proposes two dimensions of teaching practice: *Press*, consisting of Control and Challenge, and *Support*, consisting of the other five C's. Conceptual analysis, multilevel confirmatory factor analysis results, and the work of Kuhfeld (2016) led to a restructuring based on a reduced set of items. The final structure consists of two dimensions of teaching practice, with 5 CONTROL items forming one dimension, and 14 items from the other six C's forming a general INSTRUCTION dimension.

I also conducted exploratory and confirmatory factor analyses to arrive at a two-factor model for the Framework For Teaching, with the factors similar to those outlined by (Mihaly et al, 2013) and named CLASSROOM MANAGEMENT and INSTRUCTIONAL SUPPORT. Classroom Management consists of the domains of *Creating an environment of respect and rapport*¸ *Managing classroom procedures,* and *Managing student behavior*. Instructional Support is made up of *Communicating with students, Engaging students in learning, Using questioning and discussion techniques, Using assessments in instruction, and Establishing a culture of learning*.

Research Question 3: In estimating the structural model, I seek to understand to what extent teaching practice mediates the relationship between teacher social capital and student achievement. In addition, the structural relationships provide evidence of the validity of these conceptualizations of teacher social capital and teaching practice by estimating the relationships between these theoretically related constructs. Chapter VII examines research question 3: What are the relationships among teacher social capital, teaching practice and student achievement? The focus here is on discussing the substantive findings from the multilevel multidimensional structural equation mediation model. Figure 5.1 shows the structural model, with each of the latent factors depicted in the large circles and defined by the measurement models discussed in chapter VI.

To address question 3a, the structural relationships are re-examined in chapter VIII by comparing the results from the multilevel multidimensional structural equation mediation model to an alternative model that ignores the clustered nature of the data. Research question 3a considers how the relationships among these constructs change when estimated with a multilevel structural equation model versus a single-level structural equation model. In the single-level model, the students' nesting within teachers is ignored and the Between and Within effects are conflated. The procedure for estimating this model follows straightforwardly from the above description of the multilevel SEM – the only change needed in estimating this model in Mplus 7.4 is to remove the syntax identifying the analysis as multilevel and identifying level 1 variables as Within and level 2 variables as Between. The comparison presented here pertains to one of the central methodological considerations in this dissertation: to what extent are inferences about the structural relationships among these constructs dependent on modeling assumptions regarding the clustering of data?

The multilevel structural equation model separates the model into Between and Within portions, with the Between portion of the model including all school and classroom level variables and factors and the estimated relationships among the Between factors. In this MSEM Model, at Level 2, the MATH10 outcome is regressed on teacher social capital factors, teaching practice factors, School percent Black and Hispanic, classroom prior math, and district fixed effects. The Within level includes only the measurement models for CONTROL and INSTRUCTION Within classrooms, the Within Classroom correlation between CONTROL and INSTRUCTION, the Within variance of MATH10, and the effects of student level covariates on the MATH10 outcome, namely: individual prior math scores, grade level, gender, race, and English learner status. The single-level structural equation model includes the same factors and covariates but treats all variables and factors at the same level.

Next, in chapter IX, research question 3b is addressed and the results from the latent model are compared to results from a manifest model that ignores the measurement error in the indicators. The manifest MSEM model employs composite indicators in place of each teacher social capital and teaching practice factor. The composite indicators are formed from a simple average of all the items found to load on each factor in the confirmatory factor analyses discussed in Chapter VI. The structural model is re-estimated using those composite variables in place of the latent factors, omitting the measurement portion of the model. Measurement always involves error, and the comparison across the latent and manifest models provides insight into the extent to which this measurement error impacts the inferences drawn about the relationships among constructs.

In both models, there are four teacher social capital factors/manifest composites at the Between teacher level: TRUST (or TRUSTMAN): extent to which teachers trust one another and feel trusted; CARES (or CARESMAN): extent to which teachers perceive the administration as responsive to their needs; PARENT Networks (or PARENTMAN): quality of the communication between the school and parents; and Access to Expertise (EXPERT or EXPERTMAN): teachers' access to support personnel and quality professional development. These teacher social capital factors are hypothesized to impact two teaching practice factors at the Between teacher level: CONTROL (or CONTROLMAN): a measure of the respect demonstrated by students within the classroom; and INSTRUCTION (OR INSTRUCTIONMAN): a broad measure of students' perceptions of the quality of the teacher's explanations, feedback, provision of cognitive challenge, respect for student thinking, and emotional support. In the latent model, Control and Instruction are the Between level latent factors formed from student responses at Level 1 to Tripod survey items.

In the manifest model, the variances of the manifest indicators, CONTROLMAN and INSTRUCTIONMAN, are estimated at both the Between and Within levels. This results in the formation of Between level factors, CONTROLMAN and INSTRUCTIONMAN, that are latent in the sense of being Level 2 aggregations of the Level 1 constructs. Marsh and colleagues refer to this as a Manifest-Latent model and the type of model I refer to as the Latent Model as a Doubly-Latent model (Marsh, Lüdtke, Robitzsch, Trautwein, Asparouhov, Muthén, &

Nagengast, 2009). Others refer to these as simply multilevel models, since in all multilevel models, if variables are measured at Level 1 and their Level 2 variance is estimated, the model creates estimates of the Level 2 constructs. The Level 2 aggregations are estimated by the model – they are not directly observed and so they can be considered latent in the same sense that all multilevel models can be viewed as latent models. Recall that the manifest composites, CONTROLMAN and INSTRUCTIONMAN are formed by averaging among the Level 1 student survey items. These composites are entered into the multilevel structural equation model as observed variables having variance at both Level 1 and Level 2. In this way, these manifest composites are treated in the same way as the MATH10 outcome variable. The model estimates the Level 2 as well as the Level 1 variance for each of these variables and thus forms Level 2 latent factors for each of these variables that are observed (manifest) at Level 1. The four teacher social capital factors/manifest composites are correlated at the Between level. CONTROL and INSTRUCTION are correlated at the Within and Between levels. All Level 2 factors are hypothesized to effect students' state math test score outcomes, MATH10 at Level 2.

In both models, school racial composition (School percent Black and Hispanic) effects on teacher social capital are estimated at Level 2; in addition, the effects of classroom average prior math achievement on CONTROL and INSTRUCTION are estimated at Level 2. Student math achievement is also regressed on these composition covariates at Level 2, as well as on district fixed effects to account for idiosyncratic differences across districts. At Level 1, the MATH10 outcome is regressed on students' individual prior math scores, grade level, gender, race, and English learner status. In both models, the relationships of interest lie at the Between level. In the Latent model, the Within level includes only the measurement models for CONTROL and INSTRUCTION Within classrooms, the Within Classroom correlation between CONTROL and

INSTRUCTION, the Within variance of MATH10, and the effects of student level covariates on the MATH10 outcome. In the Manifest model, the Within level includes the Within level variance of CONTROLMAN, INSTRUCTIONMAN, and MATH10, the Within correlation of CONTROLMAN and INSTRUCTIONMAN, and the effects of student level covariates on the MATH10 outcome.

Chapter X then turns to consider a model that measures teaching practice using observation ratings rather than student survey items and discusses how inferences change depending on the use of these different measures. This chapter examines research question 3c: How do the relationships among the constructs differ when teaching practice is measured by expert observation ratings rather than students' survey responses? The comparisons are of policy interest because of the widespread use of both the Tripod and the FFT as measures for teacher evaluation and development. These comparisons are also expected to lead to hypotheses regarding the strengths and limitations of each measure, the potentially unique aspects of teaching that each measure best reveals, and the processes that underlie the relationships among the constructs.

Both models (see Figures 10.1 and 10.2) include four teacher social capital factors at the Between teacher level: TRUST: extent to which teachers trust one another and feel trusted; CARES: extent to which teachers perceive the administration as responsive to their needs; PARENT Networks: quality of the communication between the school and parents; and Access to Expertise (EXPERT): teachers' access to support personnel and quality professional development. These factors are correlated at the Between level.

As measured by the Tripod student perception survey items, the teaching practice factors are: CONTROL: a measure of the respect demonstrated by students within the classroom; and

INSTRUCTION: a broad measure of students' perceptions of the quality of the teacher's explanations, feedback, provision of cognitive challenge, respect for student thinking, and emotional support. When the Framework For Teaching is used to measure teaching practices, the factors are named: CLASSROOM MANAGEMENT: formed from ratings on Creating an Environment of Respect and Rapport, Managing Classroom Procedures, and Managing Student Behavior; and INSTRUCTIONAL SUPPORT: formed from ratings on Establishing a Culture for Learning, Using Questioning and Discussion Techniques, Communicating with Students, Engaging Students in Learning, and Using Assessment in Instruction.

The teaching practice factors are conceptually as similar as possible, but are not, of course, defined identically or measured in the same manner across the Tripod and FFT. In the Tripod model, CONTROL and INSTRUCTION are measured through *students'* responses to survey items and the factors are correlated at the student level (Within classrooms) and teacher level (Between classrooms/teachers). In the FFT model, four expert ratings on each domain are averaged to form a Between level manifest domain rating that is akin to a survey item at the teacher (Between) level. The teaching practice factors in the FFT model exist only at Level 2 and they are correlated. In the Tripod model, the teaching practice factors are measured at Level 1 and factors and variances are estimated at both Level 1 and Level 2 and are correlated at both levels.

In both models, all factors are hypothesized to effect students' state math test score outcomes, MATH10, at Level 2. The models include school racial composition effects on teacher social capital factors and the effects of classroom average prior math achievement on teaching practices at Level 2. The MATH10 outcome variable is also regressed on school racial composition, classroom average prior math achievement, and district fixed effects at Level 2 and student level covariates at Level 1: MATH09, race, English language learner status, male, and grade level. The Within level in the model employing the Tripod includes only the measurement models for CONTROL and INSTRUCTION Within classrooms, the Within Classroom correlation between CONTROL and INSTRUCTION, the Within variance of MATH10, and the effects of student level covariates on the MATH10 outcome. In the FFT model, the Within level includes only the Within classroom/teacher variance of MATH10 and the effects of student level covariates on MATH10. The models were estimated using Mplus 7.4 with syntax provided in Appendix D.

The analyses in chapters VIII – X provide empirical examples of the extent to which estimates are impacted by clustering, by measurement error, and by the choice of measurement instrument. Estimated effects and loadings from each model are compared and the estimates are compared with the findings of other research on teacher social capital and teaching practice. In chapter X, the comparison between results is paired with a conceptual comparison of the instruments that examines ways in which the two measures reflect distinct versus overlapping aspects of teaching practice. Reflecting on results from all of these models provides insight into the mechanisms through which social capital impacts teachers' practices and students' achievement.

Chapter VI.

Measurement Results

This chapter examines the issues involved in conceptualizing and measuring teacher social capital and teaching practice, addressing the following research questions:

1. What factors of teacher social capital can be identified using the Teacher Working Conditions Survey?
2. What factors of teaching practice can be identified using the Tripod student perception survey items? What factors of teaching practice can be identified using the Framework For Teaching observation ratings?

Measurement of Teacher Social Capital

The measurement model for teacher social capital included eighteen items from the Teacher Working Conditions Survey to approximate four factors/dimensions (see Table 6.1):

- Trust: the extent to which teachers feel trusted and respected by others and express confidence in other teachers and school leaders to solve problems;
- Cares: the degree to which teachers perceive an honest effort from the administration in addressing teacher concerns;
- Parent Networks: the relative strength of networks between teachers and parents in terms of clear two-way communication and parents being involved and informed;
- Access to Expertise: a measure of the access teachers have to quality professional development and support from experts.

Table 6.1

Descriptive Statistics for Teacher Social Capital Items, Grouped by Factor

Factors / Items	Obs	Mean	SD
Norms: Trust			
Teachers are recognized as educational experts	381	2.94	0.75
There is an atmosphere of trust and mutual respect in this school	382	2.69	0.88
Teachers feel comfortable raising issues/concerns important to them	381	2.69	0.88
Effective faculty process for making group decisions to solve problems	381	2.78	0.91
In this school we take steps to solve problems	380	2.91	0.82
Teachers are effective leaders in this school	381	2.96	0.77
Norms: Caring			
Leadership makes a sustained effort to address teacher concerns re:			
Leadership issues	380	2.95	1.03
The use of time in my school	380	2.89	0.92
Teacher leadership	380	3.01	0.91
Instructional practices and support	381	3.02	0.80
Parent Networks (strength of teacher-parent networks)			
Clear, two-way communication w/ parents/guardians & community	378	3.03	0.82
This school does a good job of encouraging parent/guardian	381	3.06	0.77
Parents/guardians know what is going on in this school	381	2.93	0.86
Access To Expertise (colleagues, support personnel, PD)			
In this school, follow up is provided from professional development	380	2.79	0.94
Sufficient resources are available for professional development	381	2.93	0.75
Professional development deepens teachers' content knowledge	378	2.83	0.84
Appropriate amount of time provided for professional development	379	2.83	0.70
PD provides ongoing opps. to work w/ colleagues to refine teaching	380	2.91	0.79

The measurement model for teacher social capital was developed by conceptually mapping the Teacher Working Conditions Survey items onto the hypothesized latent dimensions of teacher social capital as described in chapter 3: Networks, Norms, Access to Expertise, and Depth of Interactions. This hypothesized factor structure was refined through exploration of item correlations (see Appendix B for patterns of correlations among items), and single-level and multi-level confirmatory factor analyses (CFAs and MCFAs) in Mplus 7.4. CFAs were fit treating the variables first as categorical, i.e., using a robust weighted least squares estimator, WLSMV, and analyzing a polychoric correlation matrix (Muthén & Muthén, 1988-2015), and then treating the response variables as continuous and employing a maximum likelihood robust estimator, MLR. Categorical estimates were deemed to vary only negligibly from the continuous estimates and the MLR estimator was used in final models. Items were dropped on the basis of poor fit and conceptual ambiguity (i.e., unclear which factor they belonged in). The goal of item deletion was to arrive at a multidimensional MSEM model with adequate construct coverage but fewer parameters so that it would be estimable and able to converge (see Appendix C for details on decisions about dropped items). The CFA measurement model is depicted in Figure 6.1; the four teacher social capital factors are correlated to allow for the estimation of the potential influence of each factor on teaching practice and student achievement, thus potentially providing a more nuanced understanding of the relationships among these constructs. A unidimensional model for teacher social capital was also investigated due to the strong correlations among the factors.

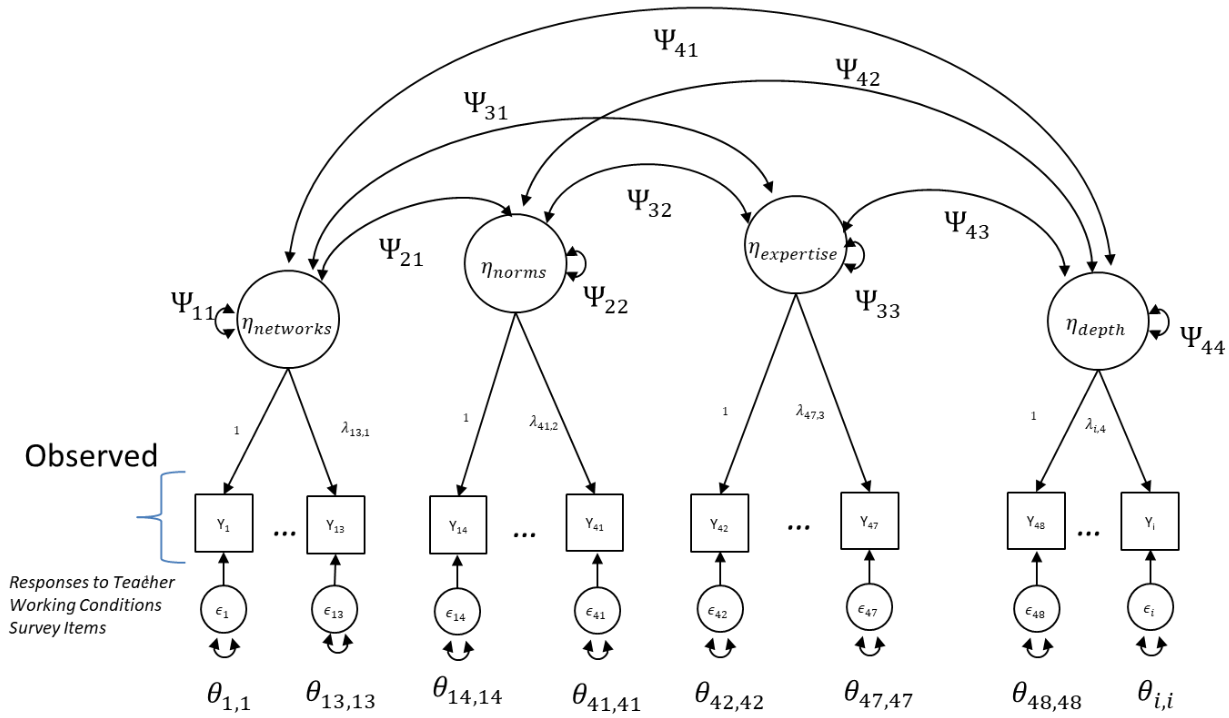

Figure 6.1. Correlated factors CFA model for teacher social capital.

This model had adequate to good fit as indicated by MCFA fit indices: RMSEA: 0.008; CFI: 0.948; TLI: 0.941; SRMR Within: 0.047; SRMR Between: 0.058. Cut-off thresholds for fit indices for multilevel confirmatory factor analyses have not been firmly established by research (Bovaird, 2007; Dyer, Hanges, & Hall, 2005; Ryu & West, 2009), but based on the generally accepted thresholds for single-level factor analyses, the estimated RMSEA is well below the 0.05 threshold and thus suggests very good fit, CFI/TLI values are close to meeting the 0.95 threshold considered good fit, and the SRMR (standardized root mean residual) Within and Between residuals are less than the 1.0 or 0.08 thresholds that have been suggested as indicating adequate fit for the single-level weighted root mean square residual (Cook, Kallen, Amtmann, 2009; Hu & Bentler, 1999; Schmitt, 2011; Yu, 2002).

The factor loadings (see Table 6.2) are strong across all dimensions, ranging from 0.73 to 0.92 at the Between level. Particularly strong and consistent loadings are found for the items in the Cares dimension (0.90 – 0.92 Between), likely reflecting a combination of strong conceptual similarities among the items and a method effect arising from the items all sharing the same statement stem. The strong loadings for all dimensions and the fit indices for the overall model suggest that these dimensions of teacher social capital are adequately measured by these eighteen items from the Teacher Working Conditions Survey. The ICCs displayed in Table 6.2 suggest that a substantial portion (0.39 – 0.48) of the variance of each of these factors "lives" at the school level. Therefore, the teacher social capital factors can be conceptualized as constructs partly shared among all the teachers surveyed at a particular school. Residual variances for the items suggest larger proportions of variance are left unexplained at the Within level (0.24 – 0.58) than at the Between level (0.15 – 0.47), and the items in the Cares (0.15 – 0.20 Between) and Parent (0.21 – 0.47 Between) factors are somewhat better explained by the model than the items in the Trust (0.22 – 0.45 Between) and Expert (0.33 – 0.46 Between) factors.

Table 6.2
Standardized (StdYX) Factor Loadings, ICCs, and Residual Variances for Teacher Social Capital Dimensions: Multilevel Confirmatory Factor Analysis

Factors / Items	**Loading**	**SE**	**Est/SE**	**p-val**	**ICC**	**Res Var**
TRUST Within						
Teachers are recognized as educational experts	0.65	0.04	16.30	0.00	NA	0.58
Atmosphere of trust and mutual respect in this school	0.67	0.04	16.53	0.00	NA	0.55
Teachers feel comfortable raising issues important to them	0.67	0.04	18.25	0.00	NA	0.56
Effective process for group decisions to solve problems	0.83	0.03	30.11	0.00	NA	0.31
In this school, we take steps to solve problems	0.86	0.03	33.60	0.00	NA	0.26
Teachers are effective leaders in this school	0.79	0.04	21.80	0.00	NA	0.37
CARES Within						
Ldrship makes sustained effort on teacher concerns re:						
Leadership issues	0.87	0.02	36.66	0.00	NA	0.24

The use of time in my school	0.78	0.04	20.19	0.00	NA	0.40
Teacher leadership	0.87	0.02	39.04	0.00	NA	0.25
Instructional practices and support	0.81	0.04	22.69	0.00	NA	0.35
PARENT Within						
Clear, two-way communication with parents and teachers	0.83	0.03	26.53	0.00	NA	0.32
School does a good job of encouraging parent involvement	0.80	0.04	21.86	0.00	NA	0.37
Parents/guardians know what is going on in this school	0.65	0.05	14.59	0.00	NA	0.57
EXPERT Within						
Followup provided from professional development	0.78	0.03	25.29	0.00	NA	0.39
Sufficient resources available for professional development	0.76	0.03	23.42	0.00	NA	0.43
Prof. dev. deepens teachers' content knowledge	0.73	0.04	19.31	0.00	NA	0.46
Appropriate amount of time provided for prof. dev.	0.70	0.04	18.59	0.00	NA	0.51
Prof. dev. provides opps. to work with colleagues	0.78	0.04	22.22	0.00	NA	0.40
TRUST Between					**0.43**	
Teachers are recognized as educational experts	0.74	0.05	13.63	0.00	0.40	0.45
Atmosphere of trust and mutual respect in this school	0.79	0.05	14.97	0.00	0.37	0.37
Teachers feel comfortable raising issues important to them	0.81	0.05	15.52	0.00	0.39	0.35
Effective process for group decisions to solve problems	0.89	0.03	33.58	0.00	0.39	0.22
In this school we take steps to solve problems	0.87	0.03	29.38	0.00	0.37	0.24
Teachers are effective leaders in this school	0.87	0.03	26.93	0.00	0.38	0.24
CARES Between					**0.48**	
Leadership makes sustained effort on teacher concerns re:						
Leadership issues	0.90	0.03	30.36	0.00	0.43	0.20
The use of time in my school	0.92	0.03	32.52	0.00	0.37	0.15
Teacher leadership	0.91	0.03	27.59	0.00	0.47	0.17
Instructional practices and support	0.91	0.03	28.04	0.00	0.46	0.17
PARENT Between					**0.48**	
Clear, two-way communication with parents and teachers	0.89	0.03	28.00	0.00	0.43	0.21
School does a good job of encouraging parent involvement	0.88	0.04	21.48	0.00	0.43	0.22
Parents/guardians know what is going on in this school	0.73	0.06	11.91	0.00	0.46	0.47
EXPERT Between					**0.39**	
Followup provided from professional development	0.82	0.05	15.52	0.00	0.35	0.33
Sufficient resources available for professional development	0.77	0.07	10.36	0.00	0.42	0.41
Prof. dev. deepens teachers' content knowledge	0.73	0.07	11.12	0.00	0.36	0.46
Appropriate amount of time provided for prof. dev.	0.77	0.07	11.41	0.00	0.36	0.41
Prof. dev. provides opps. to work with colleagues	0.80	0.06	13.46	0.00	0.37	0.37

18 items (level 1 N = 12,888 students, level 2 N = 90 schools)
RMSEA: 0.008; CFI: 0.948; TLI: 0.941; SRMR Within: 0.047; SRMR Between: 0.058

The model also shows strong correlations among the factors, ranging from 0.62 to 0.82 (see Table 6.3). High factor correlations are expected, especially in the correlation between the two subdimensions of norms, Trust and Cares. The correlation between Cares and Access to Expertise is also high, likely reflecting the fact that caring measures teachers' perceptions of the administration's efforts to meet teacher concerns and expertise measures teachers' perceptions of the professional support they receive, which relies in part on the administration's efforts. However, the high factor correlations also suggest that the full structural model may be subjected to issues of multicollinearity among the factors resulting in possible suppression effects that could make the interpretation of results difficult. Due to these high correlations, a unidimensional model is also considered in chapter VII.

Table 6.3
Teacher Social Capital Factor Correlations, at Between Level (Within Level)

	TRUST	CARES	PARENT	EXPERT
TRUST	1			
CARES	0.80 (0.77)	1		
PARENT	0.77 (0.59)	0.62 (0.49)	1	
EXPERT	0.72 (0.66)	0.82 (0.72)	0.65 (0.46)	1

Measurement Model for Teaching Practice

Tripod Student Perceptions Survey

The measurement model for teaching practice using the Tripod student perception survey items consisted of two factors: CONTROL – measured by five items, and INSTRUCTION – measured by 14 items (see Table 6.4 for the list of items grouped by factor, item wordings, and descriptive statistics). The Control factor measured teachers' classroom management practices and the extent to which students exhibited respectful and orderly behavior. It consisted of five of the seven items in the Tripod developers' original conceptualized dimension, therefore I retained the name Control for this factor. The Instruction factor consisted of 14 items from the other six dimensions in the developers' original Seven Cs conceptualization. Collectively, they measured the quality of teachers' instructional practice in terms of emotional and instructional support, effectiveness in delivering material clearly and in multiple ways, high expectations for student learning, engagement of students, and checking for student understanding.

Table 6.4
Tripod Student Survey Items in Final Model, Grouped by Factor

Var.	Item Wording	Obs	Mean	SD
	CONTROL			
B112	Student behavior in this class is under control.	16,383	3.33	1.27
B138	Student behavior in this class is a problem.*	16,455	2.74	1.31
B46	My classmates behave the way my teacher wants them to.	16,374	3.07	1.23
B49	Students in this class treat the teacher with respect.	16,647	3.53	1.18
B6	Our class stays busy and doesn't waste time.	16,612	3.48	1.19
	INSTRUCTION			
A10	My teacher in this class makes me feel that he/she really cares about me.	16,557	3.64	1.25
B34	My teacher really tries to understand how students feel about things.	16,308	3.45	1.23
B1	If you don't understand something, my teacher explains it another way.	16,600	4.04	1.08
B17	My teacher has several good ways to explain each topic we cover in this class.	16,525	3.90	1.10
B80	My teacher explains difficult things clearly.	16,216	3.83	1.13
B70	In this class, we learn a lot almost every day.	16,356	3.97	1.06
B90	In this class, we learn to correct our mistakes.	16,188	4.05	1.04
B29	My teacher makes learning enjoyable.	16,578	3.49	1.29
B44	My teacher makes lessons interesting.	16,375	3.48	1.26
B154	My teacher gives us time to explain our ideas.	16,104	3.64	1.14
A54	My teacher respects my ideas and suggestions.	16,127	3.67	1.15
B147	My teacher checks to make sure we understand what s/he is teaching us.	16,280	4.08	1.07
B58	We get helpful comments to let us know what we did wrong on assignments.	16,401	3.67	1.20
B83	The comments that I get on my work in this class help me understand how to improve.	16,419	3.68	1.17

19 items; N = 16,887 students, 520 teachers; Item response Ns range from 16,104 – 16,647
Shading indicates developers' 7Cs (7 dimensions) conceptualization.

Conceptually, this model appears to provide similar construct coverage to the full 36-item Tripod Survey, with five *Control* items, and two or three items from each of the other original six Cs (shaded to separate the original Seven Cs conceptualization in Table 6.4). However, results from this model should not be directly compared to results obtained using all Tripod items because nearly half of the items were dropped. Items were dropped on conceptual, technical, and empirical grounds (see Appendix C for detailed information on how items were dropped).

This Control + Instruction model was arrived at after attempting to fit the model proposed by Ferguson & Danielson (2014). Their conceptualization grouped the 36 items into two factors, a PRESS factor, made up of items originally grouped into Control and Challenge constructs, and a SUPPORT factor, made up of the other five constructs: Care, Clarify, Captivate, Confer, and Consolidate. However, initial multilevel confirmatory factor analysis using PRESS and SUPPORT did not converge nor produce an output file of any sort. Separate models for each factor suggested that fit for the PRESS factor alone was not acceptable (RMSEA: 0.077, CFI: 0.648, TLI: 0.590, SRMR Within: 0.104 and Between: 0.208). Conceptual analysis of individual items suggested that the items in the Challenge construct did not align well with the items in the Control construct. The items in the Control construct revolve around student behaviors (i.e., B112: Student behavior in this class is under control). The items in the Challenge construct, on the other hand, measure teacher behaviors and teacher desires. Further, several of these teacher behaviors and desires appear to exist independently of well-behaved students (i.e., B45: My teacher wants us to use our thinking skills, not just memorize things). It is not difficult to imagine a set of about equally well-behaved classrooms whose teachers vary widely in the extent to which they expect their students to use thinking skills. Further, several of the Challenge items share important characteristics with items grouped with the other five constructs collectively labeled Support. In particular, Challenge item B128: My teacher asks questions to be sure we are following along when s/he is teaching, aligns closely with Consolidate item B147: My teacher checks to make sure we understand what s/he is teaching us. Also, Challenge item B90: In this class, we learn to correct our mistakes, shares a great deal in common with Consolidate items

B58: We get helpful comments to let us know what we did wrong on assignments, and B83: The comments that I get on my work in this class help me understand how to improve. In addition, Challenge item B59: My teacher wants me to explain my answers – why I think what I think, overlaps importantly with Confer items B129: My teacher wants us to share our thoughts, B154: My teacher gives us time to explain our ideas, and A54: My teacher respects my ideas and suggestions. When so many items in purportedly separate dimensions share nontrivial conceptual similarities, the dimensions they are intended to measure are likely to collapse into a general factor.

Further evidence against the PRESS + SUPPORT conceptualization came from Kuhfeld (under review), and Wallace, Kelcey, & Ruzek (2016), as well as from the MET User's Guide, all of which point in favor of a model that treats Control as its own separate factor. Kuhfeld investigated the structure of the Tripod survey using multilevel Item Response Theory and arrived at a model that maintained Control as its own separate factor and grouped all the other items from the other six constructs together as a second factor (CONTROL + 6Cs). Similarly, the MET User Guide reports that factor analyses found that the Control items did not load on the same factor as the other 6 Cs (Bill & Melinda Gates Foundation, 2012). Because of these findings, the MET team created a composite variable made up of the other six Cs (and a second composite variable was also created based on the Control items, as the MET team had created a separate composite for each of the 7Cs), but the two-factor CONTROL + 6Cs model implied by these analyses does not appear to have made its way into any published MET reports. Wallace and colleagues, meanwhile, recently published a multilevel IRT analysis of the Tripod that identified a bifactor structure with a GENERAL teaching factor underlying all items and a specific CONTROL factor.

I examined the fit of Kuhfeld's preferred model, keeping CONTROL as its own separate factor and grouping all other items together as a second factor, termed INSTRUCTION. Because initial estimation still had problems with convergence and misfit, I re-analyzed the items and dropped weaker items to arrive at a streamlined model. Conceptual analysis proceeded from the standpoint of identifying as problematic any item that did not appear to be based in students' natural expertise as raters of their classroom experiences. Wallace, Kelsey, and Ruzek (2016) explain that the basis for the validity of student survey ratings of teaching quality stems from the fact that students possess "naturally acquired expertise through their lived, everyday experiences in classrooms (p.3)". Students' expertise as raters is based on their sensitivity to the meaningfulness of the work they are expected to do, and the teaching practices that support or constrain their effective participation and understanding. Students' firsthand experiences "situate their assessments of teaching quality as originating primarily from the perspective of a learner versus that of the teacher (p. 4)". Therefore, items asking students to infer the state of mind of their teacher rather than on their "naturally acquired expertise" as learners in that classroom rest on a possibly flimsy foundation and these items were dropped. Technical analysis identified issues based on negative wording, negative connotations in wording, misalignment with overall teacher quality, and emotionally loaded language. Any of these issues may result in student raters' confusion. Empirical analysis was based on item-by-item correlations, loadings, modification indices, residual correlations, residual variances, and item R-square values from correlation matrices and multilevel factor analysis models. A table listing the deleted items and the conceptual, technical, and empirical problems identified is provided in Appendix C along with a brief discussion of the rationale for dropping each item.

In brief, twelve items from the INSTRUCTION factor were identified in the initial conceptual analysis as being likely to be misinterpreted by student respondents and thus perform poorly. This list of items aligned exactly with an empirical analysis of the item-by-item correlations that was used to identify the items that had the worst fit with the data. These items were dropped, the fit of the model for the Instruction factor was re-examined, and a multilevel exploratory factor analysis was conducted to shed further light on the structure and fit. This model had good to adequate fit and eigenvalues suggested two Within factors and one Between factor (see Appendix C, Table C3). However, all items loaded strongly on the first Within factor and the second Within factor was not readily interpretable. Re-analysis of the item wordings in light of the item loadings on the second Within factor suggested that this factor was picking up on negative connotations in items B128, B129, and B59. These three items also asked students to infer the state of mind of their teacher and were identified as relatively poorly performing in terms of low loadings, large modification indices, large residual correlations, high residual variances, and low item R-square values from the MCFA. Other items were identified that measured the same aspects of the construct, and the items were dropped, resulting in a final measurement model for Instruction measured by 14 items. A similar procedure was followed for the items measuring the Control factor, examining conceptual, technical, and empirical misfit for items, and resulting in the deletion of two items.

Fit indices and standardized factor loadings for the final model consisting of 5 Control items and 14 Instruction items are provided in Table 6.5.

Table 6.5

Standardized Factor Loadings for 2-Factor MCFA of Tripod (19 Items: Control & Instruction)

Var.	Factor / Item	Loading	S.E.	Est./S.E.	P-Val	ICC	Res. Var.

	CONTROL Within						
PB112	Student behavior in this class is under control	0.61	0.01	63.85	0.00	NA	0.63
PB138R	Student behavior in this class is a problem	0.51	0.01	47.36	0.00	NA	0.74
PB46	My classmates behave the way my teacher wants	0.71	0.01	98.71	0.00	NA	0.49
PB49	Students in this class treat the teacher with respect	0.70	0.01	115.88	0.00	NA	0.51
PB6	Our class stays busy and doesn’t waste time	0.54	0.01	65.94	0.00	NA	0.71
	INSTRUCTION Within						
PA10	Teacher makes me feel he/she really cares about me	0.68	0.01	123.87	0.00	NA	0.54
PB34	Teacher really tries to understand how students feel	0.68	0.01	116.23	0.00	NA	0.54
PA54	My teacher respects my ideas and suggestions	0.70	0.01	122.32	0.00	NA	0.51
PB154	My teacher gives us time to explain our ideas	0.67	0.01	114.29	0.00	NA	0.55
PB1	If you don’t understand, my T explains another way	0.65	0.01	104.43	0.00	NA	0.58
PB17	T has several good ways to explain each topic	0.70	0.01	122.38	0.00	NA	0.51
PB80	My teacher explains difficult things clearly	0.70	0.01	113.62	0.00	NA	0.50
PB29	My teacher makes learning enjoyable	0.74	0.01	139.15	0.00	NA	0.46
PB44	My teacher makes lessons interesting	0.74	0.01	144.96	0.00	NA	0.46
PB147	T checks to make sure we understand the teaching	0.70	0.01	114.82	0.00	NA	0.52
PB58	We get helpful comments on assignments	0.64	0.01	97.89	0.00	NA	0.60
PB83	Comments on work help me understand…improve	0.66	0.01	115.77	0.00	NA	0.56
PB70	In this class, we learn a lot almost every day	0.58	0.01	69.64	0.00	NA	0.66
PB90	In this class, we learn to correct our mistakes	0.65	0.01	104.62	0.00	NA	0.58
	Factor Correlation						
	INSTRUCTION WITH CONTROL (Within)	0.54	0.01	50.23	0.00	NA	NA
	CONTROL Between					**0.37**	
PB112	Student behavior in this class is under control	0.99	0.01	218.89	0.00	0.17	0.02
PB138R	Student behavior in this class is a problem	0.91	0.01	68.16	0.00	0.18	0.17
PB46	My classmates behave the way my teacher wants	1.00	0.00	319.45	0.00	0.21	0.00
PB49	Students in this class treat the teacher with respect	0.95	0.01	131.25	0.00	0.26	0.10
PB6	Our class stays busy and doesn’t waste time	0.91	0.01	77.09	0.00	0.17	0.18
	INSTRUCTION Between					**0.26**	
PA10	Teacher makes me feel he/she really cares about me	0.94	0.01	112.20	0.00	0.18	0.15
PB34	Teacher really tries to understand how students feel	0.97	0.01	158.87	0.00	0.16	0.12
PA54	My teacher respects my ideas and suggestions	0.95	0.01	125.03	0.00	0.13	0.07
PB154	My teacher gives us time to explain our ideas	0.96	0.01	136.59	0.00	0.15	0.10
PB1	If you don’t understand, my T explains another way	0.97	0.01	189.34	0.00	0.14	0.08
PB17	T has several good ways to explain each topic	0.97	0.01	174.04	0.00	0.17	0.05
PB80	My teacher explains difficult things clearly	0.90	0.01	78.71	0.00	0.15	0.06
PB29	My teacher makes learning enjoyable	0.94	0.01	110.68	0.00	0.24	0.18
PB44	My teacher makes lessons interesting	0.97	0.01	176.69	0.00	0.22	0.12
PB147	T checks to make sure we understand the teaching	0.95	0.01	110.73	0.00	0.15	0.06
PB58	We get helpful comments on assignments	0.96	0.01	135.16	0.00	0.12	0.10
PB83	Comments on work help me understand…improve	0.84	0.02	46.63	0.00	0.12	0.09
PB70	In this class, we learn a lot almost every day	0.94	0.01	108.76	0.00	0.13	0.30

PB90	In this class, we learn to correct our mistakes	0.92	0.01	94.42	0.00	0.13	0.12
	Factor Correlation						
	INSTRUCTION WITH CONTROL (Between)	0.65	0.03	20.33	0.00	NA	NA

N = 16,887 Students, 520 Teachers
RMSEA: 0.037; CFI:0.935; TLI: 0.930; SRMR Within: 0.040; SRMR Between: 0.072

This MCFA model has adequate fit as indicated by RMSEA: 0.037; CFI:0.935; TLI: 0.930; SRMR Within: 0.040; SRMR Between: 0.072. Based on the thresholds for single-level factor analyses, the estimated RMSEA is below the 0.05 threshold and thus suggests good fit, and the SRMR (standardized root mean residual) Within and Between residuals are less than the 1.0 or 0.08 thresholds that have been suggested as indicating adequate fit for the single-level weighted root mean square residual; however, the CFI/TLI values are below the 0.95 threshold considered good fit and may indicate less than adequate fit (Cook, Kallen, Amtmann, 2009; Hu & Bentler, 1999; Schmitt, 2011; Yu, 2002). Substantial amounts of variance for these items is unaccounted for at the Within level, with residual variances at the Within level ranging from 0.46 to 0.74, whereas at the Between level, the residual variances are much smaller: 0.30 or less in all cases with most under 0.15, suggesting the model has better fit at the Between level.

The factor loadings (see Table 6.5) are strong across all dimensions, ranging from 0.51 to 0.74 at the Within level and from 0.84 to 0.998 at the Between level. Particularly strong loadings are found for the items in the Control dimension at the Between level. Indeed, the loadings for item B46: My classmates behave the way my teacher wants them to (0.71 Within and 0.998 Between), are so strong as to suggest that a single-item measure of classroom control might be explored (see, e.g., Peterson, Wahlquist, & Bone, 2000). The overall strong loadings for all dimensions and the fit indices for the overall model suggest that these dimensions of teaching practice are adequately measured by these nineteen items from the Tripod Student Perceptions Survey. The model also shows strong correlations among the factors, 0.54 Within and .65

Between. The strong factor correlations, especially in light of the even stronger factor correlations for teacher social capital, suggest that the full structural model may be subjected to issues of multicollinearity/suppression. Due to these high correlations, a unidimensional model is explored in Chapter VII.

Framework For Teaching Observation Ratings

The Framework For Teaching measurement model attempted to mirror, as closely as practical, the two-factor structure established for the Tripod. Confirmatory factor analyses resulted in two factors, CLASSROOM MANAGEMENT, composed of three components: Creating and Environment of Respect and Rapport, Managing Classroom Procedures, and Managing Student Behavior; and INSTRUCTIONAL SUPPORT, composed of five components: Establishing a Culture for Learning, Communicating with Students, Using Questioning and Discussion Techniques, Engaging Students in Learning, and Using Assessment in Instruction. This model was established after moving Establishing a Culture for Learning from the CLASSROOM MANAGEMENT factor to the INSTRUCTIONAL SUPPORT factor based on the factor loadings, modification indices, residual variances and overall model fit, as well as conceptual analysis that suggested that Establishing a Culture for Learning assessed the teacher's level of expectations, similar to the Challenge items in the Tripod. This amended FFT model has good fit to the data based on the thresholds established for CFI, TLI, and SRMR, though the RMSEA value is above the 0.08 cut-off for acceptable fit and thus suggests inadequate fit. The two factors of CLASSROOM MANAGEMENT and INSTRUCTIONAL SUPPORT are correlated 0.849, suggesting that a unidimensional model might be appropriate, but the two-factor model was retained to allow for close comparison with the Tripod model. As seen in Table 6.6, factor loadings are strong for all components, particularly for the CLASSROOM

MANAGEMENT components, and residual variances also suggest that the model somewhat better explains the variance in the CLASSROOM MANAGEMENT components than in the INSTRUCTIONAL SUPPORT components.

Table 6.6
Standardized Factor Loadings and Residual Variances for 2-Factor CFA of FFT

Factor / Component	Loading	S.E.	Est./S.E.	P-Value	Residual Variance
CLASSROOM MANAGEMENT					
2a Creating an Environment of Respect and Rapport	0.92	0.00	359.91	0.00	0.16
2c Managing Classroom Procedures	0.81	0.00	206.56	0.00	0.35
2d Managing Student Behavior	0.81	0.00	208.86	0.00	0.35
INSTRUCTIONAL SUPPORT					
2b Establishing a Culture for Learning	0.71	0.01	138.49	0.00	0.50
3a Communicating with Students	0.86	0.00	286.40	0.00	0.25
3b Using Questioning and Discussion Techniques	0.76	0.00	172.68	0.00	0.42
3c Engaging Students in Learning	0.79	0.00	201.49	0.00	0.37
3d Using Assessment in Instruction	0.77	0.00	178.54	0.00	0.41
Correlation: CR Management w/ Instr. Support	0.85	0.00	225.73	0.00	NA

RMSEA: 0.097; CFI: 0.966; TLI: 0.950; SRMR: 0.032

Discussion

As discussed in detail in chapters II and III, the published literature on teacher social capital and teaching practice are far from speaking with one voice in regard to the structures of these constructs. Even so, the multidimensional measurement structures found here are in accord with particularly relevant work using the Teacher Working Conditions and Tripod surveys. In regard to the teacher social capital factors, the structure identified here is based on an independent conceptual analysis of the teacher working conditions items in light of the teacher social capital and related literature discussed above and it too relies on a reduced set of items; even so, the factors identified can best be characterized as representing slight adjustments to a subset of the factors identified in the structural analysis of the Teacher Working Conditions

Survey (New Teacher Center, 2014). In regard to the teaching practice factors, the results in this dissertation align with other recent analyses of the Tripod survey in the MET data. Most pertinently, Kuhfeld (under review), found support for a similar two-factor structure (CONTROL and other 6 Cs), and the structure in this dissertation is also broadly in accord with the results from Wallace, Kelcey, and Ruzek (2016), who identified a bifactor structure with a GENERAL teaching factor and a specific CONTROL factor. An important distinction between this study and the Kuhfeld and Wallace et al analyses is that this study relies on a substantially reduced set of items. The FFT analyses produced a two-factor model that broadly mirrors the Tripod structure; further analysis of the measurement properties of the FFT vis-à-vis the Tripod is provided in Chapter X.

The measurement models also provide information about the relative strength of the measurement of these constructs. In terms of teacher social capital, the CARES factor and the PARENT NETWORK factor show the strongest factor loadings and the highest ICCs. With regard to the CARES factor, these items all share the same statement stem, suggesting that a method effect is likely the cause of some degree of the high inter-item correlations. In comparing the two teaching practice factors, the loadings are stronger for CONTROL and ICCs are higher. These results suggest that students can better measure the appropriateness of student behavior than the quality of a teacher’s feedback, the clarity of explanations, the sincerity of caring, and the like. It is also quite possible that the relatively weaker loadings and greater amount of variance among students within classrooms on INSTRUCTION items reflects true differences in the teaching quality experienced by different students. For instance, in some classrooms, the quality of feedback provided to a student identified as gifted might systematically differ from the quality experienced by other students. In the FFT analyses, the CLASSROOM MANAGEMENT factor also showed stronger measurement properties than the INSTRUCTIONAL SUPPORT factor. The causes of these differences in the strength of the measures cannot be definitively proven from these data, but the results are more readily predictable: we should expect to find stronger relationships among the factors that are measured more accurately, namely CARES, PARENT-TEACHER NETWORKS, CONTROL and CLASSROOM MANAGEMENT.

In the following chapters, I discuss the structural relationships among these teacher social capital factors, teaching practice factors, and student math achievement in grades 6-8.

Chapter VII.

Results from Multilevel Structural Equation Model

This chapter examines the evidence regarding the structural relationships among teacher social capital, teaching practice, and student achievement – both direct effects of social capital on student achievement, and indirect effects through teaching practice. The results address the third research question guiding this study, namely: What are the relationships among teacher social capital, teaching practice and student achievement?

The relationships among these constructs are investigated using the multilevel multidimensional structural equation model discussed in detail in chapter V. The model (see Figure 7.1) includes four teacher social capital factors: Trust: the extent to which teachers trust one another and feel trusted; Cares: the extent to which teachers perceive the administration as responsive to their needs; Parent Networks: the quality of the communication between the school and parents; and Access to Expertise: teachers' access to support personnel and quality professional development. The teacher social capital factors are hypothesized to impact two teaching practice factors: Control: a measure of the respect demonstrated by students within the classroom; and Instruction: a broad measure of students' perceptions of the quality of the teacher's explanations, feedback, provision of cognitive challenge, respect for student thinking, and emotional support. The four teacher social capital factors are correlated as are the two teaching practice factors and all Level 2 latent factors are hypothesized to effect students' state math test score outcomes, MATH10 at Level 2.

Teacher social capital factors are regressed on school racial composition (School percent Black and Hispanic) to account for differences in schools in terms of the populations of students served along with unmeasured school differences that are typically associated with levels of school segregation. In addition, teaching practices at Level 2 are regressed on classroom average prior math achievement and classroom racial composition (Classroom percent Black and Hispanic) to account for the effects of classroom composition. At Level 2, the MATH10 outcome is regressed on School percent Black and Hispanic and classroom prior math, as well as district fixed effects to account for unmeasured systematic differences across the separate districts.[4] At Level 1, MATH10 is regressed on students' individual prior math scores, grade level, gender, race, and English learner status. Equations for the model shown in Figure 7.1 and the Mplus syntax are detailed in Appendix D.

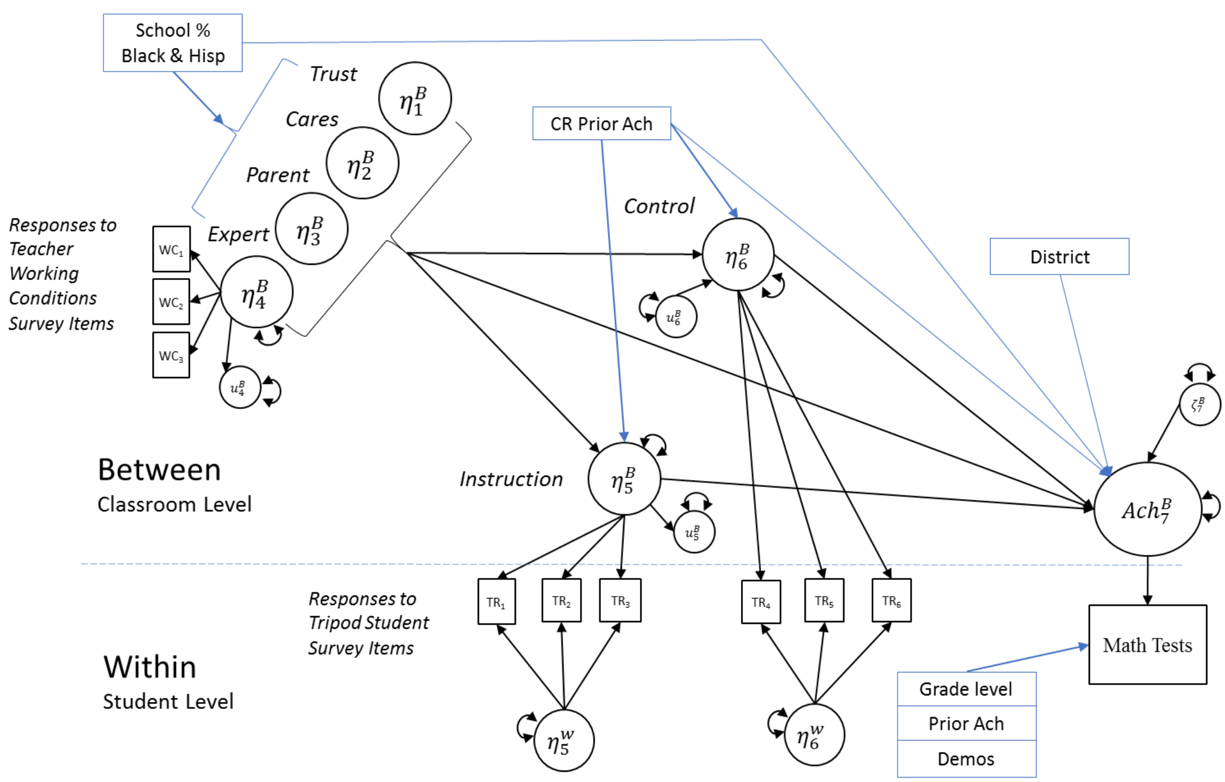

Figure 7.1. Multilevel Multidimensional Structural Equation Model. Teacher Social Capital factors are correlated, as are Teaching Practice factors. Working Conditions Survey Items for $\eta_{1\text{-}3}$ not depicted. Model estimated using state math tests as the outcome.

[4] Note that district fixed effect estimates are not available for release due to the MET data use agreement.

Results

Results from the multidimensional multilevel structural equation model are shown in Tables 7.1 and 7.2. Covariates' effects on the MATH10 outcome were all significant at <.01 level, except Male, grade level, and school and classroom racial composition, which were not significant. District fixed effect estimates are not available for release due to the MET data use agreement.

As noted below Table 7.1, the RMSEA (0.02) indicates good fit by the generally accepted threshold of 0.05 in single-level analyses, though the CFI (0.93) and TLI (0.93) estimates are somewhat below the conventional threshold of 0.95 indicating good fit. The SRMR (standardized root mean residual) Within (0.04) and Between (0.07) residuals are less than the 1.0 or 0.08 thresholds that have been suggested as indicating adequate fit for the single-level weighted root mean square residual (Cook, Kallen, Amtmann, 2009; Hu & Bentler, 1999; Schmitt, 2011; Yu, 2002). In this multilevel context, the SRMR values suggest the model fits somewhat better at the Within level than at the Between level. Large residual variances suggest that most of the variance in these factors remains unexplained by the model, particularly for Instruction.

Table 7.1

Variances, Residual Variances, and Model Fit in Multidimensional MSEM Model

	MSEM Multidimensional			
	Est	SE	Est/SE	p-val
Variances				
CONTROL	0.61	0.01	45.21	0.00
INSTRUCT	0.50	0.02	31.81	0.00
Residual Variances				
TRUST	0.94	0.02	38.92	0.00
CARES	0.96	0.02	54.62	0.00

PARENTS	0.95	0.03	38.24	0.00
EXPERT	0.96	0.02	57.57	0.00
CONTROL Between	0.91	0.03	32.90	0.00
INSTRUCT Between	0.98	0.02	56.20	0.00
MATH10 Within	0.37	0.01	42.14	0.00
MATH10 Between	0.54	0..04	13.54	0.00

Model Fit: RMSEA: 0.022; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.07
Variances for Control and Instruct estimated as invariant across levels. Residual variances are standardized values.

As shown in Table 7.2, in this model, the estimated effect of the school percent Black and Hispanic on teacher social capital factors ranges from negative 0.20 to negative 0.24 and is statistically significant at $p < 0.01$, suggesting that schools with higher percentages of Black and Hispanic students tend to have lower ratings on teacher social capital factors. Also, the estimated strength of the teacher's Control is stronger in classrooms with higher average prior achievement (effect estimate: 0.25), but this association does not hold in the students' ratings of the quality of Instruction. Teaching practice factors are not significantly impacted by the classroom racial composition (Classroom percent Black & Hispanic).

The results from the multidimensional model show that Control has a significant association with student math outcomes (0.37 at $p < .01$) while Instruction has no significant association with student math outcomes. Among teacher social capital factors, only Parent Networks shows a significant positive direct effect on student math outcomes (0.18). Access to Expertise has a marginally significant direct effect on Control (0.16) and a significant effect on Instruction (0.22). No significant effects were found for the Trust or Cares factors and no indirect effects from teacher social capital factors through teaching practice factors were significant.

Table 7.2
Structural Effects in Multidimensional Multilevel Structural Equation Model (MSEM); N = 15,644 Students, 520 Teachers.

	Est	SE	Est/SE	p-val
CONTROL Between ON				

TRUST	-0.11	0.11	-1.01	0.31	
CARES	-0.02	0.12	-0.12	0.90	
PARENT	0.11	0.08	1.29	0.20	
EXPERT	0.16	0.10	1.66	0.10	*
INSTRUCT Between ON					
TRUST	-0.06	0.12	-0.47	0.64	
CARES	-0.10	0.12	-0.77	0.45	
PARENT	-0.01	0.09	-0.12	0.91	
EXPERT	0.22	0.10	2.29	0.02	**
MATH10 ON					
CONTROL Between	0.37	0.07	5.37	0.00	***
TRUST	0.04	0.09	0.38	0.71	
CARES	0.15	0.11	1.38	0.17	
PARENT	0.18	0.08	2.18	0.03	**
EXPERT	-0.21	0.09	-2.24	0.03	**
INSTRUCT Between	0.02	0.07	0.31	0.76	
Factors Regressed on Covariates					
…ON School % Black & Hispanic					
TRUST	-0.24	0.05	-4.87	0.00	***
CARES	-0.20	0.04	-4.43	0.00	***
PARENTS	-0.22	0.06	-3.95	0.00	***
EXPERT	-0.20	0.04	-4.66	0.00	***
…ON Classroom Prior Achievement					
CONTROL Between	0.25	0.06	4.60	0.00	***
INSTRUCT Between	0.02	0.06	0.33	0.74	
…ON Classroom % Black & Hispanic					
CONTROL Between	0.02	0.06	0.30	0.77	
INSTRUCT Between	0.08	0.06	1.31	0.19	
MATH10 on Covariates					
MATH09	0.76	0.01	107.82	0.00	***
S_BLACK	-0.09	0.01	-11.59	0.00	***
S_HISP	-0.05	0.01	-6.85	0.00	***
S_ELL	-0.02	0.01	-3.05	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20	
GRADE6	0.01	0.02	0.52	0.61	
GRADE7	0.00	0.01	-0.33	0.74	
School % Black & Hispanic	-0.21	0.13	-1.54	0.12	
Class Prior Math Achieve	0.41	0.06	6.72	0.00	***
Class % Black & Hispanic	0.16	0.14	1.14	0.25	

*** = significant at p < 0 .01; ** = significant at p < 0.05; * = marginally significant at p < 0.1

Perhaps the most puzzling and intriguing result is that Access to Expertise shows a significant negative direct effect (-0.21) on student state math test scores. Also counter to the theoretical model I have proposed, levels of Trust among teachers and the extent to which the administration Cares about teacher concerns show no significant effects on teaching practices or student outcomes; in fact, the estimated effects of Trust and Cares on teaching practices are negative in direction (-0.02 to -0.11), though not significant. Considering that the model implies strong inter-correlations among the four teacher social capital factors (ranging from 0.51 to 0.78) and the two teaching practice factors (0.54 at Within level; 0.68 at Between level), the presence of these puzzling and counterintuitive negative estimates, particularly the significant negative effect of Expertise on student math outcomes, suggests the possibility that the model suffers from multicollinearity among the factors and the negative effects are caused by multicollinearity/suppression rather than representing true effects.

Sensitivity Analyses: Alternative Models Considered

To investigate the sensitivity of these results to possible multicollinearity and suppression effects, I conduct two sensitivity checks. First, I re-examine the relationships among these constructs within a simplified multilevel mediation model employing a unidimensional teacher social capital factor, a unidimensional teaching practices factor, and students’ state math test scores (see Figure 7.2).

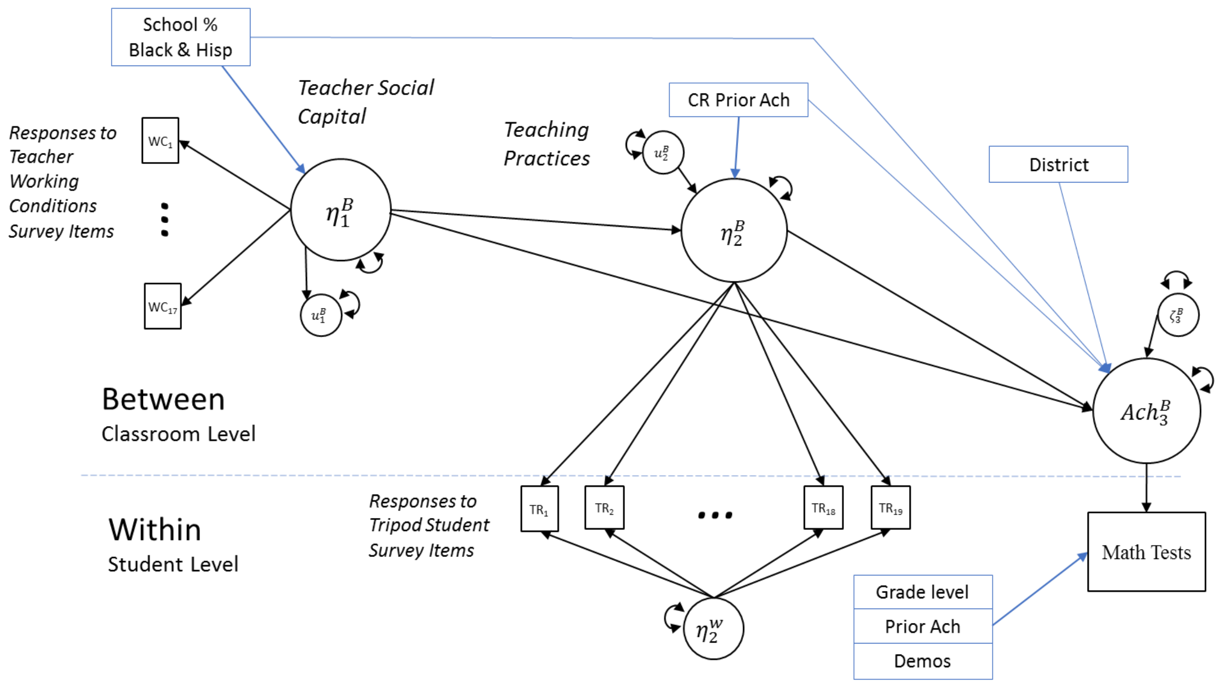

Figure 7.2. Multilevel Structural Equation Model (MSEM) with a Unidimensional Teacher Social Capital Factor and a Unidimensional Teaching Practice Factor

Second, I re-re-examine the relationships using a series of multilevel mediation models that isolate the effects of each of the social capital and teaching practice factors (see Figure 7.3). For the teacher social capital factors, first the effects of TRUST on CONTROL and MATH10 are estimated, then the effects of TRUST on INSTRUCTION and MATH10 are estimated in a separate model. The next model estimates the effects of CARES on CONTROL and MATH10, and so on. For the teaching practice factors, I estimate two models: one with all four teacher social capital factors on CONTROL and MATH10 and another model with all four teacher social capital factors on INSTRUCTION and MATH10. The results from these models are compared to the multidimensional results and implications are discussed. For these comparisons, the classroom percent Black and Hispanic covariate is dropped from the models because it was found to have no significant effects in the multidimensional and unidimensional analyses.

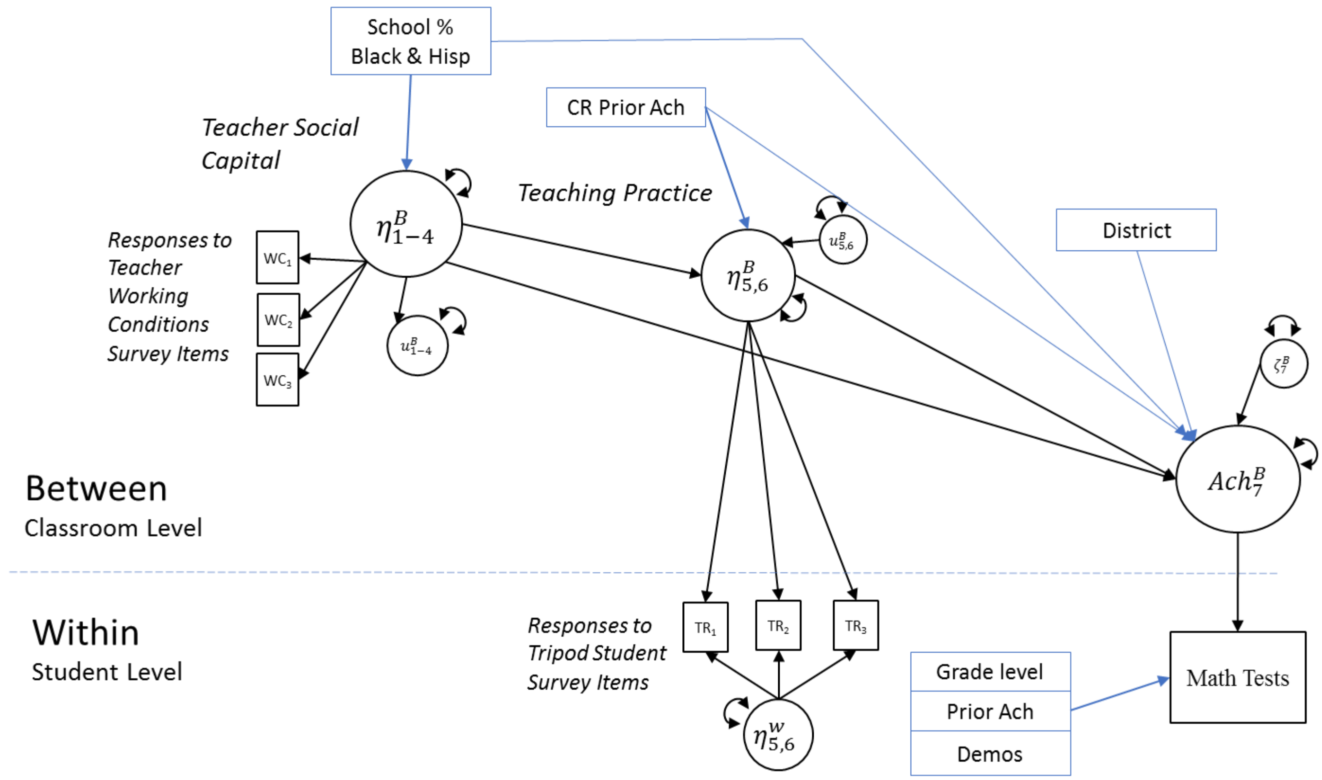

Figure 7.3. MSEM Models with Factors Considered One-by-One. Teacher social capital and teaching practice factors are substituted in place of one another and model is re-estimated.

Results from Model with Unidimensional Latent Factors

The unidimensional teacher social capital factor is hypothesized to explain the covariances among all eighteen teacher working conditions items used in the prior model and, similarly, the unidimensional teaching practice factor is hypothesized to explain the covariances among all nineteen Tripod items used in the prior model. Results are displayed in Table 7.3 and can be compared to results from the multidimensional model in Table 7.2.

Table 7.3
Structural Effects in Unidimensional MSEM Model (Between); 15,644 Students, 520 Teachers.

	Est	SE	Est/SE	p-val
Teaching Practice Between ON				

Teacher Social Capital	0.03	0.06	0.60	0.55	
MATH10 ON					
Teaching Practice Between	0.28	0.05	6.08	0.00	***
Teacher Social Capital	0.13	0.05	2.37	0.02	**
Factors regressed on Covariates					
Teacher Social Capital on School % Black & Hispanic	-0.19	0.05	-3.81	0.00	***
Teaching Practice Btw on Class Prior Math Achieve	0.02	0.06	0.39	0.70	
Teaching Practice Btw on Class % Black & Hispanic	0.08	0.06	1.37	0.17	
MATH10 ON Covariates					
MATH09	0.76	0.01	107.92	0.00	***
S_BLACK	-0.09	0.01	-11.62	0.00	***
S_HISP	-0.05	0.01	-6.78	0.00	***
S_ELL	-0.02	0.01	-3.09	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20	
GRADE6	-0.01	0.02	-0.79	0.43	
GRADE7	-0.01	0.01	-0.66	0.51	
School % Black & Hispanic	-0.22	0.14	-1.58	0.12	
Class Prior Math Achievement	0.53	0.06	9.03	0.00	***
Class % Black & Hispanic	0.17	0.15	1.17	0.24	
Variances					
Teaching Practice	0.58	0.01	42.50	0.00	
Residual Variances					
Teacher Social Capital	0.97	0.02	52.45	0.00	
Teaching Practice	0.995	0.01	149.10	0.00	
MATH10 Within	0.37	0.01	42.14	0.00	
MATH10 Between	0.63	0.04	14.25	0.00	

*** = significant at p < 0 .01; ** = significant at p < 0.05; * = marginally significant at p < 0.1
RMSEA: 0.033; CFI: 0.846; TLI: 0.839; SRMR Within: 0.053; SRMR Between: 0.119

The first thing to note is that the fit of the unidimensional multilevel confirmatory factor analysis is markedly worse than the fit of the corresponding multidimensional model. CFI (0.846), TLI (0.839), and SRMR Between (0.119) estimates fall short of the conventionally used thresholds for adequate fit. However, the RMSEA (0.033) and the SRMR Within (0.053) values still suggest acceptable fit for the MSEM model employing unidimensional factors.

This unidimensional model suggests that both teaching practices (effect estimate: 0.28) and teacher social capital (0.13) are positively associated with student math outcomes, but that teacher social capital is not significantly related to teaching practices. These results also show that schools with higher proportions of Black and Hispanic students tend to have lower ratings on Teacher Social Capital (-.0.19). Classroom average prior math is strongly associated with the MATH10 outcome (effect estimate: 0.53) while no other covariate effects on the outcome or the factors are significant. These findings are in contrast to the multidimensional results displayed in Table 7.2, which indicate a number of additional significant associations that involve some but not all of the dimensions of these larger constructs of Teacher Social Capital and Teaching Practice.

Results from One-by-One Models

Next, I present the results of alternative models that help to assess the sensitivity of the multidimensional model to multicollinearity while at the same time helping to provide an up-close view of the relationships involving each separate dimension. First, a series of models with one teacher social capital and one teaching practice factor at a time investigate the possible effects of multicollinearity among the teacher social capital factors. In Appendix E, Tables E1 and E2 display a summary of the variances, residual variances, and model fit indices for these models. Model fit indices suggested that the one-by-one models had similar fit to the multidimensional model; residual variances for the factors are large across all models (ranging from 0.85 – 0.999), again suggesting that large portions of the variance in these factors remains unaccounted for regardless of model. Most residual variances are slightly larger in the one-by one models, though the residual variance for Parent is noticeably smaller in the one-by-one models (0.85 as compared to 0.95 in the multidimensional model). The estimated residual variance for Instruction is extremely large across all one-by-one models (0.991 – 0.999), suggesting that these models explain little to none of the variance in this factor.

Structural effect estimates in the one-by-one models are shown in Table 7.4. There are a number of notable differences from the multidimensional results. Regarding Parent Networks, a similar association is found with student math achievement (estimated effect: 0.18 in model with Control; 0.21 in model with Instruction; 0.17 in multidimensional model). But in the one-by-one models, Parent Networks also shows a marginally significant positive association with Control (0.10), an association that is not significant in the multidimensional model. In addition, in the one-by-one models, both Trust and Cares have significant positive direct effects on students' state math test scores (0.13 and 0.14 respectively), whereas these associations are not significant in the multidimensional model. Access to expertise shows a positive association with Control (0.13), in line with the results from the multidimensional model (0.16), though in the one-by-one models there is no significant association with Instruction (estimate in multidimensional model is 0.23). Perhaps most noteworthy, the negative association with student outcomes found in the multidimensional model (-0.20) disappears when factors are considered one by one, leaving no significant association between Expertise and student math.

In the one-by-one models, no teacher social capital factors have significant associations with Instruction. The one-by-one models indicate a similar negative association between school racial composition and teacher social capital factors; schools with greater percentages of disadvantaged minorities have lower levels of teacher social capital. However, the one-by-one models suggest greater variation across dimensions; associations range from -0.08 to -0.40 in one-by-one models while estimates range from -0.20 to -0.24 in multidimensional model; also, the association between School Percent Black and Hispanic and Expertise is not significant in the one-by-one models.

Table 7.4

Structural Effects in Multidimensional Model vs Separate Models with One Factor at a Time; Between Level: (N = 15,644 students; 520 teachers)

	Full Multidimensional Model					Factors one-by-one				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL Between ON										
TRUST	-0.11	0.11	-0.98	0.33		0.06	0.05	1.13	0.26	
CARES	-0.02	0.12	-0.14	0.89		0.08	0.06	1.49	0.14	
PARENT	0.10	0.08	1.25	0.21		0.10	0.05	1.91	0.06	*
EXPERT	0.16	0.10	1.69	0.09	*	0.13	0.05	2.46	0.01	**
INSTRUCT Between ON										
TRUST	-0.05	0.12	-0.38	0.70		0.00	0.06	0.05	0.96	
CARES	-0.10	0.12	-0.81	0.42		0.00	0.06	0.02	0.99	
PARENT	-0.04	0.09	-0.48	0.64		-0.02	0.06	-0.41	0.68	
EXPERT	0.23	0.10	2.37	0.02	**	0.09	0.06	1.59	0.11	
MATH10 ON						$\eta_{1\text{-}4}$ – Control – MATH10				
TRUST	0.04	0.09	0.47	0.64		0.13	0.05	2.55	0.01	**
CARES	0.15	0.11	1.33	0.18		0.14	0.05	2.73	0.01	***
PARENT	0.17	0.08	2.06	0.04	**	0.18	0.06	2.84	0.01	***
EXPERT	-0.20	0.09	-2.15	0.03	**	0.02	0.05	0.40	0.69	
						$\eta_{1\text{-}4}$ – Instruct – MATH10				
TRUST	See above estimates. Model estimated with all factors at same time.					0.13	0.05	2.36	0.02	**
CARES						0.14	0.05	2.62	0.01	**
PARENT						0.21	0.07	3.16	0.00	***
EXPERT						0.02	0.05	0.31	0.76	
FACTORS Regressed on Covariates										
...ON School Black & Hispanic %						$\eta_{1\text{-}4}$ – Control – MATH10				
TRUST	-0.24	0.05	-4.88	0.00	***	-0.16	0.05	-3.10	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.14	0.05	-2.76	0.01	***
PARENTS	-0.22	0.06	-3.95	0.00	***	-0.39	0.05	-8.30	0.00	***
EXPERT	-0.20	0.04	-4.67	0.00	***	-0.08	0.05	-1.60	0.11	
...ON School Black & Hispanic %						$\eta_{1\text{-}4}$ – Instruct – MATH10				
TRUST	See above estimates. Model estimated with all factors at same time.					-0.17	0.05	-3.31	0.00	***
CARES						-0.15	0.05	-3.01	0.00	***
PARENTS						-0.40	0.05	-8.79	0.00	***
EXPERT						-0.08	0.05	-1.60	0.11	

Estimates for direct effects of TSC factors on MATH10 outcome are shown first as the estimates in the models with Control as the mediator and second in the models with Instruction as the mediator. Estimated covariate effects on MATH10 outcome are provided in Appendix E.

One-by-one model estimates regarding covariate effects on MATH10 are provided in Tables E3 and E4 in Appendix E. A negative effect of School Percent Black and Hispanic on

MATH10 is found in the EXPERT-CONTROL-MATH10 (-0.12) model, and a marginally negative effect is found in the CARES-CONTROL-MATH10 (-0.10) model and TRUST-CONTROL-MATH (-0.11) models. The estimated effect was negative but not significant in other one-by-one models.

Results from Four-by-One Models

In addition to estimating this series of models with one teacher social capital and one teaching practice factor at a time, two models with four teacher social capital factors and one teaching practice factor were fit to explore the possible effects of suppression among the teaching practice factors. Model fit indices, displayed below Table 7.5, suggest that the four-by-one models had similar fit to the multidimensional model as well as to the one-by one models. Overall, these indices suggest adequate to good fit, but for the INSTRUCT model, some of the estimates for CFI and TLI fall below the conventionally recommended 0.95 threshold and most indices suggest better fit for the CONTROL model. As shown in Table 7.5, residual variances for the factors are large across all models, again suggesting that large portions of the variance in these factors remains unaccounted for regardless of model. Residual variances in the four-by one models are similar to those found in the multidimensional model.

Table 7.5
Variances, Residual Variances and Model Fit: MSEM Multidimensional vs. Four-by-one Models

	MSEM Multidimensional				Factors 4x1			
	Est	SE	Est/SE	p-val	Est	SE	Est/SE	p-val
Variances								
CONTROL	0.61	0.01	45.21	0.00	0.50	0.02	31.77	0.00
INSTRUCT	0.50	0.02	31.81	0.00	0.60	0.01	45.07	0.00
Residual Variances					TSC – CONTROL – MATH10			
TRUST	0.94	0.02	38.92	0.00	0.94	0.02	39.29	0.00
CARES	0.96	0.02	54.62	0.00	0.96	0.02	55.47	0.00
PARENTS	0.95	0.03	38.24	0.00	0.95	0.03	38.69	0.00
EXPERT	0.96	0.02	57.57	0.00	0.96	0.02	57.89	0.00
CONTROL Btw	0.91	0.03	32.90	0.00	0.91	0.03	32.80	0.00

	Est	SE	Est/SE	p-val	Est	SE	Est/SE	p-val
MATH10 Within	0.37	0.01	42.14	0.00	0.37	0.01	42.16	0.00
MATH10 Btw	0.54	0..04	13.54	0.00	0.54	0.04	13.48	0.00
					TSC – INSTRUCT – MATH10			
TRUST	See above estimates. Model estimated with all factors.				0.94	0.02	38.92	0.00
CARES					0.96	0.02	54.70	0.00
PARENTS					0.95	0.03	38.25	0.00
EXPERT					0.96	0.02	57.67	0.00
INSTRUCT Btw	0.98	0.02	56.20	0.00	0.98	0.17	57.35	0.00
MATH10 Within					0.37	0.01	42.10	0.00
MATH10 Btw					0.61	0.05	13.46	0.00

Multidimensional Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.07
CONTROL Model Fit: RMSEA: 0.011; CFI: 0.976; TLI: 0.972; SRMR Within: 0.019; SRMR Between: 0.072.
INSTRUCT Model Fit: RMSEA: 0.023; CFI: 0.938; TLI: 0.933; SRMR Within: 0.035; SRMR Between: 0.067.
Variances restricted as invariant across levels. Largest and smallest standardized residual variance estimates shown.

Results from the four-by-one models are compared to estimates from the multidimensional model in Tables 7.6 and 7.7. Results from the model with only Control largely mirror the full multidimensional model results. However, in the model with only Instruction (see Table 7.7), a drastic difference is seen.

Table 7.6
Structural Effects in Multidimensional Model vs Models with Four Teacher Social Capital Factors–by-CONTROL Factor (N = 15,644 students; 520 teachers)

	Full Multidimensional Model					Factors Four-by-one				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL Between ON						4 TSC – Control – MATH10				
TRUST	-0.11	0.11	-0.98	0.33		-0.11	0.11	-1.00	0.32	
CARES	-0.02	0.12	-0.14	0.89		-0.01	0.12	-0.10	0.93	
PARENT	0.10	0.08	1.25	0.21		0.10	0.08	1.35	0.18	
EXPERT	0.16	0.10	1.69	0.09	*	0.15	0.10	1.58	0.11	
MATH10 ON										
CONTROL Btw	0.37	0.07	5.52	0.00	***	0.39	0.05	8.49	0.00	***
TRUST	0.04	0.09	0.47	0.64		0.04	0.09	0.47	0.64	
CARES	0.15	0.11	1.33	0.18		0.15	0.11	1.35	0.18	
PARENT	0.17	0.08	2.06	0.04	**	0.17	0.08	2.11	0.04	**
EXPERT	-0.20	0.09	-2.15	0.03	**	-0.20	0.09	-2.19	0.03	**
FACTORS Regressed on Covariates										
...ON School Black & Hispanic %										
TRUST	-0.24	0.05	-4.88	0.00	***	-0.24	0.05	-4.84	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.20	0.04	-4.40	0.00	***
PARENTS	-0.22	0.06	-3.95	0.00	***	-0.22	0.06	-3.93	0.00	***

EXPERT	-0.20	0.04	-4.67	0.00	***	-0.20	0.04	-4.63	0.00	***
…ON Classroom Prior Achievement										
CONTROL Between	0.24	0.05	5.40	0.00	***	0.24	0.05	5.45	0.00	***
MATH10 ON										
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	107.91	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.55	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.71	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.06	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.29	0.20	
GRADE6	0.01	0.02	0.61	0.54		0.01	0.01	0.74	0.46	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	-0.20	0.84	
School % Black & Hisp	-0.07	0.06	-1.18	0.24		-0.07	0.06	-1.20	0.23	
Class Prior Math Achieve	0.39	0.06	6.83	0.00	***	0.38	0.06	6.94	0.00	***

Whereas the multidimensional model shows a significant effect of Control on student math achievement but no effect of Instruction, with the factors included four teacher social capital-by-one teaching practice, both Control and Instruction show highly significant positive associations with student math outcomes. The estimated effect of Instruction is about three-tenths of a standard deviation (0.30), somewhat smaller than the estimated effect of Control (0.39), but approximately fifteen times larger than the effect estimate for Instruction in the multidimensional model (0.02). In both of the four-by-one models, estimated effects of teacher social capital factors on MATH10 are similar to results from the multidimensional model.

Table 7.7

Structural Effects in Multidimensional Model vs Models with Four Teacher Social Capital Factors–by-INSTRUCTION Factor (N = 15,644 students; 520 teachers)

	Full Multidimensional Model					Factors Four-by-one				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
INSTRUCT Between ON						4 TSC – Instruct – MATH10				
TRUST	-0.05	0.12	-0.38	0.70		-0.05	0.12	-0.36	0.72	
CARES	-0.10	0.12	-0.81	0.42		-0.10	0.12	-0.83	0.40	
PARENT	-0.04	0.09	-0.48	0.64		-0.04	0.09	-0.48	0.63	
EXPERT	0.23	0.10	2.37	0.02	**	0.22	0.10	2.30	0.02	**
MATH10 ON										

INSTRUCT Between	0.02	0.07	0.27	0.79		0.30	0.05	6.24	0.00	***
TRUST	0.04	0.09	0.47	0.64		0.01	0.10	0.12	0.91	
CARES	0.15	0.11	1.33	0.18		0.15	0.12	1.26	0.21	
PARENT	0.17	0.08	2.06	0.04	**	0.23	0.09	2.61	0.01	**
EXPERT	-0.20	0.09	-2.15	0.03	**	-0.22	0.10	-2.12	0.03	**
FACTORS Regressed on Covariates										
...ON School Black & Hispanic %										
TRUST	-0.24	0.05	-4.88	0.00	***	-0.25	0.05	-4.96	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.20	0.04	-4.51	0.00	***
PARENT	-0.22	0.06	-3.95	0.00	***	-0.22	0.06	-4.02	0.00	***
EXPERT	-0.20	0.04	-4.67	0.00	***	-0.20	0.04	-4.75	0.00	***
…ON Classroom Prior Achievement										
INSTRUCT Between	-0.03	0.05	-0.54	0.59		-0.03	0.05	-0.55	0.58	
MATH10 ON										
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	107.91	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.55	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.71	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.06	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.29	0.20	
GRADE6	0.01	0.02	0.61	0.54		0.01	0.01	0.74	0.46	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	-0.20	0.84	
School % Black & Hisp	-0.07	0.06	-1.18	0.24		-0.07	0.06	-1.20	0.23	
Class Prior Math Achieve	0.39	0.06	6.83	0.00	***	0.38	0.06	6.94	0.00	***

Discussion

This chapter examined the relationships among teacher social capital, teaching practice, and student achievement through a multilevel multidimensional latent mediation model, and provided sensitivity checks to these results by way of a unidimensional model and a series of models with the four teacher social capital factors and two teaching practice factors inserted one-at-a-time. With the factors included one-by-one or four-by-one, the estimates in the multidimensional model that ran counter to theory disappeared, reinforcing the idea that some of the multidimensional estimates may be artifacts of multicollinearity/suppression effects rather than substantively interpretable findings.

Across all models, schools with higher percentages of Black and Hispanic students have lower ratings on teacher social capital factors, by about one-fifth of a standard deviation in the multidimensional and unidimensional models. In schools serving larger percentages of Black and Hispanic students, teachers tend to rate their school as less trusting and their administration as putting less effort into meeting teachers' needs. These teachers have less confidence in the quality of the school's communication and involvement efforts with parents, and rate their schools as doing a worse job in providing quality professional development and support / Expertise to teachers. Schools with high percentages of Black and Hispanic students tend to also serve larger percentages of socially disadvantaged students (Fiel, 2013), but the models here do not account for varying levels of student poverty across schools so distinctions between impacts of race and class cannot be estimated.

In considering relationships among the constructs, the largest and most consistent effect estimate is that teacher's Control is positively associated with students' math outcomes. These associations are moderate in magnitude, over one-third of a standard deviation and reinforce earlier MET findings and findings from the broader literature that have shown a strong relationship between teachers' classroom management/control skills and students' outcomes (e.g., Brophy, 1984; Ferguson & Danielson, 2014; Kane & Cantrell, 2010; Raudenbush & Jean, 2014). Read straightforwardly, this association suggests that having a teacher one standard deviation above the mean on Control predicts a classroom with student math outcomes one-third of a standard deviation higher than the mean. This finding also reflects positive associations between student behavior and student academic outcomes. The extent to which students behave respectfully and stay on task represents complex interactions between students and teacher. Factors other than the teacher, particularly parents, school and community culture, and school leaders, also impact students' behavior. Importantly, both Control and student achievement are conditioned on classroom average prior achievement. A high-achieving classroom is also more likely to be filled with motivated, attentive, and well-behaved students. In other words, it is likely to be what teachers refer to simply as a "good class." Therefore, the association between Control and end-of-year student achievement is observed even after controlling for the effects of "good classes" on teachers' levels of control and the effects of these "good classes" on students' end-of-year achievement.

Students' perceptions of Instructional quality are also related to improved student math outcomes, but only in models which do not include Control. These results may help to explain the puzzling results found in Ferguson & Danielson's (2014) MET study. Based on a multiple regression predicting classroom average test score growth from Tripod composite factors, they conclude that, when holding constant Control and Challenge, the predicted result of increasing teachers' Support (measured by a composite of the 5 Cs other than Challenge and Control) is to decrease the students' expected test scores. Ferguson and Danielson take off from their findings to speculate that

> if a teacher is already quite challenging and the class is almost always well behaved and on task, then becoming more supportive might actually lower the sense of urgency in the class and some students might relax and learn less (p. 121),

but the findings shown here suggest a simpler explanation: Control and Instruction are correlated (0.65 at the Between level in MCFA) and the measurement of Control is more reliable than the measurement of Instruction (residual variances, factor loadings, and ICCs all suggest more measurement error for the Instruction factor than the Control factor). Therefore, it may not be possible to disentangle the positive effects of Instruction on student test outcomes from the positive effects of strong Control. In other words, the null estimate found in the multidimensional model here, and the negative estimates found by Ferguson and Danielson, may be explained as a result of suppression producing estimates that cannot be meaningfully interpreted rather than as a result of imagined negative effects of students' relaxing from receiving feedback that is too helpful, explanations that are too clear, or emotional understanding that is too supportive.

With regard to the relationships involving teacher social capital, the strength of a school's Parental Networks is directly and positively associated with student math outcomes. When teachers say their school provides clear communication to parents and encourages parental involvement, and rate parents as well-informed about school happenings, stronger student math outcomes are predicted. This finding reinforces the findings from parental involvement literature that suggest that 1) the efforts of teachers and other school personnel can influence the level and quality of parental involvement (Bermúdez & Márquez, 1996; Epstein, 1988), and 2) the level of parental involvement is positively associated with improved student outcomes (Fan & Chen, 2001; Sui-Chi & Wilms, 1996).

In the models with one teacher social capital factor included at a time, strong Parent Networks are directly associated with student outcomes and with Control, which is itself related to positive student outcomes. This positive association between Parent Networks and Control might be taken as a relationship between parental *communication* and *teachers'* levels of classroom management/control, or a relationship between parental *involvement* and *students'* levels of behavior. Parent Networks might be interpreted as reflecting the strength of a school's efforts and success in communicating to parents; but, it could also reflect the strength of the parents at that neighborhood school who are perhaps actively involved, avid boosters, etc., Similarly, high ratings on Control might be interpreted as suggestive of strong classroom management skills by teachers, but also could have their origin in respectful student behavior which perhaps is itself a product of strong parenting.

Next, teachers' Access to Expertise is positively associated with greater degrees of classroom control in all models. These findings may suggest that Expert support helps teachers achieve stronger management, but may also in part reflect a positive association between schools that provide quality professional support and classrooms that exhibit high levels of student behavior – in other words well-run schools might on average demonstrate both higher ratings on Expertise and higher ratings on Control while some other factor associated with well-run schools might be driving both of those high ratings.

Expertise is also associated with higher quality Instruction in the multidimensional model. The latter result suggests that in schools with equal levels of Trust, Caring, and Parental Networks, providing teachers with Access to Expert professional support is associated with higher ratings of Instruction. Literature suggests that professional development efforts produce few or uneven benefits to students (Jacobs & Lefgren, 2004; Kennedy, 1999b) and most professional development is low intensity, without continuity or quality (Corcoran, 1995; Little, 1993). Instances of high-quality professional development produce measurable impacts on student outcomes (Angrist & Lavy, 2001), but the more consistent finding is that focused professional development can positively impact teachers' practices (Desimone Porter, Garet, Yoon, & Birman, 2002). The findings here add a bit more weight to this conclusion – giving teachers access to high quality support appears associated with higher quality classroom practices. However, the evidence here is only suggestive; the effect estimate is not significant when the factors are included one-by-one, and thus the multidimensional estimate may reflect multicollinearity effects rather than substantively meaningful impacts of Access to Expertise on instructional practice. The estimates regarding Access to Expertise should be interpreted cautiously because of the possibility that they reflect multicollinearity effects.

One last finding regarding teacher social capital bears mention: one-by-one models suggest that the norms of a school (levels of Trust among teachers and the extent to which the administration Cares about teacher concerns) are positively associated with students' math outcomes. In a school in which teachers trust one another and perceive the school leadership as acting in their interests, students perform better on state math tests, perhaps due to greater confidence and/or lessened anxiety contributing to learning more or simply performing better on high-stakes tests. These results might be interpreted as modest reinforcement for the previously discussed findings that schools with high trust are more likely to demonstrate large gains in student learning (Bryk and Schneider, 2003). However, when all teacher social capital factors are included in the multidimensional model, only the Parent Network factor shows a significant effect on student outcomes. This finding, together with the high correlations among all four teacher social capital factors, suggest that Trusting and Caring norms are associated with strong communication with Parents, and this networking with Parents predicts positive student math outcomes, a conclusion aligned with Bryk & Shneider's (1996) conclusion that schools with more relational trust are associated with teachers more willing to involve parents.

The following chapters investigate how the relationships among these constructs change when models and measures are changed.

Chapter VIII.

Results for Multilevel vs. Single-level Models

In this chapter, the multilevel multidimensional structural equation mediation model presented in Chapter VII is compared to a structural equation model estimated in a single-level to address research question 4a: How do the relationships among teacher social capital, teaching practice, and student achievement change when estimated with a multilevel structural equation model versus a single-level structural equation model?

When data is nested, as in the situation here with students nested within classrooms, a single-level model is known to lead to downwardly biased standard errors because the interdependencies among the clustered observations are ignored (Lüdtke et al, 2011; Raudenbush & Bryk, 2002). In addition, single-level models conflate any differences between level two and level one effects, possibly resulting in the occurrence of the well-known ecological fallacy, in which relationships seen for groups are incorrectly assumed to hold for individuals (Kaplan & Elliot, 1997b). The comparisons in this chapter provide an empirical test of the extent to which inferences about the structural relationships among these constructs are dependent on modeling assumptions regarding the clustering of data. Further detail on these models is provided in chapter V and equations and Mplus syntax are provided in Appendix D.

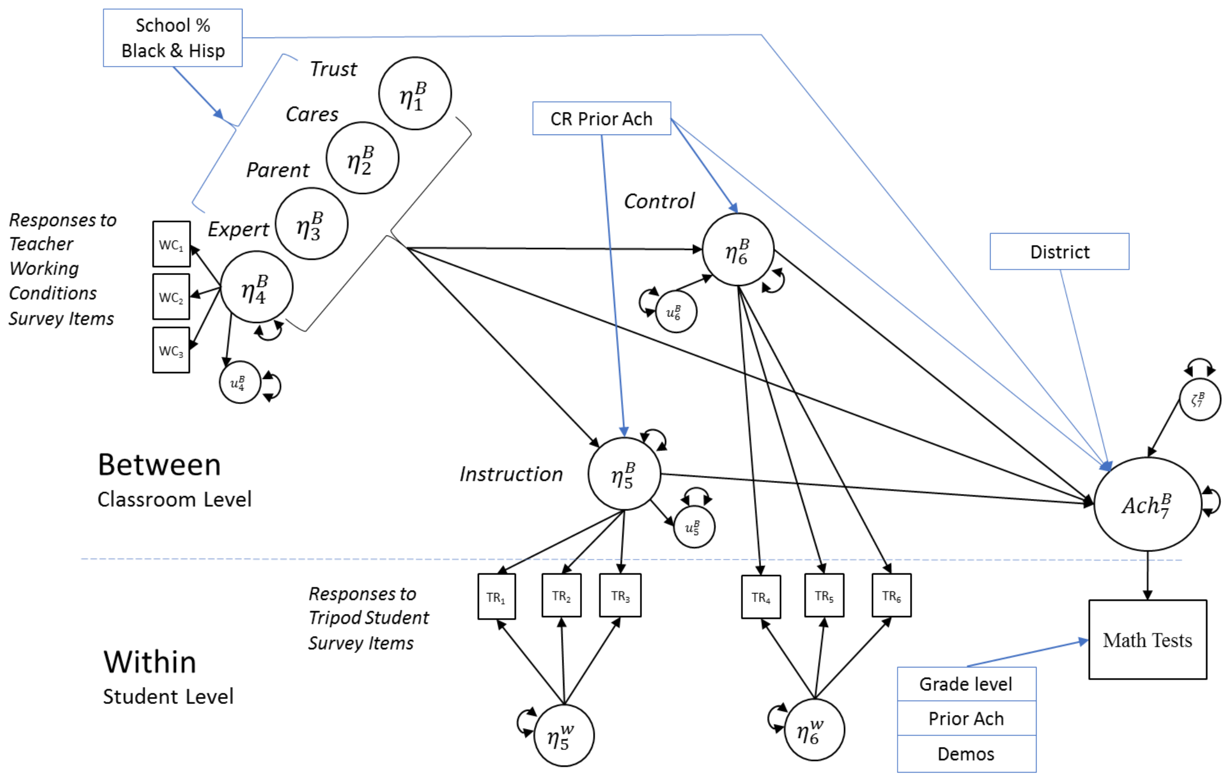

Figure 8.1. Multilevel Multidimensional Structural Equation Model (MSEM). The 4 Teacher Social Capital factors are correlated, as are the 2 Teaching Practice factors. The model was estimated using state math tests as the outcome.

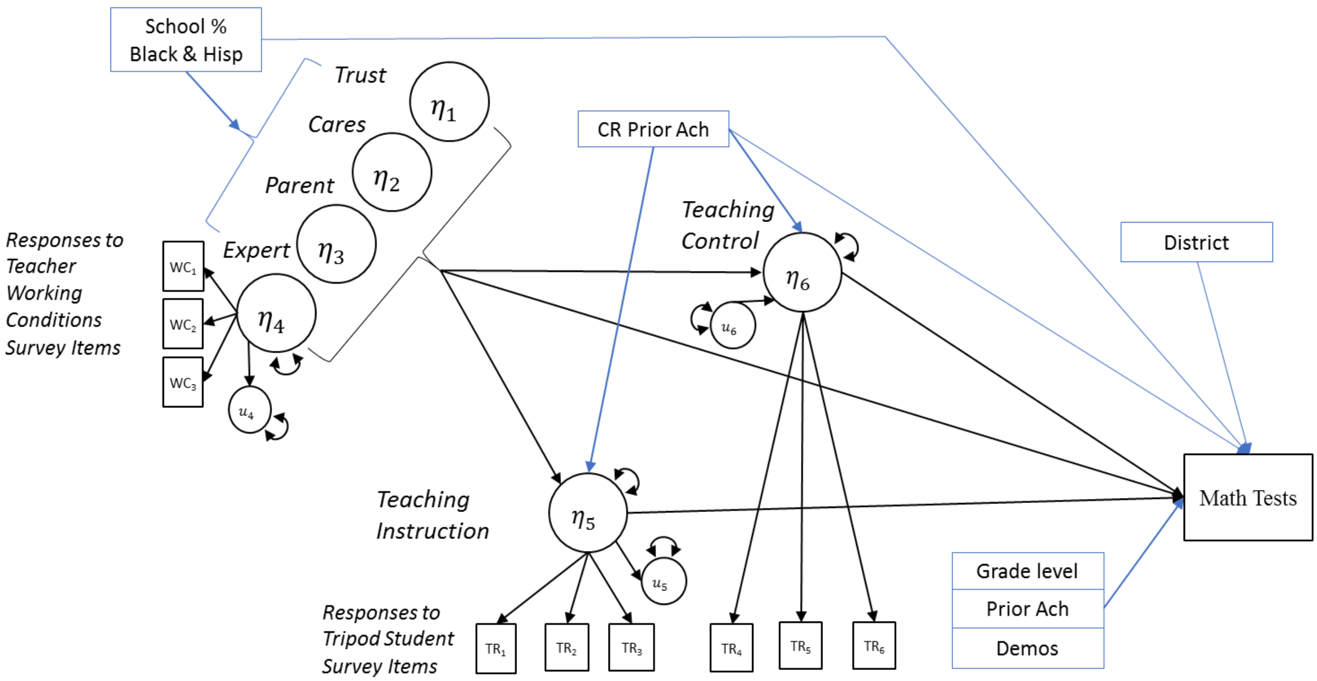

Figure 8.2. Single-Level Multidimensional Structural Equation Model (SEM). The 4 Teacher Social Capital factors are correlated, as are the 2 Teaching Practice factors. The model was estimated using state math tests as the outcome.

I present results from the MSEM Model and the SEM model below in Tables 8.1 and 8.2. Structural effect estimates displayed for the MSEM Model are at the Between level. Covariates' effects on outcome variable were all significant at <.01 level, except Male, grade level, and school percent Black and Hispanic. D district effect estimatesare not available for release due to the MET data use agreement.

Model fit indices suggest adequate fit for the multilevel model; the CFI and TLI estimates are about 0.93, somewhat below the conventional 0.95 threshold indicating good fit; the estimates for RMSEA (0.02) and SRMR Within (0.04) and Between (0.07) suggest acceptable to good fit. The fit of the single-level model appears somewhat worse by way of CFI and TLI estimates (0.91 and 0.90), but still adequate in terms of the estimates for RMSEA (0.04) and SRMR (0.05). Residual variances are large for all factors across models, suggesting a great deal of variance in these factors remains unaccounted for regardless of model. The residual variance for the CONTROL factor is notably larger in the single-level model, suggesting that the multilevel model better accounts for the variance in CONTROL.

Table 8.1
Variances, Residual Variances, and Model Fit: MSEM vs. SEM Models

	MSEM				SEM			
	Est	SE	est/SE	p-val	Est	SE	est/SE	p-val
Variances								
CONTROL	0.61	0.01	45.21	0.00				
INSTRUCT	0.50	0.02	31.81	0.00				
Residual Variances								
TRUST	0.94	0.02	38.92	0.00	0.94	0.00	213.81	0.00
CARES	0.96	0.02	54.62	0.00	0.96	0.00	295.42	0.00
PARENTS	0.95	0.03	38.24	0.00	0.95	0.01	211.93	0.00
EXPERT	0.96	0.02	57.57	0.00	0.96	0.00	323.95	0.00
CONTROL	0.91	0.03	32.90	0.00	0.97	0.00	302.98	0.00
INSTRUCT	0.98	0.02	56.20	0.00	0.99	0.00	597.72	0.00
MATH10 Within	0.37	0.01	42.14	0.00	0.34	0.01	66.97	0.00
MATH10 Between	0.54	0..04	13.54	0.00	No separate between and within models			

MSEM Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.07
SEM Model Fit: RMSEA: 0.040; CFI: 0.906; TLI: 0.900; SRMR: 0.049. Residual variances are standardized values.
In MSEM Model, variances for CONTROL and INSTRUCT estimated as invariant across levels.

The results from both models suggest the same inferences in a number of areas. First, the estimated effect of the school percent Black and Hispanic on teacher social capital factors is negative for all factors. These effects are about one-fifth of a standard deviation in magnitude and statistically significant at $p < 0.01$, suggesting that schools with higher percentages of Black and Hispanic students tend to have lower ratings on teacher social capital factors. Also, teacher's Control is stronger in classrooms with higher average prior achievement (effect estimate: 0.24 in MSEM; 0.14 in SEM), but there is no significant association between Classroom Average Prior Achievement and the quality of Instruction. Results also indicate that Control has a significant effect on student math outcomes (0.37 in MSEM; 0.04 in SEM). And, Access to Expertise has a significant or marginally significant direct effect on both Control (0.16 in MSEM; 0.08 in SEM) and Instruction (0.23 in MSEM: 0.11 in SEM). Counter to theory, the results across both models also indicate that Access to Expertise has a significant negative direct effect on student state math test scores (-0.20 in MSEM; -0.04 in SEM). This counterintuitive negative estimate suggests that both models suffer from multicollinearity/suppression among the factors, a hypothesis that was discussed in the prior chapter; the present investigation is concerned primarily with how inferences change from the multilevel to the single-level model.

Table 8.2

Structural Effects in Multilevel (MSEM) Model (Between Level) vs Single-Level (SEM) Model. N = 15,644 students; 520 teachers

	MSEM					SEM				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL ON										
TRUST	-0.11	0.11	-0.98	0.33		-0.06	0.02	-2.65	0.01	***
CARES	-0.02	0.12	-0.14	0.89		-0.01	0.02	-0.61	0.54	
PARENT	0.10	0.08	1.25	0.21		0.07	0.02	4.34	0.00	***
EXPERT	0.16	0.10	1.69	0.09	*	0.08	0.02	4.36	0.00	***
INSTRUCT ON										
TRUST	-0.05	0.12	-0.38	0.70		-0.02	0.02	-1.08	0.28	
CARES	-0.10	0.12	-0.81	0.42		-0.04	0.02	-2.08	0.04	**
PARENT	-0.04	0.09	-0.48	0.64		-0.04	0.02	-2.39	0.02	**
EXPERT	0.23	0.10	2.37	0.02	**	0.11	0.02	6.59	0.00	***
MATH10 ON										
CONTROL	0.37	0.07	5.52	0.00	***	0.04	0.01	6.13	0.00	***
INSTRUCT	0.02	0.07	0.27	0.79		0.04	0.01	6.49	0.00	***
TRUST	0.04	0.09	0.47	0.64		0.01	0.01	0.57	0.57	
CARES	0.15	0.11	1.33	0.18		0.03	0.01	2.72	0.01	***
PARENT	0.17	0.08	2.06	0.04	**	0.04	0.01	4.32	0.00	***
EXPERT	-0.20	0.09	-2.15	0.03	**	-0.04	0.01	-4.12	0.00	***
FACTORS Regressed on Covariates										
...ON School % Black & Hispanic										
TRUST	-0.24	0.05	-4.88	0.00	***	-0.24	0.01	-25.65	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.20	0.01	-23.40	0.00	***
PARENTS	-0.22	0.06	-3.95	0.00	***	-0.21	0.01	-20.30	0.00	***
EXPERT	-0.20	0.04	-4.67	0.00	***	-0.19	0.01	-24.43	0.00	***
…ON Classroom Prior Achievement										
CONTROL	0.24	0.05	5.40	0.00	***	0.14	0.01	16.27	0.00	***

INSTRUCT	-0.03	0.05	-0.54	0.59		0.00	0.01	-0.35	0.73	
MATH10 ON Covariates										
MATH09	0.76	0.01	107.95	0.00	***	0.69	0.01	103.25	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-13.19	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.04	0.01	-6.82	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.01	0.01	-2.54	0.01	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.35	0.18	
GRADE6	0.01	0.02	0.61	0.54		0.00	0.01	0.64	0.52	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	0.74	0.46	
School % BlHisp	-0.07	0.06	-1.18	0.24		-0.01	0.01	-1.85	0.06	*
Class Prior Ach	0.39	0.06	6.83	0.00	***	0.12	0.01	16.10	0.00	***

*** = significant at p < 0 .01; ** = significant at p < 0.05; * = marginally significant at p < 0.1

The results also suggest notable differences across models in the standard errors and in the estimated magnitude of effects. First, standard errors in the single-level model are systematically smaller (ranging from 0.01 – 0.02) than the standard error estimates in the multilevel model (0.04 – 0.12). This is as expected because single-level models are known to underestimate standard errors due to ignoring the dependencies among individuals clustered in groups.

The smaller standard errors in the single-level model are largely responsible for the fact that there are many more statistically significant parameter estimates in the SEM results than in the MSEM results. For instance, in the multilevel model, no significant effects were found for the levels of Trust among teachers nor for the extent to which the administration Cares about teacher concerns. In the single-level SEM, on the other hand, I find a significant positive direct effect of Cares on student state math test outcomes (0.03) and a negative effect of Cares on Instruction (-0.04). In addition, Trust shows a negative effect on teachers' Control (-0.06). Both models find that Parent Networks has a significant positive effect on student math outcomes (0.17 in MSEM; 0.04 in SEM), but the SEM model also shows a positive effect of Parent on teachers' Control (0.07) and a negative effect on teachers' Instruction (-0.04). Finally, in the multilevel model, Instruction has no significant effect on student math outcomes but in the single-level model this effect is significant (0.04). In brief, the inferences regarding Instruction, Cares, Trust, and to some degree Parent Networks vary from model to model.

Another key comparison is of the magnitude of effect estimates across MSEM versus SEM models. At first glance, effect estimates of Teacher Social Capital factors on Teaching Practice factors and on student outcomes are similar in direction and magnitude across models. The similarities are even more apparent when examining the unstandardized coefficients (see Table 8.3 below). However, the magnitude of Teaching Practice effect estimates on student outcomes differs meaningfully across MSEM and SEM models. Comparing the unstandardized estimates, the MSEM model estimates an effect of about 0.17 for Control and 0.01 for Instruction on MATH10. The SEM estimates an effect of about 0.04 for Control and 0.04 for Instruction. The MSEM ascribes nearly all of the effect on the outcome to the Control factor, whereas the SEM ascribes the effect to both Control and Instruction.

Table 8.3
Unstandardized Effect Estimates for MSEM and SEM Models

	MSEM					**SEM**				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL										
TRUST	-0.12	0.12	-0.98	0.33		-0.10	0.04	-2.65	0.01	***
CARES	-0.02	0.10	-0.14	0.89		-0.02	0.03	-0.61	0.54	
PARENT	0.09	0.08	1.24	0.22		0.10	0.02	4.32	0.00	***
EXPERT	0.14	0.08	1.68	0.09	*	0.11	0.03	4.35	0.00	***
INSTRUCT										
TRUST	-0.04	0.11	-0.38	0.71		-0.04	0.04	-1.07	0.28	
CARES	-0.08	0.09	-0.81	0.42		-0.07	0.03	-2.08	0.04	**
PARENT	-0.03	0.07	-0.48	0.63		-0.06	0.02	-2.40	0.02	**
EXPERT	0.17	0.07	2.36	0.02	**	0.16	0.03	6.57	0.00	***
MATH10 ON										
CONTROL	0.17	0.03	5.14	0.00	***	0.04	0.01	6.12	0.00	***
INSTRUCT	0.01	0.04	0.27	0.79		0.04	0.01	6.48	0.00	***
TRUST	0.02	0.04	0.47	0.64		0.01	0.02	0.57	0.57	

CARES	0.06	0.04	1.32	0.19		0.05	0.02	2.71	0.01	**
PARENT	0.07	0.04	2.08	0.04	**	0.07	0.02	4.33	0.00	***
EXPERT	-0.08	0.04	-2.16	0.03	**	-0.06	0.02	-4.13	0.00	***

*** = significant at p < 0 .01; ** = significant at p < 0.05; * = marginally significant at p < 0.1

Discussion

This chapter set out to examine how inferences regarding the relationships among teacher social capital, teaching practice, and student achievement change depending on whether clustering of students within classrooms is taken into account or ignored. Using data drawn from almost 16,000 students in grade 6 – 8 math classrooms, I compared the results from a multilevel multidimensional structural equation mediation model to the results from a single-level model that was in all other respects equivalent to the first model.

It is well known that ignoring the nesting of data leads to downwardly biased standard errors (Raudenbush & Bryk, 2002). In addition, Muthén (1994) and Morin and Marsh and colleagues (2014) discuss an additional advantage that multilevel structural equation models enjoy over their single-level SEM brethren when data are nested: multilevel SEM models separate structural estimates into Between and Within portions, whereas single-level SEM models conflate these structural estimates. The extent to which these issues matter in empirical examples has been explored much less fully. This chapter provided an example in which the differences between MSEM and SEM models result in meaningfully different inferences in regard to four of the six latent factors included in the model. I found evidence that the inferences drawn about Instruction, Trust, Cares, and Parent Networks differ substantially from the MSEM model to the SEM model.

The different parameter estimates stem first from systematically smaller standard errors in the SEM model for all variables (0.01 – 0.02 in SEM; 0.04 – 0.12 in MSEM). This is to be expected because the single-level model ignores the nesting of students within classrooms/teachers, and thus will tend to underestimate standard errors. Due to the smaller standard errors, the single-level model shows a number of significant effect estimates that are not significant in the multilevel models.

The smaller standard errors in the SEM model lead to changes in the inferences regarding several teacher social capital factors. The strength of a school's Parental Networks is associated with higher student state math test outcomes in both models (0.17 in MSEM; 0.04 in SEM), in accord with research that has shown that parental involvement is associated with positive student outcomes (Fan & Chen, 2001; Sui-Chi & Wilms, 1996) and that schools can impact the level of parental involvement (Bermúdez & Márquez, 1996; Epstein, 1988). However, the single-level model also shows a significant positive association between Parent Networks and Control (0.07) along with a negative association between Parent Networks and Instructional quality (-0.04), associations which are not significant in the multilevel model. Both models produce negative estimates for Trust and Cares on Control and Instruction, but while none of these estimates are significant in the multilevel model, the negative associations between Trust and Control (-0.06) and between Cares and Instruction (-0.04) are statistically significant in the SEM model.

The second source of the differences involves the magnitude of the effect estimates. The estimated directions of the effects are mainly the same across the multi- and single-level models, and the magnitude of the estimated effects are also mainly similar when the unstandardized coefficients are examined. However, there is one important exception. Teaching Practice effect estimates on student outcomes differ across MSEM and SEM models. Comparing the unstandardized estimates, the MSEM model estimates an effect of about 0.17 for Control and 0.01 for Instruction on MATH10. The SEM estimates an effect of about 0.04 for Control and

0.04 for Instruction. There is a marked shift in how the models apportion the effect of teaching practice on student math achievement: the multilevel model ascribes nearly all of the effect to the Control factor, whereas the SEM ascribes the effect about equally to both Control and Instruction.

In the multilevel model, the effect of Control on student math outcomes (0.37) is statistically significant and moderate in magnitude, approaching one-third of a standard deviation. These estimates appear to echo the findings from earlier MET studies and the broader literature on the effects of teaching practices on student outcomes (Brophy & Good, 1984; Ferguson & Danielson, 2014; Shouse, 1996). However, in the single-level model, the estimated effect of Control is much smaller than in the multilevel model (0.04). It is also notable that the estimated residual variance for Control is markedly larger in the single-level model (0.97) than in the multilevel model (0.91), suggesting that the multilevel model better accounts for the variance of the Control factor.

Meanwhile, inferences regarding the estimated impact of Instructional quality on student math outcomes vary even more depending on the model. In the multilevel model, the estimated effect of Instruction on student math achievement is much smaller in magnitude than the effect of Control (0.01 compared to 0.17 in the unstandardized estimates) and not statistically significant. However, in the single-level model, the effect estimates for both Control and Instruction are similar in magnitude (about .04, unstandardized) and the smaller standard errors make both estimates significant at $p < 0.01$. These differences across models bring to mind two considerations: First, the two factors of Control and Instruction are strongly correlated (MCFA correlation Within: 0.54; Between: 0.65; MSEM correlation Within: 0.54; Between: 0.68; SEM correlation: 0.58). Second, the single-level model conflates the Within and Between variances of each of these factors and the Within and Between covariances/correlations. In these results, it appears that one consequence of this conflation is that the single-level model attributes the effects of Teaching Practice on student math outcomes to both Control and Instruction factors while the multi-level model attributes the Between classroom effects of Teaching Practice on student math outcomes to the Control factor almost exclusively. In the MSEM model nearly all of the effect on student math achievement is attributed to the Control factor; in the SEM model, the effect on student math achievement is apportioned about equally to Control and Instruction.

In part, these varying effect estimates may also reflect underlying weaknesses in the measures of Instructional quality discussed in previous chapters and therefore may reinforce the difficulties many researchers have found in attempting to reliably and validly measure classroom instruction (e.g., Thomas, 1929 in Chavez, 1984; Correnti & Martinez, 2012). In chapter VI, based on ICCs and factor loadings, I found that Control is measured more reliably at the classroom level than Instruction. This finding suggests that students are better able to measure the teacher's control of the classroom/appropriateness of student behavior than the quality of a teacher's feedback, the clarity of explanations, the sincerity of caring, and the like. The multilevel model finds stronger relationships regarding the factor that is measured more accurately at the Between level, namely CONTROL. The single-level model does not make this distinction – the Within and Between level effects are conflated, and the estimates in the SEM model show approximately equal effects of Control and Instruction on student math outcomes.

Because SEM models are known to suffer from biased standard errors and to conflate effect estimates when used to estimate relationships involving nested data, and because the evidence here suggests that the MSEM model provides better fit to the data (e.g., RMSEA: 0.02 in MSEM, 0.04 in SEM; CFI: 0.93 in MSEM, 0.91 in SEM) and in particular explains a greater portion of the variance in the Control factor (residual variance for Control: 0.91 in MSEM, 0,97 in SEM), caution should be exercised in interpreting substantive importance to the results from the SEM model here; because this empirical example demonstrates that the inferences we draw can vary greatly when nesting is ignored, future researchers should exercise caution in applying single-level models to nested data structures.

Chapter IX.

Results for Latent vs. Manifest Models

This chapter addresses research question 3b: How do the relationships among the constructs change when estimated with a multilevel structural equation model that uses latent factors to represent the social capital and teaching practice constructs (see Figure 9.1) versus a model that uses manifest (observed) indicators (see Figure 9.2)? The latent model estimates the measurement error that is a part of the observed item indicators and incorporates those estimates into its calculation of the effect estimates. In the manifest model, the extent of that measurement error is unknown and ignored. For the manifest model shown in Figure 9.2, the composite indicators for the four teacher social capital factors (Trust, Cares, Parent, and Expert), and the two teaching practice factors (Control and Instruction) are formed from simple averages of the items found to load on each factor in the multilevel confirmatory factor analyses discussed in Chapter VI. Those composite indicators are then brought into the model in place of the latent factors, and the observed item indicators are not included. As a consequence, the manifest model does not take into account the measurement error in the observed survey item responses and instead treats the composite indicators as if they were observed without error.

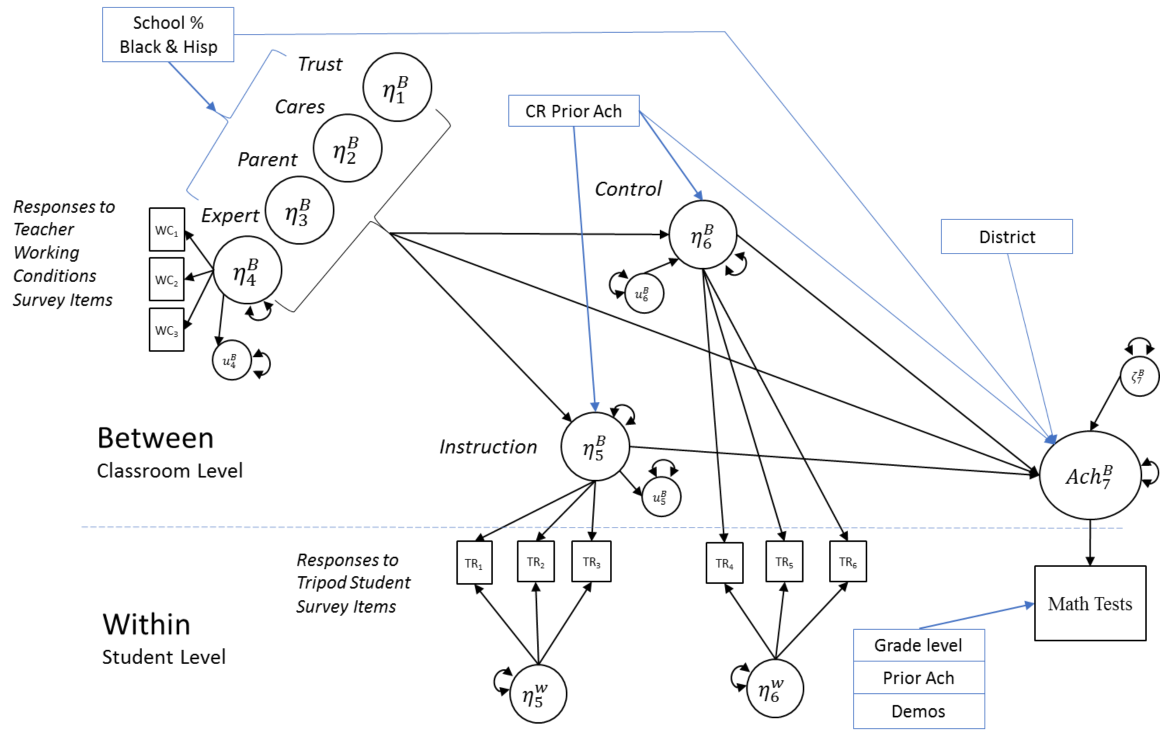

Figure 9.1. Multilevel Structural Equation Model (Latent Model). Teacher Social Capital factors are correlated, as are Teaching Practice factors. State math tests used as the outcome.

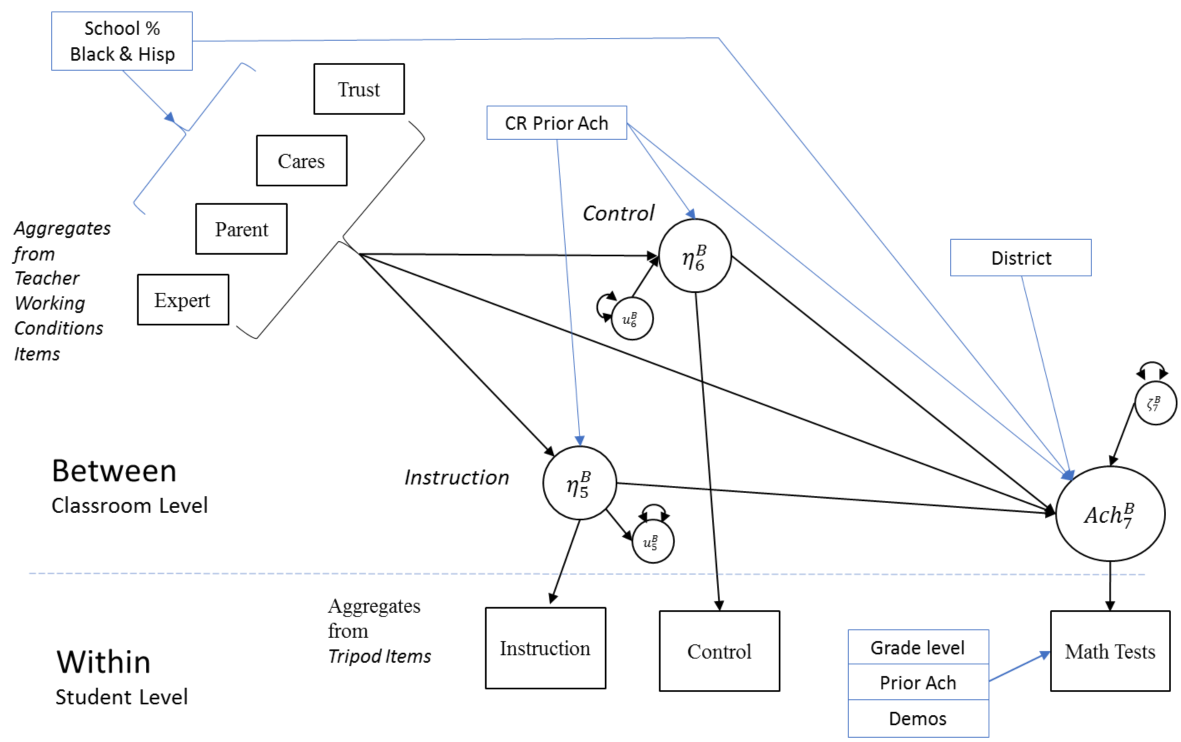

Figure 9.2. Multilevel Structural Equation Model (Manifest Model). Teacher Social Capital factors are correlated, as are Teaching Practice factors. State math tests used as the outcome.

Many investigations of the relationships addressed in this dissertation have been conducted using models with observed composite indicators that ignore measurement error (see, e.g., Ferguson & Danielson, 2014; Kane & Staiger, 2012; Kraft & Papay, 2014; Ladd, 2009; Loeb, Darling-Hammond, & Luczak, 2005). This chapter contrasts such a multi-stage approach with a structural equation model which estimates the measurement model simultaneously with the structural predictive model and thereby incorporates estimates of measurement error into the model's effect estimates. The comparisons in this chapter provide an empirical test of the extent to which inferences about the structural relationships among these constructs are dependent on modeling assumptions regarding the measurement error in indicators. By extension, the results here may provide limited insight into the extent to which previous research that ignores measurement error might be subject to similar biases and errors of inference.

I present random effects and model fit indices from the Latent model and the Manifest model below in Table 9.1. Model fit indices suggest adequate fit for these models, with no clear advantage for one model or the other. For the latent model, the estimates for RMSEA (0.02) and SRMR Within (0.04) and Between (0.07) suggest acceptable to good fit by conventional cut-offs, though the CFI and TLI estimates are about 0.93, somewhat below the conventional 0.95 threshold indicating good fit. The fit of the manifest model also appears acceptable, somewhat better by way of CFI and TLI estimates (0.97 and 0.95), and the SRMR Within (0.02), and acceptable in terms of the estimates for RMSEA (0.02) and SRMR Between (0.09). Residual variances of the factors are large across models, with slightly larger values in the manifest model, suggesting that large portions of the variance remain unexplained by these models.

Table 9.1

Variances, Residual Variances, and Model Fit: Latent vs. Manifest Models

	MSEM Latent				MSEM Manifest			
	Est	SE	Est/SE	p-val	Est	SE	Est/SE	p-val
Variances								
CONTROL	0.61	0.01	45.21	0.00	0.61	0.01	59.01	0.00
INSTRUCT	0.50	0.02	31.81	0.00	0.56	0.01	51.80	0.00
Residual Variances								
TRUST	0.94	0.02	38.92	0.00	0.96	0.02	62.51	0.00
CARES	0.96	0.02	54.62	0.00	0.96	0.02	62.06	0.00
PARENTS	0.95	0.03	38.24	0.00	0.96	0.02	55.08	0.00
EXPERT	0.96	0.02	57.57	0.00	0.96	0.02	57.16	0.00
CONTROL	0.91	0.03	32.90	0.00	0.91	0.03	32.58	0.00
INSTRUCT	0.98	0.02	56.20	0.00	0.99	0.01	77.36	0.00
MATH10 Within	0.37	0.01	42.14	0.00	0.37	0.01	42.15	0.00
MATH10 Between	0.54	0..04	13.54	0.00	0.54	0.04	13.66	0.00

Latent Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.07
Manifest Model Fit: RMSEA: 0.023; CFI: 0.974; TLI: 0.952; SRMR Within: 0.024; SRMR Between: 0.085
Variances restricted as invariant across levels. Residual variances are standardized values.

In Table 9.2, the factors show stronger inter-correlations in the latent model (0.51 – 0.78) than in the manifest model (0.44 – 0.73). The estimated correlations between the teaching practice constructs differ by only 0.02 at the Between level, but by 0.09 at the Within level. Among the teacher social capital constructs at the Between level, differences range from 0.04 to 0.07.

Table 9.2.

Factor Correlations in Latent vs. Manifest Models; Between Level (Within Level)

	Latent Model				**Manifest Model**			
	CONTROL	INSTRUCT			CONTROL	INSTRUCT		
CONTROL	1				1			
INSTRUCT	0.68 (0.54)	1			0.66 (0.45)	1		
	TRUST	CARES	PARENT	EXPERT	TRUST	CARES	PARENT	EXPERT
TRUST	1				1			
CARES	0.78	1			0.73	1		
PARENT	0.62	0.53	1		0.56	0.48	1	
EXPERT	0.68	0.74	0.51	1	0.64	0.69	0.44	1

The results from the model with manifest indicators are similar in most respects to the latent model results, but there are important differences (see Table 9.3). Both models result in similar inferences being drawn for Control, namely a significant association with student math outcomes of moderate magnitude (0.37 in Latent model; 0.39 in Manifest model). Also, Instruction has no significant relationship with student math outcomes in either model. For Access to Expertise, both models show marginally positive associations with Control (0.16 in both models), and both indicate positive associations with Instruction, though in the latent model the association is significant (0.23) while in the manifest model the association between Expertise and student math outcomes is marginally significant (0.16; p = 0.06). Both models also show a negative association between Expertise and student outcomes (-0.20 in Latent; -0.17 in Manifest). No significant effects were found in either model for Trust, Cares, or Parent Networks on teaching practices, and no indirect effects from teacher social capital factors through teaching practice factors were significant. However, the latent model estimates a significant association between Parent Networks and student math outcomes (0.17) while this association is not significant in the manifest model (0.11; p = 0.11); and the manifest model shows a marginally significant association between Trust and student math outcomes (0.12; p = 0.09) that is not found in the latent model. Covariates' effects on outcome variable were all significant at <0.01 level except Male, grade level, and School Percent Black and Hispanic. District effects are not available for release due to the MET data use agreement.

Table 9.3
Structural Effects in MSEM Latent vs. Manifest Models; (Latent Model: N = 15,644 students, 520 Teachers; Manifest Model: N = 15,628 students; 520 Teachers)

	MSEM Latent	MSEM Manifest

	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL Between ON										
TRUST	-0.11	0.11	-0.98	0.33		-0.06	0.09	-0.72	0.48	
CARES	-0.02	0.12	-0.14	0.89		-0.04	0.11	-0.35	0.73	
PARENT	0.10	0.08	1.25	0.21		0.10	0.06	1.58	0.11	
EXPERT	0.16	0.10	1.69	0.09	*	0.16	0.08	1.87	0.06	*
INSTRUCT Between ON										
TRUST	-0.05	0.12	-0.38	0.70		-0.08	0.10	-0.78	0.44	
CARES	-0.10	0.12	-0.81	0.42		-0.05	0.11	-0.45	0.66	
PARENT	-0.04	0.09	-0.48	0.64		0.00	0.07	0.03	0.98	
EXPERT	0.23	0.10	2.37	0.02	**	0.16	0.09	1.88	0.06	*
MATH10 ON										
CONTROL Between	0.37	0.07	5.52	0.00	***	0.39	0.06	6.29	0.00	***
INSTRUCT Between	0.02	0.07	0.27	0.79		0.00	0.07	0.02	0.99	
TRUST	0.04	0.09	0.47	0.64		0.12	0.07	1.68	0.09	*
CARES	0.15	0.11	1.33	0.18		0.10	0.09	1.12	0.26	
PARENT	0.17	0.08	2.06	0.04	**	0.11	0.07	1.59	0.11	
EXPERT	-0.20	0.09	-2.15	0.03	**	-0.17	0.08	-2.19	0.03	**
School BlackHisp	-0.07	0.06	-1.18	0.24		-0.08	0.06	-1.37	0.17	
CRMATH09	0.39	0.06	6.83	0.00	***	0.38	0.06	6.78	0.00	***
FACTORS Regressed on Covariates										
…ON School Black & Hispanic %										
TRUST	-0.24	0.05	-4.88	0.00	***	-0.19	0.04	-4.85	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.20	0.04	-4.92	0.00	***
PARENT	-0.22	0.06	-3.95	0.00	***	-0.20	0.04	-4.63	00.0	***
EXPERT	-0.20	0.04	-4.67	0.00	***	-0.20	0.04	-4.97	0.00	***
…ON Classroom prior math achievement										
CONTROL Between	0.24	0.05	5.40	0.00	***	0.24	0.04	5.48	0.00	***
INSTRUCT Between	-0.03	0.05	-0.54	0.59		-0.04	0.05	-0.73	0.47	
MATH10 ON Covariates										
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	107.92	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.55	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.73	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.05	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.31	0.19	
GRAD E6	0.01	0.02	0.61	0.54		0.01	0.02	0.81	0.42	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	-0.04	0.97	
School % BlHisp	-0.07	0.06	-1.18	0.24		-0.08	0.06	-1.37	0.17	
Class Prior Ach	0.39	0.06	6.83	0.00	***	0.38	0.06	6.78	0.00	***

* = p < 0.10;   ** = p < 0.05;   *** = p < 0.01

The manifest model has consistently smaller standard errors than the latent model, with the reductions ranging from 0.01 to 0.02, about seven to twenty-two percent. The structural coefficients are, on the whole, similar across the models, though estimates for individual parameters shift in the manifest model from the estimates in the latent model by as much as 0.08 standard deviations. The changes from one model to the other were small but resulted in meaningfully different inferences in regard to two of the six latent factors included in the model, Trust and Parent Networks. The latent model estimated a positive effect of Parent Networks on student math outcomes and no significant effects of Trust, whereas the manifest model found a marginally significant effect of Trust on student math outcomes while finding that the association between Parent Networks and student outcomes was not significantly different from zero (p-value: 0.11).

Discussion

This chapter set out to examine how inferences regarding the relationships among teacher social capital, teaching practice, and student achievement change depending on whether measurement error is taken into account or ignored. The two models compared were multilevel multidimensional structural equation mediation models, with the first model employing latent factors and the second one employing manifest composite indicators in the place of those factors.

The first noticeable difference is that the standard errors are slightly smaller in the manifest model. This follows from the fact that this model ignores the measurement error in the indicators and instead utilizes aggregate composites stripped of any measurement error. The differences in standard errors range from 0.01 – 0.02, depending on the dimension, and do not appear to contribute meaningfully to differences in substantive inferences regarding the structural relationships. In the largest difference, the estimation of the effect of TRUST on INSTRUCTION, the manifest indicator has a standard error that is twenty-two percent smaller than that of the respective latent factor. One might imagine scenarios in which differences of this degree could result in different substantive inferences. For instance, the estimated effect of Parent Networks on Control is approximately 0.10 standard deviations in both models. The standard error in the latent model is approximately 0.08 and in the manifest model approximately 0.06. This is not a large enough shift to conclude that this is a significant effect estimate, but the p-value shifts from 0.21 to 0.11; a change of that degree could, under slightly different circumstances, result in a change in interpretation of the substantive meaning of the parameter. It is also worth recalling that both of these models rely on survey items that I selected on the basis of the strength of their measurement characteristics: survey items with weak factor loadings and large residual correlations were dropped (see Chapter VI and Appendix C); therefore it may be that the comparison here underestimates the differences that might be found between other latent and manifest models in cases in which the indicator variables are not so rigorously screened based on the quality of their measurement.

The other difference in the estimates produced by these models lies in the parameter estimates. At first glance, the parameter estimates appear similar from one model to the next – the largest effects are for Control on student math achievement and for Expertise on Instruction and Control, along with the large negative effect of Expertise on student math outcomes. However, the estimate for Parent Networks on student math outcomes is substantially smaller in the manifest model, and the estimate for the effect of Trust on student outcomes is correspondingly larger. Closer inspection of the parameters suggests that the estimates for the effects of these different factors shift to a degree that is decidedly nontrivial. The four teacher social capital factors are relatively strongly correlated, as are the two teaching practice factors, and while the latent model attributes the teacher social capital effect on student math achievement to the influence of Parent Networks, it appears that the manifest model attributes a larger portion of this effect to Trust. Smaller changes in effect estimates among factors can be seen in examining other parameter estimates, but the shift in the effect estimate of Parent on student math achievement and effect estimate of Trust on student achievement, together with the smaller standard errors in the manifest model, are large enough to result in different substantive inferences.

The manifest model estimates larger (more positive) effects for Control and Trust on student outcomes, and a less negative effect for Expert in comparison to the estimates from the latent model for these parameters. But the manifest model estimates smaller effects for Instruction, Cares, and Parent. Here it appears pertinent to recall the differences seen in the reliability of the measurement of these constructs. First, residual variances suggest that, across both models, a larger portion of the variance of Instruction is not captured than of the variance of Control. In addition, table 6.5 reported higher ICCs for Control items and the Control factor than for the Instruction items and factor. In addition, the items measuring Control show generally stronger factor loadings than those for the Instruction construct, particularly at the Between level. Higher ICCs for the Control factor and items mean that a greater amount of the variance in ratings of Control is variance Between classrooms rather than among students within a classroom. For students' ratings of Instruction, however a relatively larger proportion of the variance lies among students within the same classroom. Stronger factor loadings for the Control items on the Control factor signify less measurement error as compared to the Instruction items and factor, and this is particularly true in measuring Between classrooms. All together, these results suggest that students report more reliably on teachers' levels of Control than on the quality of their Instruction.

The greater reliability for Control than Instruction might be related to the different estimates seen in the latent and manifest models. In the manifest model, when these constructs are formed by averaging the item responses to form manifest construct indicators, the Control construct has even stronger effect estimates than in the latent model, and the Instruction construct correspondingly weaker effect estimates. In the latent model, when these constructs are estimated as latent factors with distinct factor loadings for each item included as part of the model estimation, the estimated effect of Control is still much larger than that of Instruction, but the estimate for Control is slightly smaller than the estimate seen in the manifest model and the estimate for Instruction is slightly larger than in the manifest model.

The changes in effect estimates for the teacher social capital factors on student math achievement, however, appear not to fit with the above speculation. Among the teacher social capital factors, Cares has the strongest factor loadings, followed by Parent and Trust, which have factor loadings similar in magnitude, and Expert has the lowest factor loadings (see Table 6.5). However, the manifest model estimates a more positive effect for Trust on student outcomes and a less negative effect for Expert, while showing smaller effects for Cares and Parent. In other words, for teacher social capital factors, the two factors with relatively weaker measurement show stronger effect estimates in the manifest model than in the latent model. But for the teaching practice factors, the opposite appears true: the factor with the stronger measurement, Control, shows a stronger effect estimate in the manifest model than in the latent model.

Examining the effects of the teacher social capital factors on the teaching practice factors across the models, the manifest estimates for Trust are less negative on Control, but more negative on Instruction; for Cares, the manifest model estimates are more negative on Control and less negative on Instruction; for Parent, the manifest model estimate on Control is essentially unchanged from the latent model, while the estimate on Instruction moves from slightly negative to just over zero; for Expertise, the manifest model estimate on Control is approximately unchanged, while the estimated effect on Instruction is somewhat smaller.

As shown in Table 9.4, looking across all these changes, there is no clear pattern: no particular factors have consistently larger or smaller estimates in the manifest model versus the latent model. Furthermore, no clear connection emerges between the measurement model estimates and comparison of estimates across one model that accounts for measurement error and another model that ignores measurement error. In other words, it is not as if the factors with less measurement error have consistently stronger estimates within the latent model. Nor is it the case that the factors with less measurement error have consistently weaker estimates within the latent model.

Table 9.4

Changes in Effect Estimates and Factor Correlations from Latent to Manifest Model & Strength of Measurement; Between (Within)

		Moving from Latent to Manifest Model, How do the Estimates Change?		
Factor	Relative Strength of Measurement	Δ in Estimated Effect on Control	Δ in Estimated Effect on Instruction	Δ in Estimated Effect on Math
Trust	Medium	+0.05	-0.03	+0.08
Cares	Strong	-0.02	+0.05	-0.05
Parent	Medium	0.00	+0.04	-0.06
Expert	Weak	0.00	-0.07	+0.03
Control	Strong	NA	-0.02 (-0.09)	+0.02
Instruction	Weak	-0.02 (-0.09)	NA	-0.02

Estimated effects of Control on Instruction and Instruction on Control are factor correlations, so no causal direction is implied by the model. They are shown at the Between and (Within) levels.

Chapter X.

Measuring Teaching Practice Using a Student Survey vs. Observation Ratings

This chapter addresses research question 4c: How do the relationships among the constructs differ when teaching practice is measured by expert observation ratings rather than students' survey responses? The multilevel multidimensional structural equation mediation model discussed in detail in Chapter VII is compared to one using the Framework For Teaching (FFT) observation ratings as indicators of teaching practice in place of the Tripod student perception survey items. The model also changes somewhat to accommodate the different data structures, with student responses nested within teachers in the case of the use of the Tripod and expert raters nested within teachers in the case of the use of the FFT.

The comparisons in this chapter are of methodological interest as an empirical example of the extent to which inferences differ when different instruments are used to indicate mediating factors. These comparisons are also expected to lead to hypotheses regarding the strengths and limitations of each measure, the potentially unique aspects of teaching that each measure best reveals, and the processes that underlie the relationships among the constructs. In addition, these comparisons are of policy interest because of the widespread use of both the Tripod and the FFT as measures for teacher evaluation and development.

Both models (see Figures 10.1 and 10.2) include four correlated teacher social capital factors at the Between teacher level: TRUST: extent to which teachers trust one another and feel trusted; CARES: extent to which teachers perceive the administration as responsive to their needs; PARENT Networks: quality of the communication between the school and parents; and

Access to Expertise (EXPERT): teachers' access to support personnel and quality professional development. All factors are hypothesized to impact students' state math test score outcomes, MATH10, at Level 2 and the models include the same covariates as described previously. Further detail is provided in Chapter V and the Mplus 7.4 syntax is provided in Appendix D.

The teaching practice factors are conceptually as similar as possible, but are not, of course, defined identically or measured in the same manner across the Tripod and FFT. As measured by the Tripod, the teaching practice factors are measured through *students'* responses to survey items and the factors are correlated at the student level (Within classrooms) and teacher level (Between classrooms/teachers). The factors are named: CONTROL: a measure of the respect demonstrated by students within the classroom; and INSTRUCTION: a broad measure of students' perceptions of the quality of the teacher's explanations, feedback, provision of cognitive challenge, respect for student thinking, and emotional support. When the Framework For Teaching is used, expert ratings on each domain are averaged to form Between level manifest domain ratings akin to survey items at the teacher (Between) level. Two correlated teaching practice factors are formed from these domain ratings: CLASSROOM MANAGEMENT: formed from ratings on Creating an Environment of Respect and Rapport, Managing Classroom Procedures, and Managing Student Behavior; and INSTRUCTIONAL SUPPORT: formed from ratings on Establishing a Culture for Learning, Using Questioning and Discussion Techniques, Communicating with Students, Engaging Students in Learning, and Using Assessment in Instruction.

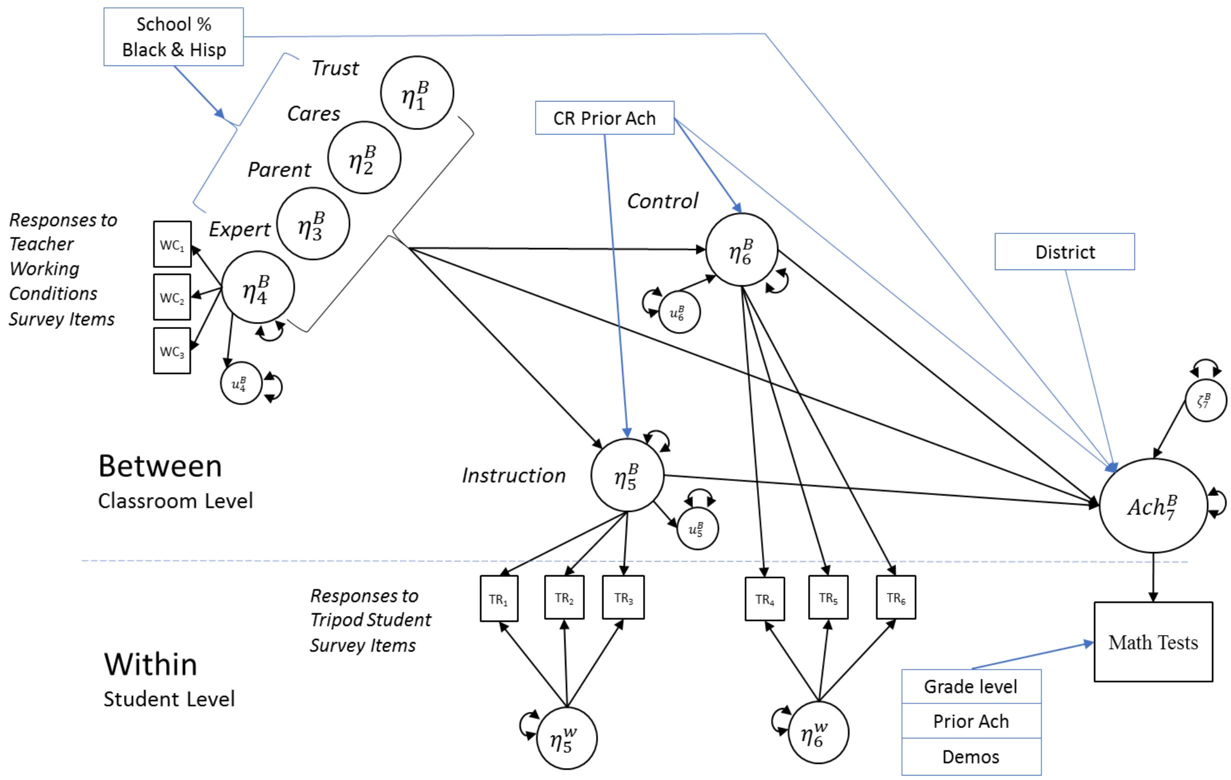

Figure 10.1. Multilevel Structural Equation Model using Tripod.

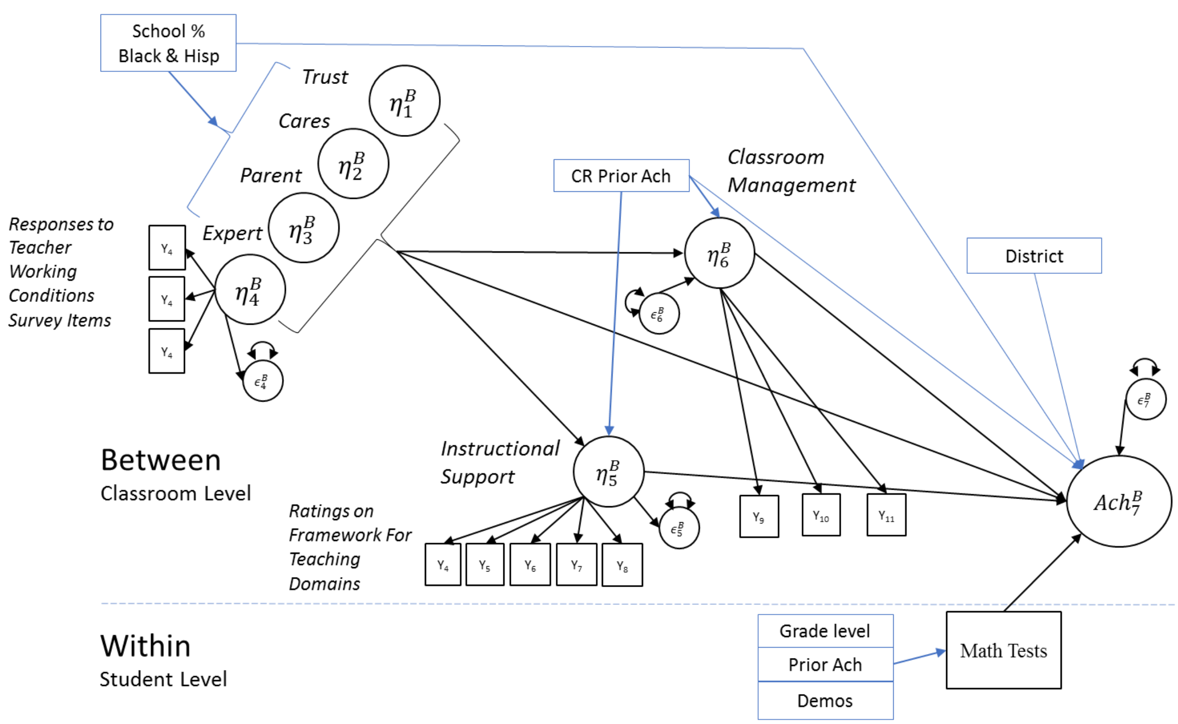

Figure 10.2. Multilevel Structural Equation Model using Framework For Teaching.

Analysis of the Measures

Before comparing the results across these models, it is important to keep in mind that any differences found may stem from at least three sources: the reliability of the measures; second, the structure of the data; and third, the aspects or constructs of classroom practice captured by each instrument.

Reliability

Previous research using the MET dataset indicated stronger reliability for the indicators of instruction derived from the Tripod, compared to indicators from the FFT (Kane & Staiger, 2012; Mihaly et al, 2013). In detail, in middle school math classrooms, Mihaly and colleagues report a year-to-year stability of 0.90 in middle school math classrooms for a composite formed from the Tripod items and year-to-year stability of 0.67 for a FFT composite in the same classrooms. However, they find very low variability for the FFT composite from one section to another in comparison to the Tripod composite, suggesting that the finding of greater reliability for the Tripod stems from aggregating the ratings over a greater number of student raters and/or from these student raters experiencing a much greater number of lessons than the expert raters who record the FFT scores. Tripod developers report highly reliable—in the range of 0.80 and above, and stable ratings for the Tripod Seven Cs (Ferguson, 2010, p. 4) and Mihaly et al (2013) also report Cronbach's alpha measures of reliability in the range of 0.80 for each of the Seven C's. However, as detailed in Chapter VI, analyses using multilevel item response theory and multilevel confirmatory factor analysis find evidence that the Tripod can reliably distinguish two, rather than seven, factors of teaching. For its part, the FFT was the most reliable observation measure in the MET report. For one observation of one lesson, the variance in the overall rating attributable to the teacher was 0.37, and after aggregating scores across four occasions with four different observers, 0.67 of the variance reflected consistent differences among teachers, which the authors refer to as an “implied reliability” (Kane & Staiger, 2012), and which matches the year-to-year stability reported by Mihaly and colleagues. The analyses in this dissertation do not directly estimate the reliability of the Tripod factors versus the FFT factors, but measurement estimates discussed in chapter VI suggest greater reliability for the Control factor than the Instruction factor when using the Tripod. Similarly for the FFT factors employed here, evidence based on the factor loadings and residual variances suggests that the Classroom Management factor has greater reliability than the Instructional Support factor. The differences in the reliabilities of the factors within each measure serve as a reminder that the previous MET findings of greater reliability for the Tripod were based on a single manifest composite formed from all items/components and therefore do not translate directly to the conceptualizations employed here with two latent factors formed from nineteen Tripod items / eight FFT components. Comparatively, evidence points in favor of the reliability of the Tripod’s Control factor and against the reliability of the FFT’s Instructional Support factor, but beyond that, the conclusion to be drawn is less clear.

Data Structure

Regarding data structure, the Tripod student survey items are measured at the student level from students nested within classrooms and the multilevel structural equation model estimates separate Within classroom and Between classroom effect estimates. The FFT ratings, however, are measured by expert raters that are aggregated to the teacher/classroom Between level for each domain. The domain aggregates are then treated as “items” at Level Two, providing for an estimate of the measurement error, but ignoring the nesting of raters within teachers by aggregating across raters. This aggregation is necessary because the two separate nesting structures of the students within teachers (Tripod) and raters within teachers (FFT) cannot both be accounted for in the same MSEM model. The inability of the model to account for the nesting of the FFT may contribute to the larger standard errors seen in the FFT estimates.

Aspects of Classroom Practice Captured by Each Instrument

The third potential source of different inferences is of most interest here: The Tripod and the FFT, while often using similar terminology and frames of reference, may in fact be measuring different aspects of teaching practice (or measuring different aspects in different depth or detail). To examine this possibility, I consider two perspectives. From the first perspective, the two instruments are measuring the same thing and so the relationships among the indicators derived from them can be viewed as providing confirmatory evidence of the validity of each measure. This view is represented by Ferguson and Danielson (2014), who provide a broad conceptual mapping of how the factors of the Tripod relate to the elements of the FFT. Their analysis finds,

> Generally, both the FfT and Tripod 7Cs frameworks have components pertaining primarily to communication about rules for time use, procedures, effort, and personal conduct. In addition, both have components pertaining primarily to communication and aspects of instruction associated very directly with implementing the curriculum (p. 107).

They further state that that educators "can cross-walk the two frameworks…" and they conclude:

> the conceptual overlap between the frameworks is substantial and…empirical patterns in the data show similarities in adult and student assessments at the classroom level….Generally, the 7Cs Press components are related to the FfT Classroom Environment components, and the 7Cs Support components are related to the FfT Instruction components (p. 128).

This perspective also is emphasized in MET project research that evaluates the strength of one measure exclusively by the extent to which it is "demonstrably related" to another measure, namely to value-added estimates of student gains (Kane & Cantrell, 2010, p. 5). This perspective focuses on the extent to which the multiple measures of teaching (including the Tripod and FFT as well as other instruments) correlate significantly with one another.

The opposing perspective emphasizes the ways in which different instruments measure different aspects of teaching. This perspective acknowledges areas of overlap in that the different instruments are still measuring the same general construct, but focuses on the unique contributions that each instrument can make to our understanding of the larger construct. This view is represented by researchers and policymakers who emphasize the value of multiple measures and remind us that “No measure is perfect” (Kane & Staiger, 2012, p. 60). This perspective also finds its place in MET project research when researchers emphasize how student survey information provides a “less sophisticated understanding of effective instruction than a trained observer,” yet at the same time offers a measure that is “less susceptible to lesson to lesson variation” (Kane & Staiger, 2012, p. 14), when they argue that, in spite of lower reliability, classroom observation holds the “potential to identify strengths and address specific weaknesses in teachers’ practice” (p. 14), and when they make the case for the use of multiple measures on the basis of their differential ability to predict different student outcomes such as happiness in class and effort (Kane & Staiger, 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013), as well as in conclusions such as “there is a common component of effective teaching shared by all indicators, but there are also substantial differences in the stable component across modes” (Mihaly, McCaffrey, Staiger, & Lockwood, 2013, p. 23).

To some extent, these two perspectives are simply different points of view and each one has validity. However, choosing to emphasize one perspective or the other entails important consequences. For instance, previous MET research found a stronger association for the Tripod than the FFT with a value-added criterion (Kane & Staiger, 2012). Viewed from the first perspective that says each instrument is measuring the same general construct of teaching, this finding suggests greater criterion validity for the Tripod. Viewed from the second perspective, which emphasizes how each instrument measures distinct aspects of teaching, this finding suggests that the Tripod and the value-added criterion measure similar aspects of teaching while the FFT may measure unique aspects not captured by either the Tripod or the value-added estimates.

There is also a degree to which this is a question that is answerable with reference to a conceptual analysis of the two instruments considering the evidence that these two instruments are measuring the same general construct(s) versus the evidence that they are measuring unique aspects of this construct. While Ferguson and Danielson (2014) provide a general conceptual mapping that stresses the similarities among the two instruments, I contend that this point of view overlooks evidence of at least five important distinctions.

Differences across the Tripod and FFT can be seen by comparing Tables 10.1 and 10.2. First, the two instruments focus on different aspects of teaching. The Tripod only weakly addresses (in items B154 and B147) the Questioning and Discussion Techniques that serve as a critical element of the FFT. And in the FFT, engaging students is measured broadly through elements including instructional materials and pacing of lessons and indicators such as *Learning tasks that require high-level student thinking and invite students to explain their thinking*, as opposed to the Captivate items in the Tripod, such as B29 and B44, which assess student engagement directly as an outcome formed from the sum of the extent to which each student finds the lessons “interesting” and “enjoyable”.

Second, the two instruments approach the measurement of teaching from distinct angles, even when they overlap substantially on the component they are measuring. For instance, in measuring Classroom Management / Control, the Tripod asks students to report on student behavioral outcomes such as B112, and students' interactions with the teacher such as B49, while the FFT expects raters to report on teacher actions such as how s/he "monitors student behavior," or "responds to misbehavior" and the extent to which the teacher's interactions with students are respectful and caring.

Third, the two instruments are based on different conceptual structures. In the Tripod, Control and Caring are distinct factors of teaching, and empirical data suggests that items within these two factors correlate relatively weakly in comparison to other items. In the FFT, on the other hand, elements of the Classroom Environment domain and of component 2a Creating an Environment of Respect and Rapport, ask raters to judge the extent to which *Teachers convey that they are invested in and care about their students*, *Teachers show respect for students' backgrounds and lives*, *convey warmth and caring*, and show *Encouragement and Active listening*.

Fourth, the two instruments come from different foundations, with the FFT arising from Constructivist theory and therefore emphasizing the discussion techniques mentioned above as well as grouping strategies (see 2c) that reflect close connections to sociocultural theory. The Tripod, on the other hand, shows the influence of process-product research and behaviorist theory in items such as B6 and in the behavioral focus on student outcomes.

Fifth, each FFT domain is itself a high inference synthesis of several components and many elements each one of which might be viewed as, and could in fact be re-written as a single, separate item. Consider two elements of component 2c. Managing Classroom Procedures: *Teachers help students to develop the skills to work purposefully and cooperatively in groups, with little supervision from the teacher;* and *Experienced teachers have all necessary materials* *to hand, and have taught students to implement routines for distribution and collection of materials with a minimum of disruption to the flow of instruction.* If these two elements were redesigned as survey items, an observer might easily respond quite differently to these two items when rating the same teacher during the same lesson, yet for the FFT ratings, the observer is expected to assign a rubric score from 1 -4 for the overall element, and this single number stands in as the holistic combination of all the elements.

Underlying all of these distinctions, and perhaps superseding them all, is the most obvious difference in the sources of the information included in these ratings: the FFT rests on averaging the judgments of several expert observers, while the Tripod relies on aggregating many more student ratings. As detailed in chapter II, this difference in the source of information entails numerous trade-offs, most notably trading off the greater expertise and outside knowledge of observers for many additional student raters and hundreds more occasions on which students can develop their inside knowledge.

The differences among these two instruments may contribute to and help to explain the differences found when estimating the relationships among the constructs using one instrument or the other.

Table 10.1
Tripod Student Survey Items, Grouped by Factor

Var.	Item Wording
	CONTROL
B112	Student behavior in this class is under control.
B138	Student behavior in this class is a problem.*
B46	My classmates behave the way my teacher wants them to.
B49	Students in this class treat the teacher with respect.
B6	Our class stays busy and doesn’t waste time.
	INSTRUCTION
A10	My teacher in this class makes me feel that he/she really cares about me.
B34	My teacher really tries to understand how students feel about things.
B1	If you don’t understand something, my teacher explains it another way.
B17	My teacher has several good ways to explain each topic that we cover in this class.
B80	My teacher explains difficult things clearly.

B70	In this class, we learn a lot almost every day.
B90	In this class, we learn to correct our mistakes.
B29	My teacher makes learning enjoyable.
B44	My teacher makes lessons interesting.
B154	My teacher gives us time to explain our ideas.
A54	My teacher respects my ideas and suggestions.
B147	My teacher checks to make sure we understand what s/he is teaching us.
B58	We get helpful comments to let us know what we did wrong on assignments.
B83	The comments that I get on my work in this class help me understand how to improve.

Table 10.2

FFT Elements and Indicators within Each Component

CLASSROOM MANAGEMENT	
2a CREATING AN ENVIRONMENT OF RESPECT AND RAPPORT	
Elements	**Indicators**
Teacher interactions with students, including both words and actions Student interactions with other students, including both words and actions	Respectful talk, active listening, and turn-taking Acknowledgement of students' backgrounds and lives outside the classroom Body language indicative of warmth and caring shown by teacher and students Physical proximity Politeness and encouragement Fairness
2c MANAGING CLASSROOM PROCEDURES	
Management of instructional groups Management of transitions Management of materials, supplies Performance of classroom routines	Smooth functioning of all routines Little or no loss of instructional time Students playing an important role in carrying out the routines Students knowing what to do, where to move
2d MANAGING STUDENT BEHAVIOR	
Expectations Monitoring of student behavior Response to student misbehavior	Clear standards of conduct, possibly posted, and possibly referred to during a lesson Absence of acrimony between teacher and students concerning behavior Teacher awareness of student conduct Preventive action when needed by the teacher Absence of misbehavior Reinforcement of positive behavior
INSTRUCTIONAL SUPPORT	
2b ESTABLISHING A CULTURE FOR LEARNING	
Importance of the content and of learning Expectations for learning and achievement Student pride in work	Belief in the value of what is being learned High expectations, supported through both verbal and nonverbal behaviors, for both learning and participation Expectation of high-quality work on the part of students Expectation and recognition of effort and persistence on the part of students High expectations for expression and work products
3a COMMUNICATING WITH STUDENTS	
Expectations for learning Directions for activities Explanations of content Use of oral and written language	Clarity of lesson purpose Clear directions and procedures specific to the lesson activities Absence of content errors and clear explanations of concepts and strategies Correct and imaginative use of language
3b USING QUESTIONING AND DISCUSSION TECHNIQUES	

Quality of questions/prompts Discussion techniques Student participation	Questions of high cognitive challenge, formulated by both students and teacher Questions with multiple correct answers or multiple approaches, even when there is a single correct response Effective use of student responses and ideas Discussion, with the teacher stepping out of the central, mediating role Focus on the reasoning exhibited by students in discussion, both in give-and-take with the teacher and with their classmates High levels of student participation in discussion
3c ENGAGING STUDENTS IN LEARNING	
Activities and assignments Grouping of students Instructional materials and resources Structure and pacing	Student enthusiasm, interest, thinking, problem solving, etc. Learning tasks that require high-level student thinking and invite students to explain their thinking Students highly motivated to work on all tasks and persistent even when the tasks are challenging Students actively "working", rather than watching while their teacher "works" Suitable pacing of the lesson: neither dragged out nor rushed, with time for closure and student reflection
3d USING ASSESSMENT IN INSTRUCTION	
Assessment criteria Monitoring of student learning Feedback to students Student self-assessment and monitoring of progress	The teacher paying close attention to evidence of student understanding The teacher posing specifically created questions to elicit evidence of student understanding The teacher circulating to monitor student learning and to offer feedback Students assessing their own work against established criteria

Results

I present results from the model employing the Tripod items and the model employing the FFT ratings below in Tables 10.3 – 10.5. For the Tripod model, the CFI and TLI estimates are about 0.93, somewhat below the conventional 0.95 threshold indicating good fit, while the estimates for RMSEA (0.02) and SRMR Within (0.04) and Between (0.07) suggest acceptable to good fit by conventional cut-offs. The fit of the FFT model appears good, better than the Tripod model by way of RMSEA (0.007), CFI and TLI estimates (0.98 and 0.98), and the SRMR Within (0.02), and approximately the same in the estimate for SRMR Between (0.07). Residual variances suggest large portions of variance in these factors remains unaccounted for in either model. Teacher social capital factors show generally smaller residuals in the Tripod model, except for Parent Networks, while the teaching practice factors have somewhat smaller residuals in the FFT model.

Table 10.3

Variances, Residual Variances, and Model Fit: MSEM using Tripod vs. FFT

	MSEM Tripod				MSEM FFT			
	Est	SE	Est/SE	p-val	Est	SE	Est/SE	p-val
Variances								
Control/CR Management	0.61	0.01	45.21	0.00	NA	NA	NA	NA
Instruct/Instruct Support	0.50	0.02	31.81	0.00	NA	NA	NA	NA
Residual Variances **(Between except where indicated)**								
TRUST	0.94	0.02	38.92	0.00	0.98	0.02	65.13	0.00
CARES	0.96	0.02	54.62	0.00	0.98	0.01	73.77	0.00
PARENTS	0.95	0.03	38.24	0.00	0.85	0.04	23.71	0.00
EXPERT	0.96	0.02	57.57	0.00	0.99	0.01	109.60	0.00
Control/Class Management	0.91	0.03	32.90	0.00	0.87	0.05	19.44	0.00
Instruct/Instruct Support	0.98	0.02	56.20	0.00	0.90	0.04	20.86	0.00
MATH10 Within	0.37	0.01	42.14	0.00	0.37	0.01	42.13	0.00
MATH10 Between	0.54	0..04	13.54	0.00	0.61	0.05	12.90	0.00

Tripod Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.07
FFT Model Fit: RMSEA: 0.007; CFI: 0.980; TLI: 0.977; SRMR Within: 0.000; SRMR Between: 0.067
Variances restricted as invariant across levels. Residual variances are standardized values.

As shown in Table 10.4, the model using Tripod items implies strong inter-correlations among the four teacher social capital factors (ranging from 0.51 to 0.78) and the two teaching practice factors (0.54 at Within level; 0.68 at Between level). The model using FFT ratings estimates ever so slightly stronger correlations among the teacher social capital factors (ranging from 0.54 to 0.78) and markedly stronger correlation among the teaching practice factors (0.83 at Between level; no Within level is estimated for teaching practice factors in this model).

Table 10.4

Factor Correlations in Tripod vs. FFT Models; Between Level (Within Level)

	Tripod Items		**FFT Ratings**	
	CONTROL	INSTRUCT	Classroom Management	Instructional Support
Measurement Model (MCFA/CFA)				
CONTROL/CR Manage	1		1	
INSTRUCT/In. Support	0.65 (0.54)	1	0.85	1
Structural Model				
CONTROL/CR Manage	1		1	

INSTRUCT/In. Support		0.68 (0.54)	1		0.83	1		
	TRUST	CARES	PARENT	EXPERT	TRUST	CARES	PARENT	EXPERT
TRUST	1				1			
CARES	0.78	1			0.78	1		
PARENT	0.62	0.53	1		0.66	0.55	1	
EXPERT	0.68	0.74	0.51	1	0.67	0.74	0.54	1

Several of the main inferences are similar across measures. The results using both Tripod items and FFT ratings show that Control/Classroom Management has a significant effect on student math outcomes (0.37 for Tripod; 0.35 for FFT) while the estimated effect for Instruction/Instructional Support on student math outcomes is not significant. Among teacher social capital factors, Parent Networks shows a positive direct effect on student math outcomes in both models (0.17 for Tripod; 0.18 for FFT), though in the model using FFT, the effect is only marginally significant (p-value: 0.06). This appears due to larger standard errors in the FFT model because the magnitude of the estimated effect is slightly larger than the estimate in the model using Tripod. Also, regardless of the measure used, no indirect effects from teacher social capital factors through teaching practice factors are significant.

Across both models, the estimated effects of the school percent Black and Hispanic on teacher social capital factors are negative (ranging from -0.20 to -0.24 for Tripod; from -0.09 to -0.38 for FFT). All of these effects are statistically significant at $p < 0.01$, with the exception of the effect on Access to Expertise in the model using FFT. That estimate (-0.09) is close to negative one-tenth of a standard deviation in magnitude, and just outside the 0.1 cut-off for a marginally significant effect estimate. Taken together, the evidence suggests that schools with higher percentages of Black and Hispanic students tend to have lower ratings on teacher social capital factors. Also in both models, the estimated strength of the teacher's Control / Classroom Management is stronger in classrooms with higher average prior achievement (0.24 for Tripod; 0.20 for FFT). All covariates' effects on MATH10 outcome variable are significant at <.01 level except Male, grade level, and School Percent Black and Hispanic. District fixed effects are not available for release due to the MET project data use agreement.

Table 10.5

Effect Estimates for MSEM Models using Tripod vs FFT to Measure Teaching Practice (Between). Tripod: N = 15,644 students, 520 Teachers; FFT: N = 15,614 students; 520 Teachers

	Tripod Items					FFT Ratings				
	Est	SE	Est/SE	p-val		Est	SE	Est/SE	p-val	
CONTROL / Classroom Management ON										
TRUST	-0.11	0.11	-0.98	0.33		-0.11	0.12	-0.93	0.35	
CARES	-0.02	0.12	-0.14	0.89		0.25	0.13	1.96	0.05	*
PARENT	0.10	0.08	1.25	0.21		0.18	0.09	1.98	0.05	**
EXPERT	0.16	0.10	1.69	0.09	*	-0.04	0.11	-0.41	0.68	
INSTRUCTION / Instructional Support ON										
TRUST	-0.05	0.12	-0.38	0.70		0.04	0.12	0.34	0.74	
CARES	-0.10	0.12	-0.81	0.42		0.17	0.13	1.26	0.21	
PARENT	-0.04	0.09	-0.48	0.64		0.09	0.10	0.91	0.37	
EXPERT	0.23	0.10	2.37	0.02	**	-0.03	0.11	-0.29	0.77	
MATH10 Outcome ON										
Control/CM	0.37	0.07	5.52	0.00	***	0.35	0.13	2.61	0.01	***
Instruct/InSupp	0.02	0.07	0.27	0.79		-0.13	0.13	-0.97	0.33	
TRUST	0.04	0.09	0.47	0.64		0.06	0.10	0.58	0.57	
CARES	0.15	0.11	1.33	0.18		0.06	0.11	0.54	0.59	
PARENT	0.17	0.08	2.06	0.04	**	0.18	0.10	1.88	0.06	*
EXPERT	-0.20	0.09	-2.15	0.03	**	-0.13	0.10	-1.35	0.18	
FACTORS regressed on covariates										
…ON School Black & Hispanic %										
TRUST	-0.24	0.05	-4.88	0.00	***	-0.15	0.05	-2.94	0.00	***
CARES	-0.20	0.05	-4.44	0.00	***	-0.13	0.05	-2.71	0.01	***
PARENT	-0.22	0.06	-3.95	0.00	***	-0.38	0.05	-8.18	0.00	***
EXPERT	-0.20	0.04	-4.67	0.00	***	-0.09	0.05	-1.64	0.10	
…ON Classroom Prior Achievement										
CONTROL/CM	0.24	0.05	5.40	0.00	***	0.20	0.06	3.34	0.00	***
INSTRUCT/InSupp	-0.03	0.05	-0.54	0.59		0.19	0.06	3.01	0.00	***
MATH10 ON Covariates										
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	108.06	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.56	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.62	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.06	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.31	0.19	
GRADE6	0.01	0.02	0.61	0.54		0.00	0.02	0.08	0.94	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	-0.26	0.79	
School % BlHisp	-0.07	0.06	-1.18	0.24		0.00	0.06	0.06	0.96	
Class Prior Ach	0.39	0.06	6.83	0.00	***	0.45	0.05	8.46	0.00	***

*** = significant at p < 0 .01; ** = significant at p < 0.05; * = marginally significant at p < 0.1

However, results also differ in several instances. In the model using Tripod, Access to Expertise has a marginally significant effect on Control (0.16), a significant effect on Instruction (0.23) and a significant negative direct effect on student state math test scores (-0.20); but, in the model employing FFT ratings, Access to Expertise shows no significant effects on teaching practice factors or on student achievement outcomes. Using Tripod, no significant effects are found for Trust, Cares, or Parent on teaching practice factors, yet using the FFT ratings, Parent Networks has a significant positive effect on Classroom Management (0.18) and Cares has a marginally significant positive effect on Classroom Management (0.25). Ratings of teachers' Instructional Support on the FFT are also higher in classrooms with higher average prior achievement (0.19), but this association does not hold in the students' (Tripod) ratings of the quality of Instruction.

Discussion

This chapter set out to examine how inferences regarding the relationships among teacher social capital, teaching practice, and student achievement change when teaching practice is measured by different instruments, specifically the Tripod student perception survey items or expert ratings on the Framework For Teaching observation rubric. Differences in estimates across these models may come from at least three distinct sources: differences in reliability across the measures, differences in the data structure across the models, or differences in the aspects of teaching being measured by the instruments. Previous research suggests somewhat greater reliability for the Tripod ratings, and, within each measurement model, for the Control and Classroom Management factors respectively. Regarding data structures, in the Tripod model, the student survey item responses were nested within teachers, but in the FFT model, the experts' ratings on each domain were averaged to the teacher level and latent factors were then formed from those Level 2 indicators treated as manifest.

Conceptual analyses of the instruments suggest that the two instruments differ in important ways related to the aspects of teaching being measured. First, they differ in their source of information, with the Tripod relying on many student responses and the FFT resting on comparatively fewer expert raters' assessments. The two instruments also approach the measurement of teaching from distinct angles, with the Tripod asking student raters to report on *student behavioral outcomes* while the FFT focuses ratings on *teacher actions*. This distinction arises most clearly in regard to the FFT's emphasis on questioning and discussion techniques and its broader measurement of teachers' *efforts* to engage students in contrast to the Tripod's focus on *students' reported levels of engagement*. The two instruments are also based on different conceptual structures, with the Tripod separating emotional support from classroom management while the FFT categorizes these aspects together. And underneath their structures, they arise from different foundations, with the FFT having roots in Constructivist and sociocultural theories and the Tripod showing the influence of Behaviorist theories and process-product research. Finally, they differ because the FFT components are high inference syntheses of many elements and indicators, whereas each Tripod item is akin to a single indicator.

Several inferences stand out from the quantitative analyses: First, teachers' level of classroom control /classroom management is positively associated with students' math outcomes in both models. Whether using the Tripod or the FFT, teachers' skill (or good fortune) in managing student behavior and keeping students on task is significantly associated with positive effects on student math achievement. These effects appear moderate in magnitude, approaching one-third of a standard deviation, and are the largest effects found in either model. A one standard deviation difference in the teacher's Control predicts one-third of a standard deviation difference in her/his students' math scores. These estimates also echo the findings from earlier MET studies drawing from the same database and reinforce the findings of the broader literature on the effects of teacher quality on student outcomes (see e.g., Brophy, 1984; Raudenbush & Jean, 2014).

In both models, teacher's Control / Classroom Management is stronger in classrooms with higher average prior achievement: there is a positive effect estimate for classroom prior math achievement on Control / Classroom Management. This suggests support for the commonplace observation that teachers tend to do a better job with classroom management and control when they are given a "good class" – a classroom with, on average, higher achieving students (if we assume that higher achieving students are also likely to be, on average, better behaved). However, this association does not preclude the possibility that teachers perceived as "better" at classroom management are more likely to be assigned classrooms full of "better" students. Importantly, the model controls for the effects of classroom average prior achievement on both Control and Student Achievement, so the estimated positive effect of Control on Student Achievement is found even after accounting for the effects of "good classes" on teachers' Control and the effects of these "good classes" on students' end-of-year achievement.

Ratings of teachers' quality of Instructional Support on the FFT are also higher in classrooms with higher average prior achievement, but no association was found in the students' (Tripod) ratings of the quality of Instruction. The presence of this association when using expert ratings combined with the absence of this association when using student ratings suggests three likely interpretations (and, of course, all of these interpretations may be true to some degree): First, if there is an association – if classrooms with higher prior achievement do, on average, receive stronger quality teaching, then student ratings may be biased in some way that prevents that association from showing up – perhaps higher achieving students tend to rate their teachers more strictly than do lower achieving students. Second, if classroom average prior achievement is not in truth associated with instructional quality, then expert raters may be biased in some way. Perhaps they rate teachers higher in classrooms which have more high achieving students. For example, classrooms with more high achieving students might tend to have more or higher quality student participation that leads observers to rate the teachers higher in instructional quality.

A third possibility, and perhaps the most interesting, is that these two instruments may be picking up on different aspects of instructional quality and perhaps teaching tends to be better in "good" classes mainly in respect to those aspects of teaching that are measured by the FFT but not by the Tripod. Considering the different aspects of teaching that are measured by the FFT versus the Tripod leads to interesting hypotheses. Recall that the FFT emphasizes the measurement of teacher actions, and includes a full component focusing on Questioning and Discussion techniques. Further, the FFT provides a broad assessment of both student engagement levels and teacher actions related to student engagement. The Tripod, on the other hand, centers on measuring student perceptions and student outcomes. This dichotomy brings to mind the distinction discussed in chapter II between *good* teaching and *successful* teaching (Fenstermacher & Richardson, 2005). Perhaps the Tripod focuses on assessing successful teaching while the FFT places more emphasis on measuring the extent of good teaching. Recall that teachers' Instructional Support on the FFT is higher in classrooms with higher average prior achievement while no association is found in Tripod ratings of Instruction – the implication, then, may be that higher achieving classrooms are experiencing more *good* teaching (as rated by the FFT), but experiencing no measurable difference in the *success* of that teaching (as rated by the Tripod). This distinction between the Tripod and FFT is not clear-cut, of course, since the two instruments also overlap in notable ways. However, drawing this distinction leads to an important recognition that the goodness of teaching practice may not directly translate into success and that student engagement and motivation are partly determined by factors prior to and outside of the classroom practice of any specific teacher.

The results here also suggest another association between some aspects of teacher social capital and teaching practices, though the precise nature of that association differs depending on the measure used. In the Tripod model, the extent to which teachers have Access to Expertise is positively associated with greater degrees of classroom Control and higher quality Instruction. In the FFT model, the strength of teachers' Networks with Parents and, with less certainty, the extent to which the administration Cares about teacher concerns are associated with stronger Classroom Management. Taken together, the findings suggest that school communication with Parents and administrative Caring about teachers' needs, in particular in terms of providing teachers with Access to Expert help, are associated with stronger teaching practices. The positive association between Expertise and Instruction in the Tripod results lends itself to the interpretation that high-quality professional development and Access to Expert support is associated with higher quality instructional practices. However, these findings should be interpreted cautiously because of the possibility discussed in chapter VII that these estimates may suffer from multicollinearity/suppression effects. Also, the associations are seen consistently across measures for effects on Control/Classroom Management but not on Instruction. Another way to interpret the positive effects of teacher social capital factors on Control/Classroom Management is that schools that support teachers and parents are more likely to feature classrooms that exhibit high levels of student behavior. Perhaps well-run schools on average demonstrate higher ratings on Classroom Management/Control and on some combination of Parent Networks, Care, or Expertise; some other factor associated with well-run schools might be driving all those high ratings.

Chapter XI. Conclusion

The quality of teachers has become a central focus of policy and research. Yet the focus on teachers has suffered from two critical limitations that this dissertation helps to address: first, an overly narrow focus on student test scores and so-called teacher effects to the exclusion of factors that impact teaching practice in the classroom, and second, a lack of attention to the conceptual, measurement, and modeling choices that are likely to impact the inferences drawn about the relationships among these factors.

The focus on teachers is clear in federal, state and district policies, including No Child Left Behind's mandate of a "highly-qualified teacher" in every classroom (U.S. Department of Education, 2002) and Race to the Top's emphasis on teacher evaluation (U.S. Department of Education, 2014), the promotion of alternative routes to teacher certification (e.g., Teach For America), revision of certification requirements, reform of licensure exams, incentives for advanced degrees and longevity, merit pay, and career ladder, apprenticeship, and residency models. Likewise, teachers have been a primary focus of education research, with attention focusing on findings that teacher quality varies a great deal (Hanushek & Rivkin, 2006; Sanders, Wright, & Horn, 1997), and has large impacts on student test outcomes (Nye, Konstantopolous & Hedges, 2004) and long-term outcomes (Hanushek 2011; Chetty, Friedman, & Rockoff, 2011; 2013).

Despite an abundance of research on the effects of teach*ers*, clear definitions have often been lacking, especially in regard to distinguishing between teach*er* quality as distinct from teach*ing* quality and between *good* teaching and *successful* teaching. In addition, little research has focused on the context and other factors that may impact the quality of teach*ing* practice. A critical aspect of the context that may impact teaching quality is the social capital available to teachers - the relationships teachers have with one another and with other key adults and the ways in which those relationships do or do not function to benefit teaching and learning. Research suggests that improved social capital will lead to improved teacher learning (Frank, Zhao, and Borman, 2004; Kraft & Papay, 2014), teacher motivation (Kraft, Papay, Charner-Laird, Johnson, & Reinhorn, 2012), and teacher retention (Ladd, 2009; 2011), all of which should contribute to improved instructional quality. This dissertation assembles evidence for the benefits of teacher social from literatures on teacher working conditions, school climate, distributive leadership, and trust, in addition to teacher social capital. By drawing connections among these literatures that have rarely if ever been connected in prior research literature, the conceptual model employed here links closely related constructs and hopefully contributes to a fuller understanding of teacher social capital and the evidence for its impacts.

This dissertation broadens the scope of the investigation into how teachers might impact student learning by hypothesizing a mediation model in which teaching practice both impacts student achievement directly and mediates the effects of the social capital available to teachers in the school environment.

Second, research on teachers has been conducted from a variety of approaches and using widely different instruments and models, yet not enough attention has been paid to the consequences of particular choices of constructs, measures, and modeling tools and assumptions. Lack of attention to construct definition reflects first, the too-often silo-ed nature of educational research agendas, which tend inexorably toward greater specialization and away from broader connections across related areas of study; researchers working on teacher working conditions and others analyzing social networks among teachers rarely if ever cite one another's work. Second, construct definition suffers from a tendency to employ terms that are inconsistently and at times inappropriately defined, such as teacher quality in place of teaching practice (Bell et al, 2012; Hiebert, 2013). The lack of applied research attention to measurement and modeling assumptions exists in spite of recent advances in multilevel and latent variable modeling that hold the promise of accounting for interdependencies in nested data (Raudenbush & Bryk, 2002) and measurement error in indicator variables (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011) and nearly one hundred years of prominent criticism of the reliability and validity of the available measures of classroom instruction (Martinez, 2012; Thomas, 1929).

This dissertation seeks to contribute to our knowledge of how methodological choices influence the estimated relationships among teacher social capital, teaching practices, and student achievement. To investigate the impact of different modeling approaches, it examined the relationships across single- vs. multilevel models, and latent trait vs. manifest indicator models. To assess the impact of measurement choices, it examined the relationships across measures derived from student survey vs. observation rating, and additionally conducted conceptual and empirical analyses of the instruments employed.

Discussion of Results

This dissertation utilizes a sample of nearly 16,000 middle school students and 520 math teachers in five districts from the Measures of Effective Teaching (MET) longitudinal database (Kane & Staiger, 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013) to investigate the relationships among teacher social capital, teaching practices, and student achievement. In the process, it examines two key areas that have received scant attention in the literature on teacher effects: first, the interplay among the many contextual factors that impact teaching and learning, and second, the consequences of conceptual, measurement, and modeling assumptions and choices in shaping our understanding of teaching and learning and the relationships among them.

I first describe the measurement findings, then results that appear consistently across models and measures, next address the different inferences stemming from different choices of models and measures, and after that consider thoughts regarding future research considerations. I discuss the limitations in these analyses and offer concluding thoughts.

Measurement Findings

For the Working Conditions Survey, I found evidence that a set of eighteen items could provide a measure approximating four correlated factors of teacher social capital: TRUST: extent to which teachers trust one another and feel trusted; CARES: extent to which teachers perceive the administration as responsive to their needs; PARENT Networks: quality of the communication between the school and parents; and Access to Expertise (EXPERT): teachers' access to support personnel and quality professional development. Multilevel and single-level confirmatory factor analyses of the four-factor model suggested adequate fit to the data (MCFA fit indices: RMSEA: 0.008; CFI: 0.948; TLI: 0.941; SRMR Within: 0.047; SRMR Between: 0.058). Structural equation models suggested criterion validity for the model in that associations with teaching practices and student math outcomes in several cases aligned with associations predicted by prior research and theory: strong Parent Networks were associated with greater student learning (standardized effect estimate: 0.17), and Parent Networks, Caring administration, and Access to Expertise showed limited evidence of positive associations with teaching practices. However, findings also suggested evidence of multicollinearity among the four factors, particularly with regard to a negative association between Access to Expertise and student math outcomes (standardized effect estimate: -0.20). These structural findings contrary to theory suggest that additional measurement work may be called for in order to develop instruments that can consistently distinguish among multiple factors of teacher social capital. In comparing across factors, factor loadings and residual variances suggested relatively stronger measurement for the CARES and PARENT dimensions.

For the Tripod, evidence supported a two-factor structure that accorded with results from Kuhfeld (under review), and unpublished analyses in the MET User's Guide (Bill and Melinda Gates Foundation, 2012), as well as broadly similar to a bifactor structure identified by Wallace, Kelcey, & Ruzek (2016). These results are not consistent with the developers' "Seven Cs" conceptualization or the PRESS + SUPPORT model from Ferguson & Danielson (2014). A reduced set of nineteen items was used to adequately approximate this two-factor model (MCFA fit indices: RMSEA: 0.037; CFI:0.935; TLI: 0.930; SRMR Within: 0.040; SRMR Between: 0.072). The two factors identified were CONTROL: a measure of the respect demonstrated by students within the classroom; and INSTRUCTION: a broad measure of students' perceptions of the quality of the teacher's explanations, feedback, provision of cognitive challenge, respect for student thinking, and emotional support.

Factor loadings, ICCs, and residual variances suggested comparatively more reliable measurement of the CONTROL dimension. Taken together, the measurement results suggest the following scenario: two students in Mrs. Jones' class are more likely to disagree on whether she provides meaningful feedback that helps them improve than they are to disagree on the extent to which students behave respectfully. They are more likely to disagree about whether Mrs. Jones cares about them as people than they are about whether their classmates follow Mrs. Jones' directions. If we understand the disagreement among students as error, then these results show that students report more reliably on teachers' levels of Control than on the quality of their Instruction. Another way of saying this is that students are better able to consistently distinguish between levels of classroom behavior and respect than between levels of quality in teachers' explanations, feedback, challenge, and support. However, it is also plausible that some portion of that student disagreement stems from some students actually receiving more meaningful feedback or more care from their teacher than other students receive (Schweig, 2014).

For the FFT, I found support for a two-factor structure that broadly mirrored the structure for teaching practice identified using the Tripod. This model identified CLASSROOM MANAGEMENT, composed of three components: Creating and Environment of Respect and Rapport, Managing Classroom Procedures, and Managing Student Behavior; and INSTRUCTIONAL SUPPORT, composed of five components: Establishing a Culture for Learning, Communicating with Students, Using Questioning and Discussion Techniques, Engaging Students in Learning, and Using Assessment in Instruction. This structure is similar to the developers' model, but moves Establishing a Culture for Learning from the CLASSROOM MANAGEMENT factor to the INSTRUCTIONAL SUPPORT factor based on the factor loadings, modification indices, residual variances, overall model fit, and conceptual analysis suggesting that this component assessed the teacher's level of expectations for student learning rather than Classroom Management elements. This model appeared to display adequate fit based on CFA fit indices: RMSEA: 0.097; CFI: 0.966; TLI: 0.950; and SRMR: 0.032. However, the two factors correlate 0.849, suggesting that a unidimensional model may be appropriate, and the RMSEA is above the 0.08 cut-off for acceptable fit. The two-factor model was retained to allow for close comparison with the Tripod model. Factor loadings were strong across all components and, together with residual variances, suggested somewhat stronger measurement of Classroom Management than Instructional Support.

After examining and refining the individual measures, I compared how the structural relationships among these constructs varied when examined across a variety of models and employing the two different measures of teaching practice: the Tripod student survey and the FFT expert observation rating. Considering these structural relationships, several findings stood out as consistent across models and measures, while other inferences shifted depending on the model or the measures employed.

Consistent Findings Across Models and Measures

There is consistent evidence across models that schools with higher percentages of Black and Hispanic students have lower ratings on teacher social capital factors, with the effect estimates averaging about one-fifth of a standard deviation. Teachers in schools with higher percentages of disadvantaged minorities tend to rate their school as less trusting and their administration as putting less effort into meeting teachers' needs. These teachers have less confidence in the quality of the school's communication and involvement efforts with parents. These teachers also rate their schools as doing a worse job in providing quality professional development and support / Expertise to teachers, though this association is weaker in some models. Schools serving high proportions of Black and Hispanic students also typically serve large percentages of poor students and achieve at lower levels on standardized tests (Fiel, 2013). In addition, these segregated schools make do with lower levels of funding, less experienced and less qualified teachers, higher levels of teacher turnover, less challenging curricula, and often inadequate facilities and learning materials (Orfield, Kuscera, & Siegel-Hawley, 2012; Phillips & Chin, 2004; Roza, Hill, Sclafani, & Speakman, 2004). In the results in this dissertation, the negative impacts associated with ethnic/racial school segregation appear to impact teachers' social capital, but do not show direct effects on student outcomes. Further research is needed to uncover the extent to which teacher social capital may mediate the impacts of school segregation on student outcomes.

The strongest and most consistent associations among the factors suggest that teachers' Control / Classroom Management is positively associated with students' math outcomes. The association between Control and student achievement is found after controlling for the effects of classroom average prior achievement on teachers' levels of control and on students' prior end-of-year achievement and it is moderate in magnitude (about one-third of a standard deviation). Effects of Control on student achievement echo the findings from earlier MET studies drawing from the same database and reinforce the findings of the broader literature on the effects of teacher quality on student outcomes. Teachers' classroom management/control skills have shown a strong relationship with positive student outcomes (see e.g., Brophy, 1984; Raudenbush & Jean, 2014). In a related vein, schools which emphasize academic success, conformity to specific standards of achievement, and a "press" for intellectual attainment have higher student outcomes (Lee, Smith, Perry, & Smylie, 1999; Phillips, 1997).

Looking in detail at the items that form the Control factor suggests that good Control is in some respects equivalent to and indistinguishable from good student behavior. The extent to which students behave respectfully and stay on task is a property of classrooms not teachers and represents complex interactions between students and teacher (Bell, Gitomer, McCaffrey, Hamre, Pianta, and Qi, 2012). Factors other than the teacher, particularly parents, school and community culture, and school leaders, also likely impact students' behavior. Therefore, the association between Control and student achievement might also be interpreted to reflect a positive association between student behavior and student academic outcomes, as opposed to or in addition to an association between teacher skill in management and student academic outcomes. Saying that good student control is associated with higher student achievement, and that good student composition is associated with good student control as well as with higher student achievement does not resolve the question of the extent to which student composition vs teaching practice (or parent influence) is causing these effects. Having said this, the fact that these associations are found consistently whether using student or expert observer ratings does lend some support to the importance of the teacher's role.

With regard to the relationship between teacher social capital and student achievement, the strength of a school's Parental Networks is directly and positively associated with student math outcomes across most models. Stronger student math outcomes are predicted in schools where teachers say their school provides clear communication to parents and encourages parental involvement, and rate parents as well-informed about school happenings. This relationship is about 0.17 standard deviations with a p-value of about 0.04 in the latent MSEM Tripod model and 0.18 standard deviations (p-value: 0.06) in the FFT model. These findings reinforce the findings from parental involvement literature that suggest that 1) the efforts of teachers and other school personnel can influence the level and quality of parental involvement (Bermúdez & Márquez, 1996; Epstein, 1988), and 2) the level of parental involvement is positively associated with improved student outcomes (Fan & Chen, 2001; Sui-Chi & Wilms, 1996).

Inconsistent Findings across Models and Measures

Multilevel vs. Single-level. In comparing the multilevel structural equation model (MSEM) results to the single-level structural equation model (SEM) results, substantively and meaningfully different inferences were found in regard to Instruction, Trust, Cares, and Parent

Networks. These differences stemmed first from systematically smaller standard errors in the SEM estimation, a finding in accord with the known tendency of single-level models to underestimate standard errors when data is nested hierarchically (Raudenbush & Bryk, 2002). Second, the differences across models come from a marked shift in how the models apportion the effect of teaching practice on student math achievement: the multilevel model ascribes nearly all of the effect to the Control factor, whereas the SEM apportions the effect about equally to both Control and Instruction. Indications from measurement results discussed above suggest that Control is measured more reliably at the classroom level than Instruction. The findings here show that the multilevel model finds stronger relationships involving the factor that is measured more reliably at the Between level (Control), whereas the single-level model, in which the Within and Between level effects are conflated (Muthén, 1994), finds approximately equal effects of Control and Instruction on student math outcomes.

Latent vs. Manifest Model. Comparing results across a latent and a manifest model revealed small differences across models. Smaller standard errors in the manifest model are expected because this model utilizes aggregate composites stripped of any measurement error. The differences in standard errors range from 0.01 – 0.02, depending on the dimension, and do not appear to contribute meaningfully to differences in substantive inferences. Findings do reflect consequential differences between the models' estimated effect of Trust on student math achievement (marginally significant and larger by 0.08 standard deviations in the manifest model) and the estimated effect of Parent Networks on math outcomes (smaller by 0.06 standard deviations and nonsignificant in the manifest model). Theory and prior evidence suggest that manifest models that ignore measurement error cause estimates of the relationships among the underlying constructs to be attenuated by the measurement error; the misspecification of the model can lead can lead to incorrect inferences about critical questions such as effect sizes (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Muthén & Asparahouv, 2007; Preacher, Zhang, & Zyphur, 2011). However, ignoring measurement error in the indicators is also common practice in a great deal of applied research and there have been few empirical examples demonstrating whether or not the theoretical advantages result in meaningful differences in actual data (Preacher, Zyphur, & Zhang, 2010). This dissertation provides one such empirical example. In this case, the differences among models were not large and appeared to stem not from underestimated standard errors, but from shifts in effect estimates that were not predictable based on the relative strength of the measurement of the factors. Specifically, for teacher social capital factors, the two factors with relatively weaker measurement show stronger effect estimates in the manifest model than in the latent model. But for the teaching practice factors, the opposite appears true, albeit to a tiny degree: the factor with the stronger measurement, Control, shows a stronger effect on math in the manifest model (by 0.02 standard deviations) than in the latent model. The different inferences across models in this case are not large and are not readily explained by differences in the strength of measurement; they might best be interpreted as statistical noise.

One further observation: both of these models rely on survey items selected on the basis of the strength of their measurement characteristics: for the most part survey items with weak factor loadings and large residual correlations were dropped (see Chapter VI and Appendix C); therefore, it may be that the comparison here underestimates the differences that might be found between other latent and manifest models in cases in which the indicator variables are not so rigorously screened based on the quality of their measurement. This observation also implies that if others find more substantial differences across latent versus manifest models, those findings might in part reflect statistical noise resulting from poorly performing indicator level variables.

Student Survey vs. Expert Observation Ratings. In comparing the results across the use of different instruments, I found three differences that shed light on their relative strengths and on the underlying relationships among the constructs. First, ratings of teachers' quality of Instructional Support on the FFT are higher in classrooms with higher average prior achievement (effect estimate 0.19), but no association was found in the students' (Tripod) ratings of the quality of Instruction. One possible cause for these different findings is that the instruments may measure different aspects of teaching. For instance, conceptual analysis of the instruments suggests that the FFT may more closely reflect the *goodness* of teaching whereas the Tripod comes closer to approximating the *success* of teaching. Further, the FFT provides a broad assessment of both student engagement levels and teacher actions related to student engagement, including high-level learning tasks, suitable pacing, and effective instructional materials and student grouping. The Tripod, on the other hand, centers on measuring student perceptions and student outcomes; in regard to student engagement, the Tripod measures this as an outcome – to what extent do the students find lessons interesting and enjoyable, while the FFT rates not only observers' perceptions of students' engagement but also the teacher actions that might be expected to result in these outcomes. Another possibility is that the ratings for one or both instruments may reflect systematic biases among the raters themselves. For example, higher achieving students may rate their teachers more strictly than low achieving students, or expert raters may rate teachers higher in classrooms which have more high achieving students because they observe higher quality student participation.

Second, the strength of Parent Networks is associated with stronger levels of classroom management in the FFT model (effect estimate 0.18) but not in the Tripod model. With the FFT measure, strong Parent Networks are marginally directly associated with positive student outcomes while also being associated with more positive levels of student behavior that are themselves strongly related to positive student academic outcomes. Because the association between Parent Networks and Classroom Management using FFT is significant while there is no significant association between Parent Networks and Control using Tripod, and noting that students' item responses on the Control factor appear to conflate respectful student behavior with effective classroom management, these results may suggest that the relationship in the FFT model should be interpreted as a positive relationship between parental communication and teachers' levels of classroom management, more so than a positive relationship between parental involvement and students' levels of behavior. However, this speculation rests on the assumption that FFT raters are able to accurately discern teachers' classroom management skills as existing independent of the student behavior they observe.

Third, administrative Caring and responsiveness to teachers' needs appears associated with stronger teaching practices, but the exact makeup of that association differs depending on the measure used. The associations are seen across both measures consistently for effects on Control/Classroom Management, but effects on Instruction are only found when using the Tripod. When using the FFT observer ratings, a marginally significant positive association is seen between Cares and Classroom Management (0.25), but the associations between Expert and Instruction and Expert and Control are not significant. When using the Tripod, significant associations are found between Access to Expertise and Control (0.16) and Instruction (0.23) but no association is seen between Cares and Control. Putting these findings together, the suggestion is that administrations which are Caring and responsive to teachers' needs, in particular in providing teachers with Access to Expert help, are associated with stronger teaching practices and/or more well-behaved students. Well-run and supportive schools are likely to include administrators who Care about and respond to teachers' needs *and* provide teachers with Access to Expert help (and the correlation between the Cares and Expert factors is 0.82). These findings, therefore suggest a positive relationship between well-run schools and classrooms that are well-managed / feature well-behaved and respectful students. It is of course quite possible that some other element associated with well-run schools might be driving the high ratings on these factors of Care/Access to Expertise and Control/Classroom Management.

Findings Inconsistent with Theory

Several of the results that appeared in the multidimensional models were inconsistent with the theoretical model and with the results that would be expected based on prior research and theory. These counterintuitive results, along with the strong inter-correlations found among the four teacher social capital factors and between the two teaching practice factors, suggested the possibility that the multidimensional models suffered from multicollinearity and/or suppression effects. To investigate this possibility, I conducted sensitivity analyses which accounted for the incongruities and showed that the relationships among the constructs were consistent with theory.

In the multidimensional models using student survey ratings to measure teaching practice, teachers' Access to Expertise appears positively associated with greater degrees of classroom control and higher quality Instruction. However, the model also shows of a negative effect of Access to Expertise on student math test scores. Read directly, the negative estimate might suggest the "reverse causation" described by Ladd (2009), in which the provision of additional supports to low performing schools may make it appear that the additional supports are "causing" low test scores. Ladd found that teachers' perceptions of professional development opportunities were predictive of negative achievement effects in reading and concluded that this likely illustrated a case of more professional development opportunities being provided to schools that are performing less well in reading than other schools with the same types of students. However, in sensitivity analyses intended to isolate the effects of each factor, the negative effect on student math outcomes disappears in all cases while the positive effects of Expertise on teaching practice are weaker and not consistent across models. Thus, the results here suggest a more cautious interpretation: in sensitivity analyses, Access to Expertise shows no significant direct association with student math outcomes, thus suggesting that the negative association found in the multidimensional model reflects multicollinearity effects rather than reverse causation. Meanwhile, the suggestions of positive impacts of Expertise on teaching practice are still present but weaker after considering the sensitivity analyses.

Sensitivity analyses also provide instructive differences in regard to students' perceptions of Instructional quality. Students' perceptions of Instructional quality are not related to student math outcomes in the multidimensional models, but *are* related to better student math outcomes in sensitivity analyses which do not include Control. These results may reflect the fact that Control and Instruction are quite strongly correlated (0.65 correlation in MCFA) and the relative accuracy of the measures suggest greater reliability for Control than Instruction (stronger factor loadings, higher ICCs, smaller residual variance). The findings from the sensitivity checks run counter to, and at the same time help to explain, earlier MET research from Ferguson & Danielson (2014). They found that, when holding constant Control and Challenge, the predicted result of increasing teachers' Support (measured by a composite of the 5 Cs other than Challenge and Control) is to decrease the students' expected test scores. Sensitivity analyses suggest that the null estimate found here in the multidimensional model and the negative estimates found by Ferguson and Danielson may be best explained as a result of suppression effects and measurement weaknesses rather than as meaningfully interpretable signals of a negative or non-existent relationship between instructional support/quality and student learning.

Finally, the multidimensional model shows no significant relationships involving Trust or Cares with teaching practices or student achievement. But the sensitivity analyses suggest that the norms of a school, as represented in the levels of Trust among teachers and the extent to which the administration Cares about teacher concerns, are positively associated with students' state math outcomes. Students perform better on state math tests in a school in which teachers trust one another and perceive the school leadership as acting in their interests, suggesting that students in Trusting and Caring atmospheres may feel more confident and/or relaxed and thus either learn more or simply perform better on high-stakes tests. Those results are not strong or robust across models, but they are in accord with prior research and theory.

In all, the sensitivity analyses show that the relationships among these constructs individually were largely as expected; some of the relationships were small and not significant, but estimates that appeared at first glance contrary to theory were subsequently explained by analyses that isolated the effects of each factor and accounted for threats of multicollinearity/suppression.

The sensitivity analyses not only clarify these initially puzzling results, they also provide a clear reminder that complex models, featuring multiple measures, mediation, nesting, and high factor correlations, can face important complications including multicollinearity, suppression, model misspecification, potentially underestimated standard errors, and the possibility of conflating effects from Between and Within levels. Thoughtful attention to measurement and modeling choices are called for to avoid finding and interpreting spurious statistical artifacts. Considering the importance of teachers and teaching in ongoing policy debates, it is all the more important in this context to focus our attention on robust effects that are likely to hold up. Our students do not have time for us to chase mistakes, so we must take our time and take care to avoid them.

Limitations

The analyses herein are limited in several important respects. First, with regard to the sample, the MET data are drawn from volunteer districts, schools, and teachers; therefore, the sample cannot be considered representative. However, MET descriptive analyses found that teachers included in the MET study were similar to teachers in the same districts who did not participate in terms of ethnicity, years of teaching experience, and a value-added measure of teacher's impacts on state assessment outcomes (Kane & Staiger, 2012, Table 2, page 17; Table 10, page 17; Bill & Melinda Gates Foundation, 2012). In addition, the sample analyzed in this dissertation is drawn from middle school mathematics classrooms; particularly because of the many unique aspects of this grade level and subject area, results may not generalize to other grade levels or subject areas. Also, the analyses are based on observed correlations in cross-sectional data; therefore the results strictly demonstrate associations (conditional correlations) among factors and any references to findings of causal "effects" remain hypothesized.

Second, there are limitations to the models employed here. For instance, the multilevel models discussed in this dissertation nonetheless do not account for the third level of nesting, that of teachers/classrooms within schools. Initial models accounted for that level of nesting (through the use of the Twolevel Complex command in Mplus 7.4), but that model proved inestimable due to the large number of parameters in relation to the number of level 3 units (the number of schools). In addition, difficulties with model convergence contributed to the decisions to eliminate survey items that had weaker fit conceptually or empirically. Each survey item that was dropped made the model simpler, resulting in fewer parameters and a model that was more likely to be estimable. The decisions to drop individual items are detailed in Appendix C, but a critical reader might judge the construct coverage to be weakened by the elimination of so many items.

This limitation also applies to the conceptual models employed. Numerous potentially interesting relationships are left unexplored by the analyses in this dissertation. For example, concerning the relationships of the constructs with covariates, in Chapter VII, neither the unidimensional nor the multidimensional model showed an association between classroom racial composition and teaching practice. In fact, for this covariate, the estimates were positive even though theory and prior research on tracking and the unequal distribution of high-quality teachers suggest a likely negative association. In other words, we might expect to find that classrooms with greater percentages of historically disadvantaged minorities would tend to experience somewhat worse teaching practice. This was not found. This non-finding might have suggested multicollinearity / suppression due to the model also including the school percent Black and Hispanic, or it might have suggested that students of different ethnicities may rate teachers in systematically different ways. For instance, perhaps Hispanic and /or Black students tend to defer to authority more in their ratings and this biases their ratings of teaching practices higher than students of other races. This possibility was not explored in this dissertation, but see Schenke, Nguyen, Watts, Sarama, and Clements (2017) for an analysis of how students of different races may perceive classroom environments differently. In these analyses, because the classroom percent Black and Hispanic variable also showed no significant associations with the MATH10 outcome and was not adding to the explanatory power of the model, it was dropped from all subsequent analyses.

Third, the measurement of the constructs is, of course, imperfect. For instance, the operationalization of teacher social capital lacks any measurement of a dimension representing the strength of teacher-teacher networks, the depth of interactions within networks, or the strength of distributed leadership, and the operationalization of teaching practice features only two dimensions whereas some studies have found evidence for as many as nine separable dimensions of teaching practice. Further, because many items were dropped, results from these models cannot be directly compared to results obtained using all Teacher Working Conditions items or all Tripod items. Future research is called for to confirm or modify these models in other samples.

The instruments themselves are limited. As detailed in chapters II and III, measuring teaching practice and teacher social capital is a challenging enterprise. The measures employed here are widely used surveys and observation frameworks, with prior work assessing and attempting to improve their validity and reliability. However, substantial noise still exists in these measures and complicates the estimation of these models considerably. In addition to the typically expected measurement error, survey response differences may reflect true differences in student or teacher experiences. For instance, if certain students actually receive more teacher support than others, then these differences, perhaps across subgroups such as sex, race, or prior achievement strata, would reflect true sources of variation that would also be of substantive interest. The present study does not investigate these issues, instead focusing on the classroom level estimates of these constructs.

It might also be wished that additional measures of these constructs were available. For instance, previous MET research compared a number of alternative observation rubrics, but alternative student surveys were not included in the MET data. One such alternative might be the Questionnaire on Teacher Interaction (QTI), which measures students' views of teacher behaviors along two axes, one of dominance and submission, and the other of cooperation and opposition (Kyriakides, 2005). Kyriakides' analyses of the QTI demonstrated reliability of about .90, found a factor structure that showed nine distinct elements of quality teaching, and showed that eight of the nine elements were associated with student achievement gains in both math and Greek language. The MET data also lack any measures of teachers' written interactions with students, such as can be collected through a portfolio (Martinez, Borko, Stecher, Luskin, & Kloser, 2012), of teachers' daily perceptions of content coverage, such as can be collected through teacher logs (Rowan, Jacob, & Correnti, 2009), or of teachers' underlying knowledge and skills, such as can be collected through innovative teacher tests asking teachers to respond to videos and vignettes of practice (Kersting et al, 2012). The prior research on these and other measures of teaching quality tells us that these types of indicators would correlate substantially with the outcome - student achievement, and predictor - teaching quality. This suggests that the final models may have unmeasured confounds.

Measuring teaching quality accurately is therefore clearly of vital importance to this study; and yet it also might be argued to be this study's Achilles' heel. Teaching is an incredibly complex endeavor, which researchers have sought for decades to measure. The measures employed in this dissertation are clearly less than ideal, but they are arguably among the best available and, perhaps most importantly, they are available.

In terms of student outcomes, the models in this dissertation rely on student standardized state math test scores as a manifest outcome variable. One of the strengths of the MET dataset is that it also includes an alternative assessment of student math achievement, the Balanced Assessment of Mathematics (BAM). The BAM is designed to measure students' mathematical problem-solving and critical thinking. Clearly, additional insight into student outcomes might be gained by comparing the extent to which results shift when using the BAM outcome and/or forming a latent student achievement outcome. In working on this dissertation, I conducted preliminary analyses of these comparisons, but set them aside due to the large percentage of missing data on the BAM outcome, the difficulties with model convergence, and the necessity of focusing this dissertation on a reasonable number of comparisons.

In addition, non-achievement outcomes may be of particular interest in terms of the relationships explored in this dissertation. If students experience higher levels of trust, caring and feelings of connection while in school, this may lead them to engage in greater communication, cooperation, and civility in the wider community (Sztompka, 1999). These non-achievement outcomes are not directly addressed within this dissertation, but this limitation implies no less importance to these outcomes and comes solely as a reflection of the fact that adequate measures of students' social outcomes were not identified.

Final Thoughts

This dissertation examined two key areas that have received scant attention in the literature on teacher effects: first, the interplay among the many contextual factors that impact teaching and learning, and second, the consequences of conceptual, measurement, and modeling assumptions and choices in shaping our understanding of teaching and learning and the relationships among them. In examining teaching, this dissertation clarified a definition of teaching quality that has in prior research been indistinguishable from a host of related terms; in examining teacher social capital, this dissertation drew connections among several research traditions closely related to teacher social capital that have previously been investigated separately.

The evidence consistently suggests that schools with higher percentages of disadvantaged minority students show lower levels of teacher social capital, and that both teachers' success with control / classroom management and the quality of teachers' networks with parents are positively associated with student math outcomes.

Examination of measures and models indicate that measurement quality, modeling assumptions, and measurement choice all contribute to meaningful differences in the conclusions likely to be drawn. Conceptual, technical, and empirical analyses of the measures indicate poor quality in a number of survey items and result in different measurement structures than have been previously used in published literature. Failing to account for multilevel structure suggests misleading inferences due to underestimated standard errors and effect estimates that appear conflated at the Between and Within level. Comparisons across latent and manifest models do not show large differences and the differences are not predictable based on measurement strength. Comparisons across the student survey and observation ratings show differences that together suggest that the survey measures the *success* of teaching to a greater extent while the observation measurement particularly reflects the *goodness* of teaching.

Results also indicate several parameter estimates that are contrary to theory. However, the model considered here features a nested structure, multiple mediators, and high factor correlations, suggesting the possibility of multicollinearity. Indeed, the results contrary to theory disappear when reconsidered through a series of simplified models conducted as robustness checks. The model complexity and the evidence of multicollinearity, especially in the context of a vital policy focus such as teacher quality, amplify the importance of thoughtful consideration of measurement quality and choices of measures and models. Careless application of methodology could easily lead to incorrect inferences and policy decisions based on spurious findings.

In addition to these important findings, this dissertation advances three critical lessons:

1) Relationships between teaching and learning are more complex than previously advertised. In particular, more complex than previously concluded by MET research, which was itself advertised as the largest, most comprehensive study of teaching yet conducted.

2) Think small. The results in this dissertation vary meaningfully depending on seemingly small choices in measures and models. Because all of the methods compared here have been relatively commonly used in related applied research, the “results may vary” conclusion is not just a trivial social scientist’s bumper sticker, but a critical reminder that the conclusions drawn from research should be considered tentative until verified under plausible alternative specifications.

3) Think big. The conclusions reached here are somewhat tenuous, at certain points conflicting, and potentially confusing. In spite of these limitations, the variety of frameworks and approaches employed here serve to advance a larger argument: as researchers we are, at all times, viewing a small part of a moving elephant from a particular angle. Enlarging our viewpoint, through complex mediation models,

consideration of multiple research literatures, and the use of alternative methodologies, holds the promise of enriching our understanding.

Appendix A.
Potential Social Capital Questions

Networks

Teacher-Teacher: Time to Collaborate

Please rate how strongly you agree or disagree with the following statements about the use of time in your school.

TWC_TML21COLLAB: Teachers have time available to collaborate with colleagues.

TWC_TML21NONINSTIME: The non-instructional time provided for teachers in my school is sufficient.

Activities during the school day (i.e., time for which you are under contract to be at the school)?

TWC_TMT46COLLABPLN: Collaborative planning time

Please rate how strongly you agree or disagree with statements about professional development in your school.

TWC_PDL21TIME: An appropriate amount of time is provided for professional development.

As a beginning teacher, I have received the following kinds of supports.

TWC_MNT10SUPCOMPLN: Common planning time with other teachers

Teacher – Parents/Community Networks

Please rate how strongly you agree or disagree with the following statements about community support and involvement in your school.

TWC_CSL21INFLUENCE: Parents/guardians are influential decision makers in this school.

TWC_CSL21COMMUNIC: This school maintains clear, two-way communication with parents/guardians and the community.

TWC_CSL21ENCINVOLVE: This school does a good job of encouraging parent/guardian involvement.

TWC_CSL21INFOLEARN: Teachers provide parents/guardians with useful information about student learning.

TWC_CSL21KNOW: Parents/guardians know what is going on in this school.

TWC_CSL21STUSUCCESS: Parents/guardians support teachers, contributing to their success with students.

TWC_CSL21COMMSUCCESS: Community members support teachers, contributing to their success with students.

TWC_CSL21SUPPORT: The community we serve is supportive of this school.

Norms

Trust

Overall Level of Trust in the School

Please rate how strongly you agree or disagree with the following statements about the staff in your school.

TWC_DPT21CULTURE: Differences in the culture and background of staff are valued.

Please rate how strongly you agree or disagree with the following statements about school leadership in your school.

TWC_EML21TRUSTRESP: There is an atmosphere of trust and mutual respect in this school.

TWC_LDL21RAISECONC: Teachers feel comfortable raising issues and concerns that are important to them.

Please rate how strongly you agree or disagree with the following statements about teacher leadership in your school.

TWC_EML21EXPERTS: Teachers are recognized as educational experts.

TWC_EML21TRUSTSOUND: Teachers are trusted to make sound professional decisions about instruction.

Trust btw Admin-Teacher

Please rate how strongly you agree or disagree with the following statements about school leadership in your school.

TWC_LDL21SHAREDVIS: The faculty and leadership have a shared vision.

TWC_LDL21TCHRSUPP: The school leadership consistently supports teachers.

Please rate how strongly you agree or disagree with the following statements.

TWC_MET21IMPINSTR: There is a detailed plan for improving instruction in our school.

Please rate how strongly you agree or disagree with the following statements about managing student conduct in your school

TWC_PTL21BEHAVSUPP: A process (i.e. team) exists for behavior support planning and problem solving.

The school leadership makes a sustained effort to address teacher concerns about:

TWC_LDL21EFFORTLD: Leadership issues

TWC_LDL21EFFORTTM: The use of time in my school

TWC_LDL21EFFORTPD: Professional development

TWC_LDL21EFFORTTL: Teacher leadership

TWC_LDL21EFFORTIP: Instructional practices and support

Please rate how strongly you agree or disagree with the following statements about teacher leadership in your school.

TWC_EML21DECMAKE: Teachers are relied upon to make decisions about educational issues.

TWC_EML21TCHLEADER: Teachers are encouraged to participate in school leadership roles.

TWC_EML21SCHINFLU: Teachers have an appropriate level of influence on decision making in this school.

Extent to which Teachers are trusted with Important Roles (Distributed Leadership Roles)

Please indicate the role teachers have at your school in each of the following areas.

TWC_EML49INSTMAT: Selecting instructional materials and resources

TWC_EML49TECHNIQ: Devising teaching techniques

TWC_EML49ASSESS: Setting grading and student assessment practices

TWC_EML49INSERVE: Determining the content of inservice professional development programs

TWC_EML49SIPLAN: School improvement planning

TWC_EML49NEWTCH: The selection of teachers new to this school

Please rate how strongly you agree or disagree with the following statements about policies and practices in District 21.

TWC_DPT21DISTCOMM: There is good communication throughout the district.

TWC_DPT21DISTSUPP: I feel supported by the district.

TWC_DPT21VALUETCH: District 21 values teachers.

I trust the other teachers in my school

Please rate how strongly you agree or disagree with the following statements about instructional practices and support in your school.

TWC_MPL21LEARN: Teachers in my school have what it takes to get the children to learn.

Please rate how strongly you agree or disagree with the following statements about teacher leadership in your school.

TWC_EML21PROCESS: The faculty has an effective process for making group decisions to solve problems.

TWC_EML21SOLVE: In this school we take steps to solve problems.

TWC_EML21EFFLEADER: Teachers are effective leaders in this school.

Please rate how strongly you agree or disagree with the following statements.

TWC_MET21HELPOTHERS: Teachers take responsibility for helping one another do well.

Expertise

Please rate how strongly you agree or disagree with the following statements about your school facilities and resources.

TWC_FRL21PROPERSON: Teachers have sufficient access to a broad range of professional support personnel.

Please rate how strongly you agree or disagree with statements about professional development in your school.

TWC_PDL21FOLLOWUP: In this school, follow up is provided from professional development.

TWC_PDL21SUFFRES: Sufficient resources are available for professional development in my school.

TWC_PDL21DEEPEFFECT: Professional development deepens teachers' content knowledge.

Please rate how strongly you agree or disagree with the following statements.

TWC_MET21STULEARN: It's easy for other teachers in this school to know what students learned in my class.

TWC_MET21CONTKNLDG: I have detailed knowledge of the content covered and instructional methods used by other teachers at this school.

Please rate how strongly you agree or disagree with the following statements about instructional practices and support in your school.

TWC_IPL21SUPPORTS: Provided supports (i.e. instructional coaching, professional learning communities, etc.) translate to improvements in instructional practices by teachers.

Please rate how strongly you agree or disagree with the following statements about school leadership in your school.

TWC_LDL21FDBKIMPR: Teachers receive feedback that can help them improve teaching.

Depth of Interactions

Please rate how strongly you agree or disagree with the following statements.

TWC_MET21PLANINSTR: I frequently plan and coordinate instruction with my students' other teachers.

TWC_MET21CONSISTENT: I collaborate with other teachers to achieve consistency on how we assess student work.

Please rate how strongly you agree or disagree with statements about professional development in your school.

TWC_PDL21COLLEAGUE: Professional development provides ongoing opportunities for teachers to work with colleagues to refine teaching practices.

Please rate how strongly you agree or disagree with the following statements about instructional practices and support in your school.

TWC_IPL21PLCINSTR: Teachers work in professional learning communities to develop and align instructional practices.

Appendix B.
Correlations among items within hypothesized Teacher Social Capital Dimensions

Norms – Trust & Caring: polychoric correlations

Teachers are recognized as educational experts														
Teachers trusted to make decisions on instruction	.86													
Effective process for decisions to solve problems	.70	.69												
In this school we take steps to solve problems	.73	.74	.89											
Teachers are effective leaders in this school	.72	.69	.80	.79										
The faculty and leadership have a shared vision	.69	.61	.71	.71	.69									
Atmosphere of trust and mutual respect in school	.55	.60	.70	.72	.69	.73								
Teachers comfortable raising issues/concerns	.57	.62	.71	.70	.72	.74	.83							
School leadership consistently supports teachers	.61	.61	.71	.66	.67	.74	.82	.78						
Leadership addresses teacher concerns re: leadership	.62	.59	.74	.72	.73	.70	.64	.66	.69					
the use of time in my school	.58	.54	.64	.59	.61	.62	.58	.56	.62	.83				
professional development	.45	.46	.62	.64	.65	.55	.57	.53	.60	.77	.73			
teacher leadership	.58	.59	.70	.67	.71	.62	.58	.61	.68	.91	.85	.89		
instructional practices and support	.63	.59	.69	.71	.73	.67	.61	.67	.70	.85	.82	.79	.88	
Teachers encouraged to participate in leadership	.64	.65	.68	.64	.70	.61	.52	.57	.59	.66	.49	.55	.70	.62

Norms – Trust & Caring: Pearson's correlations

Teachers are recognized as educational experts													
Teachers trusted to make decisions on instruction	0.65												
Effective process for decisions to solve problems	0.47	0.49											
In this school we take steps to solve problems	0.61	0.55	0.67										
Teachers are effective leaders in this school	0.48	0.56	0.56	0.54									
The faculty and leadership have a shared vision	0.48	0.58	0.49	0.59	0.48								
Atmosphere of trust and mutual respect in school	0.41	0.50	0.50	0.53	0.51	0.58							
Teachers comfortable raising issues/concerns	0.42	0.52	0.52	0.53	0.52	0.59	0.74						
School leadership consistently supports teachers	0.40	0.44	0.46	0.46	0.50	0.52	0.64	0.60					
Leadership addresses teacher concerns re: leadership	0.37	0.34	0.42	0.38	0.50	0.39	0.38	0.37	0.44				
the use of time in my school	0.30	0.31	0.37	0.29	0.37	0.35	0.32	0.34	0.37	0.66			
professional development	0.31	0.36	0.36	0.32	0.46	0.33	0.33	0.36	0.44	0.66	0.62		
teacher leadership	0.32	0.35	0.37	0.34	0.42	0.36	0.34	0.42	0.48	0.63	0.68	0.71	
instructional practices and support	0.31	0.29	0.40	0.36	0.41	0.34	0.34	0.32	0.38	0.61	0.61	0.76	0.65

Parent-Teacher networks: polychoric

Parents/guardians are influential decision makers in this school.							
School maintains clear, two-way comm. with parents/guardians and teachers	0.55						
This school does a good job of encouraging parent/guardian involvement	0.56	0.87					
Teachers provide parents/guardians with useful info about student learning	0.35	0.74	0.68				
Parents/guardians know what is going on in this school.	0.58	0.75	0.73	0.67			
Parents/guardians support teachers, contributing to success with students	0.56	0.47	0.51	0.33	0.60		
Community support teachers, contributing to success with students	0.59	0.62	0.60	0.41	0.60	0.76	
The community we serve is supportive of this school.	0.63	0.62	0.60	0.44	0.67	0.80	0.92

Parent-Teacher networks: Pearson's

Parents/guardians are influential decision makers in this school.							
School maintains clear, two-way comm. with parents/guardians and teachers	0.39						
This school does a good job of encouraging parent/guardian involvement	0.38	0.56					
Teachers provide parents/guardians with useful info about student learning	0.22	0.47	0.45				
Parents/guardians know what is going on in this school.	0.37	0.46	0.52	0.45			
Parents/guardians support teachers, contributing to success with students	0.39	0.31	0.31	0.25	0.44		
Community support teachers, contributing to success with students	0.36	0.40	0.37	0.26	0.41	0.48	
The community we serve is supportive of this school.	0.33	0.35	0.30	0.27	0.41	0.47	0.73

T – T Networks: time to collaborate: polychoric correlations

Teachers have time available to collaborate with colleagues.			
The non-instructional time provided for teachers in my school is sufficient.	0.57		
In average week, how much time to…Collaborative planning time	0.45	0.24	
An appropriate amount of time is provided for professional development	0.42	0.41	0.22

T – T Networks: time to collaborate: Pearson's correlations

Teachers have time available to collaborate with colleagues.			
The non-instructional time provided for teachers in my school is sufficient.	0.45		
In average week, how much time to…Collaborative planning time	0.35	0.17	
An appropriate amount of time is provided for professional development	0.33	0.30	0.16

Distributed Leadership Roles: *Extent to which Teachers are trusted with Important Roles:* Polychoric

Please indicate the role teachers have at your school in each of the following areas:

Selecting instructional materials			
Devising teaching techniques	0.65		
Setting grading and assessment policies	0.43	0.54	
School improvement planning	0.49	0.45	0.32

Distributed Leadership Roles: *Extent to which Teachers are trusted with Important Roles:* Pearson's

Please indicate the role teachers have at your school in each of the following areas:

Selecting instructional materials			
Devising teaching techniques	0.53		
Setting grading and assessment policies	0.36	0.44	
School improvement planning	0.48	0.38	0.32

Access to Expertise*:* polychoric

Teachers have sufficient access to a broad range of professional support personnel					
Teachers receive feedback that can help them improve	0.45				
Sufficient resources are available for professional development	0.49	0.55			
Professional development deepens teachers content knowledge	0.33	0.52	0.63		
In this school, followup is provided from professional development	0.51	0.61	0.69	0.71	
Provided supports (coaching, PLCs) translate to improvements in instruction	0.43	0.45	0.55	0.56	0.52

Access to Expertise*:* Pearson's

Teachers have sufficient access to a broad range of professional support personnel.					
Teachers receive feedback that can help them improve	0.31				
Sufficient resources are available for professional development	0.36	0.34			
Professional development deepens teachers content knowledge	0.29	0.39	0.43		
In this school, followup is provided from professional development	0.36	0.38	0.37	0.49	
provided supports (coaching, PLCs) translate to improvements in instruction	0.26	0.29	0.25	0.32	0.27

Depth of Interactions: polychoric			
PD provides ongoing opportunities to work with colleagues to refine teaching practices			
Teachers work in PLCs to develop and align instructional practices	0.59		
I frequently plan and coordinate instruction with students' other teachers	0.33	0.40	
I collaborate with other teachers to achieve consistency in how we assess student work	0.50	0.53	0.66

Depth of Interactions: Pearson's			
PD provides ongoing opportunities to work with colleagues to refine teaching practices			
Teachers work in PLCs to develop and align instructional practices	0.48		
I frequently plan and coordinate instruction with students' other teachers	0.24	0.33	
I collaborate with other teachers to achieve consistency in how we assess student work	0.33	0.41	0.57

Appendix C.
Conceptual, Technical, and Empirical Analyses of Survey Items

This appendix provides details on all survey items considered in initial analyses that were dropped from the final models on conceptual, technical, and/or empirical grounds.

Teacher Social Capital: Items Dropped from the Teacher Working Conditions Survey

Based on research literature on teacher social capital, social network theory, trust in schools, and teacher working conditions, as well as related literature on distributed leadership, school climate, and parental involvement, a working conceptual model composed of the factors of Norms, Networks, Access to Expertise, and Depth of Interactions was formed, as detailed in Chapter III. Based on this model, 40 items in the Teacher Working Conditions Survey were selected for initial analysis. Factors named Trust, Cares, Parent-Teacher Networks, Teacher-Teacher Networks, Distributed Leadership, Access to Expertise, and Depth of Interactions were hypothesized (see Table 5.3). These factors were analyzed and items dropped first on the basis of further conceptual analyses and technical analyses of potential item wording difficulties. Second, empirical analyses were conducted on item fit in multilevel and single-level confirmatory factor analyses, including consideration of factor loadings, residuals variances and covariances, R-squared for each item, and modification indices. The goal was to arrive at a parsimonious measurement model that could approximate as many dimensions of teacher social capital as possible while still requiring the estimation of a small enough number of parameters to allow the full multilevel multidimensional structural equation mediation model to be able to be estimated. The final measurement model for teacher social capital was composed of eighteen items measuring four dimensions: Trust, Cares, Parent-Teacher Networks, and Access to

Expertise. Table X1 lists the items dropped and provides a brief description of the grounds for dropping each item. A more complete explanation of these decisions follows.

Table C1.
Teacher Working Conditions Survey Deleted Items:
Summary of Conceptual, Technical, and Empirical Issues Leading to Deletion

Item	**Empirical Issues**	**Conceptual/ Technical Issues**
TEACHER-TEACHER NETWORKS		
Collaborative planning time (in categories of number of hours)	high residual variance; r-squared: .046; large modification indices	Weak relation to any factor
Teachers have time available to collaborate with colleagues	High residual correlations	Weak relation to any factor
The non-instructional time provided for teachers in my school is sufficient	High residual correlations	Weak relation to any factor
DISTRIBUTED LEADERSHIP	Measured with different scale	Non-essential factor dropped for parsimony;
Indicate the role teachers have in…Setting grading and assessment policies		District/state level policy not school policy
Selecting instructional materials and resources		Dropped for parsimony
Devising teaching techniques		Dropped for parsimony
School improvement planning		Dropped for parsimony
Teachers have time available to collaborate with colleagues		Not measuring leadership role
Noninstructional time provided for teachers is sufficient		Not measuring leadership role
PARENT-TEACHER NETWORKS		
Parents support teachers, contributing to students' success	Empirically overfit (negative residual covariances over 0.1)	
Community members support teachers, contributing to st success	Empirically overfit	
The community we serve is supportive of this school	Empirically overfit	
Parents/guardians are influential decision makers in school	Lowest loadings of parent items; high residual variances, covariances; low R-square	Measures power/influence of parents rather than communication of school
Teachers provide parents/guardians with useful information about student learning	Low loadings	Hard for teachers to accurately measure
DEPTH OF INTERACTIONS		
I frequently plan/coordinate instruction with other teachers		Measures individual teacher habit, not shared social capital
I collaborate to achieve consistency in assessing student work		Measures individual teacher habit, not shared social capital
Teachers work in professional learning communities to develop and align instruction		Too specific & dependent on consistent definition of PLCs
CARES		
The faculty and leadership have a shared vision	Lowest loadings on Cares Between & Within	Conceptual confusion: could fit with Trust or Cares
Teachers receive feedback that can help them improve teaching	Second lowest loadings on Cares; high residual variances, covariances; low R-square	Could fit with Expert or Cares

School leadership consistently supports teachers	third lowest factor loadings on Cares; high residual variances, covariances; low R-square	Conceptually loads on Trust & Cares
Leadership makes a sustained effort to address teacher concerns re: Professional Development	Relatively low loadings; high modification indices	Conceptually loads on Expertise and Cares
TRUST		
Teachers are trusted to make sound professional decisions about instruction	High modification indices, cross-loadings	Conceptually overlaps with Cares
ACCESS TO EXPERTISE		
Teachers have sufficient access to a broad range of professional support personnel	High negative residual covariances (6 over.1)	Parsimony
Provided supports (coaching, PLCs, etc.) translate to improvements in instruction	Relatively low loadings	Long causal chain; parsimony

Conceptual / Technical Issues. The Distributed Leadership construct was dropped because this construct stands conceptually apart from typical definitions of teacher social capital – distributed leadership is related to teacher social capital but has not been defined in the literature as part of the construct, whereas norms such as trust and caring, the extent and strength of networks, and the level of access that these networks provide to important social and knowledge resources are nearly always included as part of teacher social capital. In order to achieve a parsimonious and estimable model, this construct was dropped. The Distributed Leadership items were also measured using a different scale (small to large role) that likely introduced method effects that would complicate interpretations.

The Depth of Interactions dimension was dropped because several of the items were re-evaluated as measuring individual teacher habits, not shared social capital. For example, the item *I frequently plan/coordinate instruction with other teachers* measures a positive attribute of the individual teacher respondent that is likely associated with positive outcomes for students and teachers, but this item would have to be aggregated to the school level to be interpreted as measuring a shared aspect of the school's staff. Another item, *Teachers work in professional learning communities to develop and align instruction,* was considered too specific and dependent on a consistent working definition of Professional Learning Communities that might not exist in practice. For instance, PLCs could work on collaboratively assessing student work instead – with equal or greater effectiveness, but this would not be rated highly by a respondent who interpreted the question strictly.

One of the Parent-Teacher Networks items, *Parents/guardians are influential decision makers in school*, was considered for deletion because it measured the power and influence of parents rather than the quality of the school's communication efforts. In other words, this item was judged to measure, at least in part, a construct residing in parents rather than an aspect of schools. Another Parent item, *Teachers provide parents/guardians with useful information about student learning*, was dropped because it was judged to be hard for teachers to accurately measure other teachers' provision of this information.

Several items within the Cares and Trust dimensions were dropped because their inclusion effectively blurred the distinctions among dimensions. These items were not considered flawed in and of themselves, but simply did not fit cleanly into one dimension or another.

Empirical Issues. The Teacher-Teacher Networks factor was dropped because three of the four items had poor fit to the model, with high residual covariances and variances. For instance, the item measuring collaborative planning time in categories of number of hours was dropped due to a large residual variance corresponding to an R-squared value of .046, meaning that the model explained less than five percent of the variance in this item, along with modification indices that suggested this item fit the model poorly. One item, *Appropriate amount of time provided for professional development*, was re-assigned to the Expertise factor.

Examination of the residual correlations also suggested the model overfit the correlations among several indicators for the Parent-Teacher Networks factor. Three indicators in this dimension had negative residuals over 0.1, suggesting that the model imposed a stronger relationship among these indicators than is found in the data. These three items were dropped.

Teaching Practices: Items Dropped from the Tripod Student Perception Survey

Initial analyses used the full set of thirty-six Tripod items. Items were dropped based on conceptual analyses, technical analyses of potential item wording difficulties, and empirical analyses of item fit, including factor loadings, residuals variances and covariances, R-squared, and modification indices. The goal was twofold: a parsimonious measurement model requiring the estimation of a small enough number of parameters to allow for the full multilevel structural equation model to be estimated, and a model that could approximate two dimensions of teaching practice to accord with the structure found in exploratory factor analyses, as well as in prior research by Kuhfeld (under review), and findings reported in the MET User's Guide. The final model was composed of nineteen items measuring two dimensions: termed Control and Instruction. Table X2 lists the items dropped and provides a brief description of the grounds for dropping each item. A more complete explanation of these decisions follows.

Table C2.

Tripod Deleted Items: Summary of Conceptual and Technical Issues Leading to Deletion

Var.	Item Wording	Conceptual/ Technical Issues	Similar Items Retained
	INSTRUCTION		
B146	My teacher seems to know if something is bothering me.	Mind-reading	A10, B34
B130	My teacher knows when the class understands, and when we do not.	Mind-reading	B1, B147
B136	When s/he is teaching, my teacher thinks we understand even when we don't.*	Negative phrasing; mind-reading	B1, B147
B128	My teacher asks questions to be sure we are following along when s/he is teaching.	Negative connotation	B147
B133	My teacher asks students to explain more about the answers they give.	Negative connotation; potential for student misunderstanding	B154

B21	In this class, my teacher accepts nothing less than our full effort.	Negative phrasing	B70
B36	My teacher doesn't let people give up when the work gets hard.	Negative phrasing	
B45	My teacher wants us to use our thinking skills, not just memorize things.	Mind-reading; negative phrasing	B154, A54
B59	My teacher wants me to explain my answers – why I think what I think	Mind-read; negative connotation	B154, A54
B141	This class does not keep my attention – I get bored.*	Negative phrasing	B29, B44
B129	My teacher wants us to share our thoughts	Might align positively with off-task behavior; Mind-read; negative connotation	B154, A54
B135	Students get to decide how activities are done in this class.	Might align positively with out-of-control classroom	A54
B155	Students speak up and share their ideas about class work.	Might align positively with off-task behavior	A54
B145	My teacher takes the time to summarize what we learn each day	Too specific	B80
	CONTROL		
B113	I hate the way that students behave in this class.*	Negative phrasing emotionally-loaded language;	B112, B138
B114	Student behavior in this class makes the teacher angry.*	Negative phrasing; emotionally-loaded language	B112, B138

* Shading separates items by developers' original conceptualization of seven dimensions (7 Cs).

Conceptual Issues. Students' expertise as raters is based on their natural sensitivity to the meaningfulness of the work they are expected to do, and the teaching practices that support or constrain their effective participation and understanding. Several items asked students to report not on these aspects of the classroom, but on the teacher's desires, motivations, state of mind, or level of understanding. Illustrative of this type of item is B130: My teacher knows when the class understands, and when we do not, as well as B129, B130, B45 and B146. Items rest on a possibly flimsy foundation when they are based on asking students to infer the state of mind of their teacher rather than on their "naturally acquired expertise" as learners in that classroom (Wallace, Kelsey, & Ruzek, 2016, p. 3). These items were dropped.

Technical Issues. The first technical issue was items including negative phrasing (B130, B136, B21, B36, and B141). Some of these items are intended to be reverse coded, but some of them just include a negative phrase (i.e., B21: In this class, my teacher accepts nothing less than our full effort). Negatively worded items are linguistically more difficult to comprehend and interpret as intended and are thus likely to be subject to more error (DeVellis, 2003). DeVellis recommends avoiding the use of these items entirely. In an analysis of the Tripod survey using the MET dataset, Wallace, Kelsey, & Ruzek (2016) noted empirical issues with the fit of the negatively worded Tripod items and employed a bifactor model which included a nuisance factor. Only the negatively worded items loaded on this nuisance factor, thus accommodating the correlations among these items that appeared to be driven by the negative item wording rather than by conceptual similarities among the items. For this dissertation, items with negative wording were dropped from the final model.

The next technical issues concerned items that appeared likely to be understood differently by student raters than the item developers had intended. First on this list was B135: S*tudents get to decide how activities are done in this class*. This item appears intended to measure the extent to which student perspectives are honored. However, on its face, it seems that a teacher with little control over his classroom might score more highly on this item than a strict and effective teacher. This item had the highest residual correlations of any Tripod item in initial factor analyses and was also noted as poorly performing by Kuhfeld (under review).

Two other items, B129 and B155, were similarly intended to tap into teachers' regard for student perspectives, but positive responses could be read as aligning not with quality teaching but with a classroom that wastes time debating procedures and airing student complaints.

Additionally, B145: My teacher takes the time to summarize what we learn each day, was added to the list of items to potentially drop because it asks students to report on a teacher behavior that might be too specific. Summarizing learning each day is a recommended practice, but it could easily be done robotically and without being helpful. This recommendation is akin to the recently popular mandate to post learning goals and standards on the board or in lesson plans;

it is easy to imagine a good teacher who does not write learning goals on the board or summarize the learning each day and a less effective teacher who does this practice, so item B145 may not help distinguish between levels of quality teaching.

B59, B128, and B129 were identified base on empirical misfit in MCFA and MEFA analyses. A multilevel exploratory factor analysis on 17 *Instruction* items, to shed further light on the factor structure, suggested two Within factors and one Between factor, and this model had good to adequate fit. However, all items loaded strongly on the first Within factor and the second Within factor was not readily interpretable (see Table X3). Re-analysis of the item wordings in light of the item loadings on the second Within factor suggested that this factor was picking up on negative and top-down connotations in items B128, B129, and B59. The negative connotations implicit in these items stood out in contrast to items B147, B154, and A54. To give an example, B128, one of the items that displayed substantial empirical misfit, describes the effective teacher as one who “asks questions to be sure we are following along”, a negative, top-down connotation that implies student resistance to learning, whereas B147 describes the teacher as one who “checks to make sure we understand”, a positive, student-centered connotation. Similarly, B129 describes a teacher who “asks students to explain more” implying a teacher desire for something lacking in students, whereas B154 and A54 describe a teacher who “gives us time to explain” and “respects my ideas”. In the latter two items student thinking and regard for student perspectives are front and center. It appeared that the three items displaying misfit might be positively endorsed by some students in response to a nagging teacher who asked a lot of questions rather than a challenging teacher who pushed students to think more deeply.

Table C3
Factor Structure for MEFA on 17 Instruction Items (Tripod Survey)

	Five Largest Eigenvalues				
	1	2	3	4	5
WITHIN	7.729	1.113	0.769	0.746	0.677
BETWEEN	14.71	0.709	0.5	0.33	0.184

WITHIN FACTOR STRUCTURE: Factor Loadings on

Item	Factor 1	Factor 2
PA10	0.666	0.233
PB34	0.666	0.215
PA54	0.701	0.135
PB154	0.678	0.063
PB1	0.663	0.038
PB17	0.699	0.107
PB80	0.702	0.116
PB29	0.704	0.503
PB44	0.705	0.495
PB147	0.714	0.016
PB58	0.642	0.087
PB83	0.668	0.125
PB128	0.496	-0.127
PB133	0.488	-0.186
PB59	0.55	-0.148
PB70	0.588	0.034
PB90	0.671	-0.028

RMSEA: 0.038; CFI: 0.957; TLI: 0.948; SRMR Within: 0.025; SRMR Between: 0.043

Items B113, B114, and B6 also appeared to have technical problems in their wording. B113 and B114 both use emotionally-loaded language that might bias responses, and B113 asks students to respond with their own feelings (“I hate the way…”) about the classroom climate rather than with the more-objective assessment asked for in items B112 or B138. The use of the word “hate” might skew responses because it is a strong word: some might be reluctant to say they hate anything. Also, a truly out of control class may have a number of poorly behaved students who may not want to attack their own or their friends’ behavior. As for B114, this item asks students to infer the feelings of the teacher, which may prove difficult, and may also not have a direct connection to the level of classroom control that actually exists – a teacher in an out-of-control classroom might feel sad or depressed or hopeless rather than angry, or a teacher might be comfortable with a degree of chaos that is not conducive to learning. Any of these situations would lead to this item not correlating with the actual level of classroom control.

Item B6 appears to measure two separate things within the same item – whether the class stays busy and whether the class doesn't waste time. This may make it difficult for students to respond accurately to this item. It is easy to imagine, and unfortunately might not be difficult to find, a class that stays busy but wastes a lot of time on busy-work. However, item B6 addresses issues of time on task that appear to be an important part of the construct of *Control,* but that are not addressed in other items. Based on these considerations, items B113 and B114 were deleted and a multilevel confirmatory factor analysis was re-estimated for Control. This five-item conceptualization of Control had good fit and was retained in the final model along with the fourteen items measuring INSTRUCTION. Fit indices and further information on the final measurement model for teaching practice as measured by the Tripod items are provided in chapter VI.

Appendix D.
Equations and Syntax for Multidimensional Multilevel Structural Equation Model

With regard to the teacher social capital factors, a measurement model estimates each Teacher Working Conditions survey item response, WC_{kj}, as a function of the underlying latent Between-level teacher social capital factors, $\boldsymbol{\eta}_1 - \boldsymbol{\eta}_4$, and residual u_{kj}

$$WC_{kj} = \boldsymbol{\Lambda}_{\boldsymbol{k}}^{\boldsymbol{B}} \boldsymbol{\eta}^{\boldsymbol{B}} + \mathrm{e}_{kj} \qquad (7.1)$$

where *k* indexes the WC items. *B* indicates the Between classroom/teacher portion (this could be considered akin to an index for classroom *j*) and *j* indexes the classroom/teacher. $\boldsymbol{\Lambda}$ is a matrix of factor loadings of items on latent teacher social capital factors $\boldsymbol{\eta}_1 - \boldsymbol{\eta}_4$. WC_{kj} indicates the response of the *j*th teacher to the *k*th item on the Teacher Working Conditions survey. e_{kj} is the between level residual variance for the *k*th item.

For the teaching practice factors, $\boldsymbol{\eta}_5, \boldsymbol{\eta}_6$, the measurement model includes Within and Between portions. First, each student survey item response is decomposed into a Between-classroom portion and a Within-classroom deviation from that Between portion.

$$TR_{kij} = TR_{kj} + (TR_{kij} - TR_{kj}) \qquad (7.2)$$

Each Tripod student survey response TR_{kij} is estimated as a function of Between and Within teacher latent factors, η^B and η^W, and Between and Within-teacher residual variance, $\mathrm{e}_{kj} + \varepsilon_{kij}$.

$$TR_{kij} = \boldsymbol{\Lambda}_{\boldsymbol{k}}^{\boldsymbol{B}} \boldsymbol{\eta}^{\boldsymbol{B}} + \boldsymbol{\Lambda}_{\boldsymbol{k}}^{\boldsymbol{W}} \boldsymbol{\eta}^{\boldsymbol{W}} + \mathrm{e}_{kj} + \varepsilon_{kij} \qquad (7.3)$$

The structural equations predicting post mathematics achievement, MATH10 are:

Level 1: $MATH10_{ij} = Ach_7^B + \Gamma_1 \mathrm{Math09}_{ij} + \Gamma_2 \mathrm{Race}_{ij} + \Gamma_3 \mathrm{Gender}_{ij} + \Gamma_4 \mathrm{grade}_{ij} + \zeta_{ij}$

Level 2:

$$Ach_7^B = \gamma_{00} + \gamma_{01}\eta_{1j} + \gamma_{02}\eta_{2j} + \gamma_{03}\eta_{3j} + \gamma_{04}\eta_{4j} + \gamma_{05}\eta_{5j} + \gamma_{06}\eta_{6j} +$$

$$\gamma_{05} Sch\%BlackHisp_j + \gamma_{06} District_j + \gamma_{07} CRMath09_j + \gamma_{08} CRBlackHisp_j + r_j$$

$$\Gamma_{1_j} = \gamma_{10}$$

$$\Gamma_{2_j} = \gamma_{20}$$

$$\Gamma_{3_j} = \gamma_{30}$$

$$\Gamma_{4_j} = \gamma_{40}$$

$$\Gamma_{5_j} = \gamma_{50}$$

$$\Gamma_{6_j} = \gamma_{60}$$

$$\Gamma_{7_j} = \gamma_{70} \qquad (7.4)$$

where $MATH10_{ij}$ is state math test score outcome for the *i*th student in classroom *j*. Student math achievement is modeled as a function of the classroom mean/intercept, Ach_7^B; student level covariates, Math09_{ij}, Race_{ij}, Gender_{ij}, and grade_{ij}; the effects of the teacher social capital factors, η_{1-4j}: Trust, Cares, Parent, and Expert, and the teaching practice factors, $\eta_{5,6j}$; the school's racial composition, $Sch\%BlackHisp_j$; district fixed effects, $District_j$; classroom prior achievement, $CRMath09_j$; classroom racial composition, $CRBlackHisp_j$; and student and teacher level error terms, ζ_{ij} and r_j. It is important to note that the four teacher social capital factors, $\boldsymbol{\eta}_{1-4j}$ are correlated, as are the two teaching practice factors. The $\Gamma_{1_j} = \gamma_{10} \ldots \Gamma_{7_j} = \gamma_{70}$ equations include no random effects/error terms, reflecting the assumption that the effects of these variables on classroom math achievement do not vary across classrooms.

The equations predicting the Between latent factors are:

$$\boldsymbol{\eta}_j = \alpha + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{KX}_j + u_j \qquad (7.5)$$

where α is the mean/intercept; $\boldsymbol{\eta}_j$ is a 6 x 1 vector of the 6 latent factors; $\boldsymbol{\beta}$ is a matrix of the coefficients containing the effect estimates of the teacher social capital factors on the teaching practice factors, with 0s in all cells where effects are constrained to 0; $\boldsymbol{X}_j$ is a vector of covariates regressed on the latent factors, including *Sch%BlackHisp*, *CRMath09*, and *CRBlackHisp*; *K* is a 6 x 3 matrix (6 rows for the 6 latent factors and 3 columns for the 3 covariates), once again with 0s in cells indicating where that particular factor is not regressed on that covariate; u_j is the residual error term. Note that in the case of the teacher social capital factors, the equations 7.5 reduce to a regression of each factor on the Sch%BlackHisp variable, along with an intercept and error term. At the Within level, the teaching practice factors are not predicted by any other variables, and so are defined solely by the measurement models above.

Mplus Syntax
! Indicates comments in Mplus, not included as part of the command syntax

```
Model:
%Between%
  ! Teacher Social Capital -unidimensional
! norms trust
   trustB BY We21exp
    We21tres
    WL21rais
    We21proc
    We21solv
    We21lead;

  ! admin cares for Ts
   caresB BY WL21ef
    WL21eftm
    WL21eftl
    WL21efip;

 ! T-P Networks
   parentsB BY
    Wc21com
    Wc21enc
    Wc21kno ;
  ! expertise
   expertB BY
    Wp21foll
```

```
   Wp21suff
   Wp21deep
   Wp21time
   Wp21coll
    ;

 ! TRIPOD
  ! teaching control
   controlB BY
     PB112 (18)
     PB138R (19)
     PB46 (20)
     PB49 (21)
     PB6 (22);
  ! teaching instruction
  instructB BY
        PA10  (17)
        PB34 (1)
        PA54 (2)
        PB154 (3)
        PB1 (4)
        PB17 (5)
        PB80 (6)
        PB29 (7)
        PB44 (8)
        PB147 (9)
        PB58 (10)
        PB83 (11)
        PB70 (15)
        PB90 (16);

  ! correlate the Teaching practice factors
    instructB WITH controlB;
  !correlate the TSC factors
    trustB WITH caresB parentsB expertB;
    caresB WITH parentsB expertB;
    parentsB WITH expertB;

!covariates on teacher social capital
  trustB ON L3BH (tL3BH);
  caresB ON L3BH (tL3BH);
  parentsB ON L3BH (tL3BH);
  expertB ON L3BH (tL3BH);
!mediation paths
  controlB ON trustB(actru);
  controlB ON caresB(accar);
```

```
  controlB ON parentsB(acpar);
  controlB ON expertB(acexp);
  instructB ON trustB(aitru);
  instructB ON caresB(aicar);
  instructB ON parentsB(aipar);
  instructB ON expertB(aiexp);
!covariates on teaching practice
  controlB ON CRmath09 (cCRmath9);
  instructB ON CRmath09 (iCRmath9);

!outcome
  MATH10t ON dist18 dist33 dist71;
  MATH10t ON L3BH (mathL3BH);
  MATH10t ON CRmath09 (mCRmath9);
  MATH10t ON controlB (bcontrol);
  MATH10t on instructB (binstruct);
  MATH10t ON trustB (c1);
  MATH10t ON caresB (c2);
  MATH10t ON parentsB (c3);
  MATH10t ON expertB (c4);

%Within%
 ! instruction - student perceptions
   instructW BY
      PA10  (17)
      PB34 (1)
      PA54 (2)
      PB154 (3)
      PB1 (4)
      PB17 (5)
      PB80 (6)
      PB29 (7)
      PB44 (8)
      PB147 (9)
      PB58 (10)
      PB83 (11)
      PB70 (15)
      PB90 (16);
 ! control - student perceptions
   controlW BY
     PB112 (18)
     PB138R (19)
     PB46 (20)
     PB49 (21)
     PB6 (22);
```

```
 ! correlate teaching practices within
   controlW WITH instructW;

!outcome
  MATH10t ON MATH09t (prior);
  MATH10t ON S_BLACK (black);
  MATH10t ON S_HISP (hisp);
  MATH10t ON S_ELL (ell);
  MATH10t ON S_MALE (male);
  MATH10t ON grade6 (gr6);
  MATH10t ON grade7 (gr7);
```

Appendix E.
Additional Results Tables

Table E1

Variances, Residual Variances and Model Fit: MSEM Multidimensional vs. One-by-one Models

	MSEM Multidimensional				Factors 1x1			
	Est	SE	est/SE	p-val	Est	SE	est/SE	p-val
					TRUST – CONTROL – MATH10			
Variances								
CONTROL	0.61	0.01	45.21	0.00	0.50	0.02	31.77	0.00
Residual Variances								
TRUST	0.94	0.02	38.92	0.00	0.98	0.02	61.53	0.00
CONTROL Btw	0.91	0.03	32.90	0.00	0.93	0.02	40.05	0.00
MATH10 Within	0.37	0.01	42.14	0.00	0.37	0.01	42.16	0.00
MATH10 Btw	0.54	0..04	13.54	0.00	0.56	0.04	13.90	0.00
					CARES – CONTROL – MATH10			
Variances								
CONTROL					0.50	0.02	31.77	0.00
Residual Variances								
CARES	0.96	0.02	54.62	0.00	0.98	0.01	71.14	0.00
CONTROL Btw					0.93	0.03	36.93	0.00
MATH10 Within	See above estimates. Model estimated with all factors at same time.				0.37	0.01	42.16	0.00
MATH10 Btw					0.55	0.04	13.38	0.00
					PARENT – CONTROL – MATH10			
Variances								
CONTROL					0.50	0.02	31.77	0.00
Residual Variances								
PARENT	0.95	0.03	38.24	0.00	0.85	0.04	23.32	0.00
CONTROL Btw					0.93	0.02	37.92	0.00
MATH10 Within					0.37	0.01	42.16	0.00
MATH10 Btw					0.55	0.04	13.81	0.00
					EXPERT – CONTROL – MATH10			
Variances								
CONTROL					0.50	0.02	31.77	0.00
Residual Variances								
EXPERT	0.96	0.02	57.57	0.00	0.99	0.01	112.46	0.00
CONTROL Btw					0.92	0.03	34.13	0.00
MATH10 Within					0.37	0.01	42.16	0.00
MATH10 Btw					0.57	0.04	13.82	0.00

Multidimensional Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.070
TRUST-CONTROL Fit: RMSEA: 0.016; CFI: 0.979; TLI: 0.975; SRMR Within: 0.019; SRMR Between: 0.071
CARES-CONTROL Fit: RMSEA: 0.016; CFI: 0.981; TLI: 0.976; SRMR Within: 0.019; SRMR Between: 0.080
PARENT-CONTROL Fit: RMSEA: 0.017; CFI: 0.981; TLI: 0.977; SRMR Within: 0.019; SRMR Between: 0.070
EXPERT-CONTROL Fit: RMSEA: 0.017; CFI: 0.978; TLI: 0.974; SRMR Within: 0.019; SRMR Between: 0.080
Variances are restricted to be invariant across levels.

Table E2

Variances, Residual Variances and Model Fit: MSEM Multidimensional vs. One-by-one Models

	MSEM Multidimensional				Factors 1x1			
	Est	SE	est/SE	p-val	Est	SE	est/SE	p-val
					TRUST – INSTRUCT – MATH10			
Variances								
INSTRUCT	0.61	0.01	45.21	0.00	0.60	0.01	45.07	0.00
Residual Variances								
TRUST	0.94	0.02	38.92	0.00	0.98	0.02	60.60	0.00
INSTRUCT Btw	0.91	0.03	32.90	0.00	0.999	0.00	279.86	0.00
MATH10 Within	0.37	0.01	42.14	0.00	0.37	0.01	42.12	0.00
MATH10 Btw	0.54	0..04	13.54	0.00	0.64	0.04	14.56	0.00
					CARES – INSTRUCT – MATH10			
Variances								
INSTRUCT					0.60	0.01	45.07	0.00
Residual Variances								
CARES	0.96	0.02	54.62	0.00	0.98	0.01	69.04	0.00
INSTRUCT Btw					0.999	0.00	279.13	0.00
MATH10 Within	See above estimates. Model estimated				0.37	0.01	42.12	0.00
MATH10 Btw	with all factors at same time.				0.63	0.04	14.19	0.00
					PARENT – INSTRUCT – MATH10			
Variances								
INSTRUCT					0.60	0.01	45.07	0.00
Residual Variances								
PARENT	0.95	0.03	38.24	0.00	0.85	0.04	23.20	0.00
INSTRUCT Btw					0.998	0.00	228.22	0.00
MATH10 Within					0.37	0.01	42.10	0.00
MATH10 Btw					0.62	0.04	14.09	0.00
					EXPERT – INSTRUCT – MATH10			
Variances								
INSTRUCT					0.60	0.01	45.07	0.00
Residual Variances								
EXPERT	0.96	0.02	57.57	0.00	0.99	0.01	113.20	0.00
INSTRUCT Btw					0.991	0.01	98.09	0.00
MATH10 Within					0.37	0.01	42.11	0.00
MATH10 Btw					0.65	0.05	14.54	0.00

Multidimensional Model Fit: RMSEA: 0.023; CFI: 0.933; TLI: 0.928; SRMR Within: 0.038; SRMR Between: 0.070
TRUST- INSTRUCT Fit: RMSEA: 0.030; CFI: 0.938; TLI: 0.933; SRMR Within: 0.036; SRMR Between: 0.066
CARES- INSTRUCT Fit: RMSEA: 0.031; CFI: 0.938; TLI: 0.933; SRMR Within: 0.035; SRMR Between: 0.069
PARENT- INSTRUCT Fit: RMSEA: 0.031; CFI: 0.938; TLI: 0.933; SRMR Within: 0.035; SRMR Between: 0.069
EXPERT- INSTRUCT Fit: RMSEA: 0.030; CFI: 0.937; TLI: 0.932; SRMR Within: 0.035; SRMR Between: 0.071
Variances are restricted to be invariant across levels.

Table E3

Covariate Effects in Multidimensional Model vs Separate Models with One Factor at a Time; Between Level: (N = 15,644 students; 520 teachers)

MATH10 on Covariates						TRUST – Control – MATH10				
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	107.96	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.56	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.75	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.06	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.28	0.20	
GRADE6	0.01	0.02	0.61	0.54		0.01	0.01	0.76	0.45	
GRADE7	0.00	0.01	-0.22	0.83		0.00	0.01	-0.17	0.86	
School % Black & Hisp	-0.07	0.06	-1.18	0.24		-0.11	0.06	-1.89	0.06	*
CR Prior Math Achieve	0.39	0.06	6.83	0.00	***	0.39	0.05	7.07	0.00	***
						CARES – Control – MATH10				
MATH09	See above estimates. Model estimated with all factors at same time.					0.76	0.01	108.02	0.00	***
S_BLACK						-0.09	0.01	-11.57	0.00	***
S_HISP						-0.05	0.01	-6.72	0.00	***
S_ELL						-0.02	0.01	-3.05	0.00	***
S_MALE						-0.01	0.01	-1.29	0.20	
GRADE6						0.01	0.01	0.73	0.47	
GRADE7						0.00	0.01	-0.13	0.89	
School % Black & Hisp						-0.10	0.06	-1.85	0.06	*
CR Prior Math Achieve						0.39	0.05	7.15	0.00	***
						PARENT-CONTROL-MATH10				
MATH09						0.76	0.01	107.94	0.00	***
S_BLACK						-0.09	0.01	-11.55	0.00	***
S_HISP						-0.05	0.01	-6.70	0.00	***
S_ELL						-0.02	0.01	-3.10	0.00	***
S_MALE						-0.01	0.01	-1.28	0.20	
GRADE6						0.01	0.01	0.70	0.48	
GRADE7						0.00	0.01	-0.17	0.87	
School % Black & Hisp						-0.05	0.06	-0.89	0.38	
CR Prior Math Achieve						0.39	0.05	7.09	0.00	***
						EXPERT-CONTROL-MATH10				
MATH09						0.76	0.01	108.02	0.00	***
S_BLACK						-0.09	0.01	-11.55	0.00	***
S_HISP						-0.05	0.01	-6.72	0.00	***
S_ELL						-0.02	0.01	-3.08	0.00	***
S_MALE						-0.01	0.01	-1.29	0.20	
GRADE6						0.01	0.01	0.77	0.44	

GRADE7						0.00	0.01	0.00	1.00	
School % Black & Hisp						-0.12	0.06	-2.15	0.03	**
CR Prior Math Achieve						0.39	0.06	6.92	0.00	***

Table E4

Covariate Effects in Multidimensional Model vs Separate Models with One Factor at a Time; Between Level: (N = 15,644 students; 520 teachers)

MATH10 on Covariates						TRUST – Instruct – MATH10				
MATH09	0.76	0.01	107.95	0.00	***	0.76	0.01	108.12	0.00	***
S_BLACK	-0.09	0.01	-11.55	0.00	***	-0.09	0.01	-11.58	0.00	***
S_HISP	-0.05	0.01	-6.71	0.00	***	-0.05	0.01	-6.64	0.00	***
S_ELL	-0.02	0.01	-3.07	0.00	***	-0.02	0.01	-3.10	0.00	***
S_MALE	-0.01	0.01	-1.29	0.20		-0.01	0.01	-1.29	0.20	
GRADE6	0.01	0.02	0.61	0.54		-0.01	0.02	-0.70	0.49	
GRADE7	0.00	0.01	-0.22	0.83		-0.01	0.01	-0.55	0.58	
School % Black & Hisp	-0.07	0.06	-1.18	0.24		-0.08	0.06	-1.30	0.19	
CR Prior Math Achieve	0.39	0.06	6.83	0.00	***	0.50	0.05	9.70	0.00	***
						CARES – Instruct – MATH10				
MATH09	See above estimates. Model estimated with all factors at same time.					0.76	0.01	108.15	0.00	***
S_BLACK						-0.09	0.01	-11.60	0.00	***
S_HISP						-0.05	0.01	-6.62	0.00	***
S_ELL						-0.02	0.01	-3.09	0.00	***
S_MALE						-0.01	0.01	-1.30	0.19	
GRADE6						-0.01	0.02	-0.73	0.47	
GRADE7						-0.01	0.01	-0.53	0.60	
School % Black & Hisp						-0.07	0.06	-1.24	0.22	
CR Prior Math Achieve						0.51	0.05	9.79	0.00	***
						PARENT – Instruct – MATH10				
MATH09						0.76	0.01	108.08	0.00	***
S_BLACK						-0.09	0.01	-11.58	0.00	***
S_HISP						-0.05	0.01	-6.60	0.00	***
S_ELL						-0.02	0.01	-3.13	0.00	***
S_MALE						-0.01	0.01	-1.29	0.20	
GRADE6						-0.01	0.01	-0.79	0.43	
GRADE7						-0.01	0.01	-0.59	0.56	
School % Black & Hisp						-0.01	0.06	-0.21	0.83	
CR Prior Math Achieve						0.50	0.05	9.62	0.00	***
						EXPERT – Instruct – MATH10				
MATH09						0.76	0.01	108.17	0.00	***

S_BLACK	-0.09	0.01	-11.58	0.00	***
S_HISP	-0.05	0.01	-6.62	0.00	***
S_ELL	-0.02	0.01	-3.12	0.00	***
S_MALE	-0.01	0.01	-1.30	0.20	
GRADE6	-0.01	0.02	-0.69	0.49	
GRADE7	-0.01	0.01	-0.39	0.69	
School % Black & Hisp	-0.09	0.06	-1.56	0.12	
CR Prior Math Achieve	0.50	0.05	9.56	0.00	***

BIBLIOGRAPHY

Adams, P. (2016). School districts of all sizes experiencing teacher shortages. *Journal Star*. Peoria, IL. http://www.pjstar.com/news/20160113/school-districts-of-all-sizes-experiencing-teacher-shortages

Ali-Hawkins, A. (2015). The Relationship Between Middle School Configurations and Student Success. (Doctoral Dissertation) Walden University ScholarWorks.

Anderman, E. M., Maehr, M. L., & Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research & Development in Education 32*(3), 131-147.

Angrist, J. D. and Lavy, V. (2001). Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools. *Journal of Labor Economics 19*(2), 343 – 369.

Anson, A. R., Cook, T. D., Habib, F., Grady, M. K., Haynes, N., & Comer, J. P. (1991). The Comer School Development Program: A theoretical analysis. *Urban Education, 26*, 56--82.

ASCD (2007). The Whole Child Initiative: Challenged. http://www.ascd.org/programs/The-Whole-Child/Challenged.aspx

Bacolod, M. (2007). Who Teaches and Where They Choose to Teach: College Graduates of the 1990s. *Educational Evalution and Policy Analysis.* vol. 29. pp. 155 - 168. http://epa.sagepub.com/cgi/content/abstract/29/3/155

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... & Shepard, L. A. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. EPI Briefing Paper# 278. *Economic Policy Institute*.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*(2), 135–167.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological methods*, *11*(2), 142.

Becker, H. J., & Epstein, J. L. (1982). Parent Involvement: A Survey of Teacher Practices. *The Elementary School Journal*, *83*(2), 85–102. Retrieved from http://www.jstor.org/stable/1001098

Bedard, K., & Do, C. (2005). Are middle schools more effective? The impact of school structure on student outcomes. *Journal of Human Resources*, *40*(3), 660-682.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, *57*(5), 785-799.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, *17*(2-3), 62-87.

Benson, P. L. (2007). Developmental Assets: An overview of theory, research, and practice. In R. K. Silbereisen & R. M. Lerner (Eds.), Approaches to positive youth development (p. 33-58). Thousand Oaks, CA: Sage.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, *17*(1), 48-62.

Bill & Melinda Gates Foundation. (2012, October). *Measures of Effective Teaching: Grantee Files, 2009-2011 and User Guide*. Ann Arbor, MI.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, *5*(1), 7-74.

Bourdieu, P. (1986). The forms of capital. in Richardson, J., Handbook of Theory and Research for the Sociology of Education (1986), Westport, CT: Greenwood, pp. 241–58.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? Teacher attrition and student achievement* (No. w14022). National Bureau of Economic Research.

Bryk, A. S., & Schneider, B. L. (1996). *Social trust: A moral resource for school improvement*. US Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center. http://ccsr.uchicago.edu/publications/social-trust-moral-resource-school-improvement

Bryk, A. S., & Schneider, B. L. (2003). Trust in Schools: A Core Resource for School Reform. *Educational Leadership, 60(6)*, 40 – 45.

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

California Commission on Teacher Credentialing (2009). California Standards for the Teaching Profession. http://www.ctc.ca.gov/educator-prep/standards/CSTP-2009.pdf

Camburn, E. M. (2012, November). Review of "Asking Students about Teaching". Boulder, CO: National Education Policy Center.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal*, 3-20.

Carver, R. P. (1975). The Coleman Report: Using inappropriately designed achievement tests. *American Educational Research Journal*, *12*(1), 77-86.

Center for Education Policy Research (2012). Strategic Data Project Human Capital Diagnostic: Los Angeles Unified School District. Strategic Data Project. Cambridge, MA: Harvard University. http://cepr.harvard.edu/cepr-resources/files/news-events/sdp-lausd-hc.pdf

Chávez, R. C. (1984). The use of high-inference measures to study classroom climates: A review. *Review of Educational Research*, *54*(2), 237-261.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2011). The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. National Bureau of Economic Research. NBER Working Paper No. 17699. http://www.nber.org/papers/w17699NATIONAL

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2013). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. National Bureau of Economic Research. NBER Working Paper 19424. http://www.nber.org/papers/w19424

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs. *Princeton, NJ: Mathematica Policy Research*.

Clotfelter, C., Ladd, H. F., and Vigdor, J. (2007). How and Why do Teacher Credentials Matter for Student Achievement? *Calder Working Paper 2*. National Center for Analysis of Longitudinal Data in Education Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (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. L., & Wheeler, J. (2007). High Poverty Schools and the Distribution of Teachers and Principals. *Calder Working Paper 1*. National Center for Analysis of Longitudinal Data in Education Research.

Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, *30*(3), 203-235.

Coburn, C. E., Mata, W. S., & Choi, L. (2013). The Embeddedness of Teachers' Social Networks: Evidence from a Study of Mathematics Reform. *Sociology of Education*, *86*(4), 311-342.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education*, *119*(1), 137-182.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology*, S95-S120.

Coleman, J.S., Campbell, E. Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*, 2 volumes. (OE-38001; Superintendent of Documents Catalog No. FS5.238:38001). Washington, D.C.: U.S. Government Printing Office. http://files.eric.ed.gov/fulltext/ED012275.pdf

Cook, K.F., Kallen, M.A., & Amtmann, D. (2009). Having a fit: Impact of Number of Items and Distribution of Data on Traditional Criteria for Assessing IRT's Unidimensionality Assumption. *Qual Life Res.18*(4): 447–460. doi:10.1007/s11136-009-9464-4

Cook, P. J., MacCoun, R., Muschkin, C., & Vigdor, J. (2008). The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management*, *27*(1), 104-121.

Corcoran, T. B. (1995). Helping Teachers Teach Well: Transforming Professional Development. *CPRE Policy Briefs.* Consortium for Policy Research in Education. New Brunswick, NJ.

Correnti, R. & Martinez, J.F. (2012). Conceptual, Methodological, and Policy Issues in the Study of Teaching: Implications for Improving Instructional Practice at Scale. *Educational Assessment*, *17*:2-3, 51-61. http://dx.doi.org/10.1080/10627197.2012.717834

Danielson, C. (1996). Enhancing Professional Practice: A Framework for Teaching. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007). Enhancing Professional Practice: A Framework for Teaching. (2nd Ed.) Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education policy analysis archives*, *8*, 1. http://epaa.asu.edu/ojs/index.php/epaa/article/view/392

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness. *education policy analysis archives*, *13*(42), n42.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Desimone, L., & LeFloch, K. (2004). Probing the “trickle down” effect of standards and assessments: Are we asking the right questions. *Educational Evaluation and Policy Analysis*, *26*(1), 1-22.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., and Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction: Results from a Three-Year Longitudinal Study. *Educational Evaluation and Policy Analysis.* American Educational Research Association Vol. 24, No. 2. pp. 81-112. Stable URL: http://www.jstor.org/stable/3594138

Dove, M. J., Pearson, L. C., & Hooper, H. (2010). Relationship between grade span configuration and academic achievement. *Journal of Advanced Academics*, *21*(2), 272-298.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, *92*(6), 1087.

DuFour, R. (2002). The learning-centered principal. *Educational leadership*, *59*(8), 12-15.

DuFour, R. (2004). What is a" professional learning community"?. *Educational leadership*, *61*(8), 6-11.

Duncan, O. D. & Goldberger, A. (1973). Structural Equation Models in the Social Sciences, New York, NY: Seminar Press.

Eide, E.R., & Showalter, M.H. (2010). Human Capital. In: Brewer, D.J., & McEwan, P.J. (eds) (2010). *Economics of Education* (pp. 27-32). Amsterdam: Elsevier.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard educational review*, *66*(1), 1-27.

Epstein, J. L. (1988). How Do We Improve Programs for Parent Involvement?. Educational Horizons, 66(2), 58-59.

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

Farr, S. (2010). *Teaching as Leadership: The highly effective teacher's guide to closing the achievement gap*. San Francisco: Jossey-Bass.

FAST SCASS: Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards (2012). *Using the Formative Assessment Rubrics and Observation Protocols To Support Professional Reflection on Practice.*

Feldman, S. S., Wentzel, K. R., & Gehring, T. M. (1989). A comparison of the views of mothers, fathers, and pre-adolescents about family cohesion and power. *Journal of Family Psychology*, *3*(1), 39.

Fenstermacher, G., & Richardson, V. (2005). On making determinations of quality in teaching. *The Teachers College Record*, *107*(1), 186-213.

Feng, L. (2006) "Combating Teacher Shortages: Who Leaves, Who Moves, and Why." Electronic Theses, Treatises and Dissertations. Paper 4479

Ferguson, R.F. (2008). The tripod project framework. The Tripod Project. http://www.achievementseminars.com/seminar_series_2008_2009/readings/ferguson_tripod_project.pdf

Ferguson, R. F. (2010). Student perceptions of teaching effectiveness (Discussion brief from the National Center for Teacher Effectiveness and the Achievement Gap Initiative). Boston, MA: Harvard University.

Ferguson, R. F. (2012). Can student surveys measure teaching quality? Phi Delta Kappan, 94(3), 24–28.

Ferguson, R. F., & Danielson, C. (2014). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. in Kane, T. J., Kerr, K.A., and Pianta, R.C., eds., *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project*, 98-143. Bill and Melinda Gates Foundation. San Francisco, CA: Jossey Bass.

Ferguson, R.F. & Hirsch, E. (2014). How Working Conditions Predict Teaching Quality and Student Outcomes. in Kane, T. J., Kerr, K. A., & Pianta, R. C., eds., *Designing Teacher Evaluation Systems: New Guidance from the Measures of Effective Teaching Project.* 332 – 380. Bill and Melinda Gates Foundation. San Francisco, CA: Jossey Bass.

Follman, J. (1995). Elementary public school pupil rating of teacher effectiveness. *Child Study Journal, 25*(1), 57–78.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education*, *77*(2), 148-171.

Frey, C. B., & Osborne, M. A. (2013). The future of employment: how susceptible are jobs to computerization?. Working Paper. Retrieved Feb. 9, 2016. http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Futernick, K. (2010). Incompetent Teachers or Dysfunctional Systems?. *Phi Delta Kappan*, *92*(2), 59-64.

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

Gibbs, N. (2014). “Honor Thy Teacher.” *Time*.

Ginsburg, K. R., & Jablow, M. M. (2005). Building resilience in children and teens. *AAP Books*.

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.

Goe, L., Bell, C., & Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.

Goe, L., & Stickler, L. M. (2008). Teacher Quality and Student Achievement: Making the Most of Recent Research. TQ Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.

Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, *2*(1), 50-55.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job.* Brookings.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, *52*(11), 1182.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*(11), 1209-1217. doi: http://dx.doi.org/10.1037/0003-066X.52.11.1209

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of educational research, 66*(3), 361-396.

Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores [14th William H. Angoff Memorial Lecture]. Retrieved from http://www.nnstoy.org/download/vam-and-student-growth-models/VAM%20Angoff%20lecture%20ETS.pdf

Haggarty, L., & Pepin, B. (2002). An investigation of mathematics textbooks and their use in English, French and German classrooms: Who gets an opportunity to learn what?. *British Educational Research Journal*, *28*(4), 567-590. Stable URL: http://www.jstor.org/stable/1501441

Hanushek, E.A. (1970). The Production of Education, Teacher Quality, and Efficiency. in U.S. Office of Education, Do Teachers Make a Difference? , Washington, D.C.: Government Printing Office, pp. 79-99.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Hanushek, E.A. (2011). Valuing Teachers. *Education Next*. 41-45.

Hanushek, E.A. & Rivkin, S.G. (2006). Teacher Quality. in Hanushek E.A. & Welch, F. *Handbook of the Economics of Education, Volume 2.* 1052 – 1078. *DOI: 10.1016/S1574-0692(06)02018-6*

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). *Teachers, schools, and academic achievement* (No. w6691). National Bureau of Economic Research.

Hargreaves, A., & Fullan, M. (2013). The power of professional capital. *Journal of Staff Development*, *34*(3), 36-39.

Harris, A. (2002). Effective leadership in schools facing challenging contexts. School Leadership & Management, 22(1), 15-26.

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

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Haynes, N. M., Emmons, C., & Ben-Avie, M. (1997). School climate as a factor in student adjustment and achievement. *Journal of educational and psychological consultation*, *8*(3), 321-329.

Heller, R., Calderon, S., & Medrich, E. (2003). Academic Achievement in the Middle Grades: What Does Research Tell Us? A Review of the Literature.

Hiebert, J. (2013). The constantly underestimated challenge of improving mathematics instruction. In *Vital directions for mathematics education research* (pp. 45-56). Springer New York.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning*, *1*, 371-404.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal*, *42*(2), 371-406.

Hirsch, E., & Emerick, S. (2007). Teacher Working Conditions Are Student Learning Conditions: A Report on the 2006 North Carolina Teacher Working Conditions Survey. Center for Teaching Quality.

Hoetker, J., & Ahlbrand Jr, W. P. (1969). The persistence of the recitation. *American educational research journal*, *6*(2), 145-167.

Hofer, M., Kuhnle, C., Kilian, B., & Fries, S. (2012). Cognitive ability and personality variables as predictors of school grades and test scores in adolescents. *Learning and Instruction*, *22*(5), 368-375.

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, *47*(1), 181-217.

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.

Illinois Administrative Code (2013). Illinois Professional Teaching Standards. http://www.isbe.state.il.us/PEAC/pdf/IL_prof_teaching_stds.pdf

Ingersoll, R. (2003). Is there really a teacher shortage?

Ingvarson, L., & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education*, *52*(1), 5-35.

Interstate Teacher Assessment and Support Consortium (InTASC) (2013). Model Core Teaching Standards and Learning Progressions for Teachers 1.0. http://www.ccsso.org/Documents/2013/2013_INTASC_Learning_Progressions_for_Teachers.pdf

Iowa Core (2010). K-12 21st Century Skills: Essential Concepts and Skills with Details and Example. https://iowacore.gov/iowa-core/subject/21st-century-skills

Jacob, B. & Lefgren, L. (2004). The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago. *The Journal of Human Resources*, 39(1), 50-79. Stable URL: http://www.jstor.org/stable/3559005

Jacob, B. A., Lefgren, L., & Sims, D. (2008). *The persistence of teacher-induced learning gains* (No. w14065). National Bureau of Economic Research.

Jackson, C. K., & Bruegmann, E. (2009). *Teaching students and teaching each other: The importance of peer learning for teachers* (No. w15202). National Bureau of Economic Research.

Johnson, S. M. (2006). The Workplace Matters: Teacher Quality, Retention, and Effectiveness. Working Paper. *National Education Association Research Department*.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (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.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests.*Psychometrika*, *36*(2), 109-133.

Kane, T.J., & Cantrell, S. (2010). Learning about teaching: Initial findings from the measures of effective teaching project. Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. (2013). Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Seattle, WA: Bill & Melinda Gates Foundation.

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

Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). Identifying effective classroom practices using student achievement data (NBER Working Paper 15803). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w15803.pdf

Kaplan, D., & Elliott, P. R. (1997a). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(1), 1–24.

Kaplan, D., & Elliott, P. R. (1997b). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics*, *22*(3), 323-347. http://jeb.sagepub.com/content/22/3/323.full.pdf

Keller, B. (2007). Coaching Teachers to Help Students Learn. *Education week*,*27*(15), 22-24.

Kennedy, M. M. (1999a). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis*, *21*(4), 345–363.

Kennedy, M. M. (1999b). Form and Substance in Mathematics and Science Professional Development. *NISE brief*, *3*(2), n2.

Kersting, N.B., Givvin, K.B., Thompson, B.J., Santagata, R. and Stigler, J.W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568-589. http://aer.sagepub.com/content/49/3/568 DOI: 10.3102/0002831212437853

Kimball, S.M. (2002). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. Journal of Personnel Evaluation in Education 16(4), 241-268.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418-444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate behavioral research*, *36*(2), 249-277.

Kraft, M. A., Papay, J. P., Charner-Laird, M., Johnson, S. M., Ng, M., & Reinhorn, S. K. (2012). *Committed to their students but in need of support: How school context influences teacher turnover in high-poverty, urban schools*. Working paper.

Kraft, M.A. & Papay, J.P. (2014). Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. *Educational Evaluation and Policy Analysis*. *36*(4):476-500. http://scholar.harvard.edu/mkraft/publications/can-professional-environments-schools-promote-teacher-development-explaining

Krashen, S. D. (1981). Bilingual education and second language acquisition theory. *Schooling and language minority students: A theoretical framework*, 51-79.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Kuhfeld, M. (under review). "A validity analysis of the Tripod survey". *Educational Assessment.*

Kyriakides, L. (2005). Drawing from teacher effectiveness research and research into teacher interpersonal behaviour to establish a teacher evaluation system: A study on the use of student ratings to evaluate teacher behaviour. *Journal of Classroom Interaction, 40*(2), 44–66.

Ladd, H. (2009). *Teachers' perceptions of their working conditions: How predictive of policy relevant outcomes?* CALDER Working Paper 33. Washington, DC: National Center for

Analysis of Longitudinal Data in Education. http://www.urban.org/uploadedpdf/1001440-Teachers-Perceptions.pdf

Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis, 33*(2), 235-261.

Ladd, H. F., & Sorensen, L. C. (2014). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. *Center for Analysis of Longitudinal Data in Education Research Working Paper, 112*.

Lampert, M. (1985). How do teachers manage to teach? Perspectives on problems in practice. *Harvard Educational Review, 55*(2), 178-195.

Lavy, V., Silva, O., & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics, 30*(2), 367-414.

Leana, C.R. (2011, Fall). The missing link in school reform. *Stanford Social Innovation Review, 9*(4), 34.

Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science, 17*(3), 353-366.

Lee, V. E., & Smith, J. B. (1999). Social support and achievement for young adolescents in Chicago: The role of school academic press. *American Educational Research Journal, 36*(4), 907-945.

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of urban Economics, 56*(2), 169-191.

Little, J. W. (1990). Teachers as colleagues. in *Schools as Collaborative Cultures: Creating the Future Now (School Development and the management of change).* Lieberman, A., ed.

pp. 165 – 194. Bristol, PA: Falmer Press.
http://files.eric.ed.gov/fulltext/ED333064.pdf#page=170

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational evaluation and policy analysis*, *15*(2), 129-151.

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

Lortie, D. C. (1975). *School teacher: A sociological inquiry*. Chicago: University of Chicago Press.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American journal of education*, *106*(4), 532-575.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods,13*(3), 203.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological methods, 16*(4), 444.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, *34*(2), 120-131.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education*, *12*(1), 73-84.

MacKinnon, D. P. (2008). Introduction to statistical mediation analysis. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *Psychological Methods, 7*, 83–104.

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

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational evaluation and policy analysis*, *17*(3), 305-322.

McNeal, R. B. (1999). Parental involvement as social capital: Differential effectiveness on science achievement, truancy, and dropping out. *Social forces*, *78*(1), 117-144.

Marsh, H.W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009) Doubly Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error, *Multivariate Behavioral Research, 44*:6, 764-802, DOI: 10.1080/00273170903333665

Martínez, J. F., Borko, H., & Stecher, B. M. (2012). Measuring instructional practice in science using classroom artifacts: Lessons learned from two validation studies. *Journal of Research in Science Teaching*, *49*(1), 38-67.

Medley, D.M. & Mitzel, H.E. (1963). Measuring Classroom Behaviour by Systematic Observation. In Gage (Ed.), *Handbook of Research on Teaching*, Chicago: Rand McNally.

Meyer, P. (2011). The middle school mess. *Education Next. 11*(1).

Mihaly, K., McCaffrey, D. F., Staiger, D. O., & Lockwood, J. R. (2013). A Composite Estimator of Effective Teaching. MET Project. Bill and Melinda Gates Foundation.

Milanowski, A. T. (2011). Validity Research on Teacher Evaluation Systems Based on the Framework for Teaching. Paper presented at American Education Research Association annual meeting on April 10, 2011 in New Orleans, LA. http://files.eric.ed.gov/fulltext/ED520519.pdf

Ministry of Education (2004). "Marco para la Buena Enseñanza." Santiago, Chile: Ministerio de Educación.

Morin, A. J. S., Marsh, H. W., Nagengast, B., & Scalas, L. F. (2014). Doubly Latent Multilevel Analyses of Classroom Climate: An Illustration. *The Journal of Experimental Education*, *82*(2), 143–167. doi:10.1080/00220973.2013.769412

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research & Evaluation*, *12*(1), 53–74.

Murnane, R., & Cohen, D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard educational review*, *56*(1), 1-18.

Murnane, R., Singer, J. D., Willett, J. B., Kemple, J., & Olsen, R. (2009). *Who will teach?: Policies that matter*. Cambridge, MA: Harvard University Press.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376-398.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*(1), 81-117.

Muthén, B. O., du Toit, S.H.C. & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Conditionallly accepted for publication in *Psychometrika*.

Muthén, L.K. and Muthén, B.O. (1998-2015). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén

National Board for Professional Teaching Standards (2012). *Middle Childhood Generalist Standards, Third Edition.* Arlington, VA: National Board for Professional Teaching Standards. http://www.nbpts.org/national-board-standards; http://boardcertifiedteachers.org/sites/default/files/MC-GEN.pdf

National Council on Teacher Quality. (2012). *State of the states 2012: Teacher effectiveness policies.* Washington, DC: National Council on Teacher Quality.

New Teacher Center (2014). North Carolina Teacher Working Conditions Survey: Design, Reliability, & Validity, Research Brief. http://ncteachingconditions.org/uploads/File/NC%20val%20rel%20brief%20%205-14.pdf

Noddings, N. (1988). An ethic of caring and its implications for instructional arrange- ments. *American Journal of Education, 96*, 215-231.

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

Oakes, J. (1986). Educational indicators: A guide for policymakers. New Brunswick, NJ: Rutgers University, Center for Policy Research in Education. http://files.eric.ed.gov/fulltext/ED315920.pdf

Oakes, J. (2005). *Keeping track*. Yale University Press.

OECD (2009). Creating Effective Teaching and Learning Environments: First results from TALIS. Paris, France: OECD Publishing.

OECD (2013). Teachers for the 21st Century: Using Evaluation to Improve Teaching. Paris, France: OECD Publishing.

Papay, J.P., & Kraft, M.A. (2012). Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career growth. Harvard Graduate School of Education Working Paper.

Partnership for 21st Century Skills. (2007). Framework for 21st Century Learning. www.21stcenturyskills.org

Patrick, J., & Smart, R. M. (1998). An empirical evaluation of teacher effectiveness: The emergence of three critical factors. *Assessment & Evaluation in Higher Education*, *23*(2), 165-178.

Peña, D. C. (2000). Parent involvement: Influencing factors and implications. *The Journal of Educational Research*, *94*(1), 42-54.

Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*(2), 135–153.

Piaget, J. (1952). *The origins of intelligence in children* (Vol. 8, No. 5, p. 18). New York: International Universities Press.

Pianta, R.C., Hamre, B.K., Haynes, N.J., Mintz, S.L. & LaParo. K.M. (2006). *Classroom Assessment Scoring System - CLASS: Manual* Middle/Secondary Version Pilot. Charlottesville, VA: University of Virginia.

Pianta, R. C., La Paro, K. M., Hamre, B. K. *(*2007*). Classroom Assessment Scoring System—CLASS.* Baltimore, Maryland: Brookes Publishing.

Pianta, R.C, LaParo, K.M., & Hamre, B.K. (2008). *Classroom Assessment Scoring System (CLASS) Manual K-3.* Baltimore, Maryland: Brookes Publishing.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). Reformed teaching observation protocol (RTOP) reference manual. *Tempe, Arizona: Arizona Collaborative for Excellence in the Preparation of Teachers*.

Phillips, M. (1997). What makes schools effective: A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance in middle school. *American Educational Research Journal, 34,* 633-662.

Phillips, M. & Chin, T. 2004. "School Inequality: What Do We Know?" in K. Neckerman (Ed.) *Social Inequality* (pp. 467-510). New York: Russell Sage Foundation Press

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. doi:10.1080/10705511.2011.557329

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

Protheroe, N. (2002). Improving Instruction through Teacher Observation. *Principal*, *82*(1), 48-51.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of democracy*, *6*(1), 65-78.

Putnam, R. (2001). Social capital: Measurement and consequences. *Canadian Journal of Policy Research*, *2*(1), 41-51.

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

Raudenbush, S. W., & Jean, M. (2014). To what extent do student perceptions of classroom quality predict teacher value added. *Designing teacher evaluation systems*, 170-202.

Raudenbush, S. W., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness*, *1*(2), 138-154.

Reform Support Network (2012) Evaluations of Teacher Effectiveness: State Requirements for Classroom Observations. US Department of Education, https://rtt.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=2421

Rich, M. (2015). Teacher Shortages Spur a Nationwide Hiring Scramble (Credentials Optional). *The New York Times*. New York, NY. http://www.nytimes.com/2015/08/10/us/teacher-shortages-spur-a-nationwide-hiring-scramble-credentials-optional.html?_r=0

Richards, J.C. (1991). Towards Reflective Teaching. *The Teacher Trainer*. 5(3), 4-8.

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

Rockoff, J. E., & Lockwood, B. B. (2010). Stuck in the middle: Impacts of grade configuration in public schools. *Journal of Public Economics*, *94*(11), 1051-1061.

Rothstein, B., & Stolle, D. (2008). The state and social capital: An institutional theory of generalized trust. *Comparative politics*, 441-459. Stable URL: http://www.jstor.org/stable/20434095

Rothstein, R. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington, DC: Economic Policy Institute.

Rothstein, J., & Mathis, W. J. (2013). Review of two culminating reports from the MET project. National Education Policy Center. University of Colorado.

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher*, *38*(2), 120-131.

Roza, M., Hill, P. T., Sclafani, S., & Speakman, S. (2004). How within-district spending inequities help some schools to fail. *Brookings papers on education policy*, (7), 201-227.

Sadler, D. R. (1989). "Formative Assessment and the Design of Instructional Systems," *Instructional Science, vol. 18*, p. 119 - 144.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education*, *11*(1), 57-67.

Schenke, K., Nguyen, T., Watts, T. W., Sarama, J. S., & Clements, D. H. (2017). Differential effects on African American and non-African American's Mathematics Achievement. *Journal of Educational Psychology.* March 23, 2017, online first publication. http://dx.doi.org/10.1037/edu0000165

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304-321.

Schneider, M. (2003). Linking School Facility Conditions to Teacher Satisfaction and Success. National Clearinghouse for Educational Facilities. http://files.eric.ed.gov/fulltext/ED480552.pdf

Schochet, Peter Z. and Hanley S. Chiang (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. http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf

Schofield, L. S., Junker, B., Taylor, L. J., & Black, D. A. (2013). *Predictive inference using latent variables with covariates*. Submitted. Retrieved from http://cep.lse.ac.uk/seminarpapers/23-05-14-DB.pdf

Schuller, T., Bynner, J., and Feinstein, L. (2004). Capitals and capabilities, Centre for Research on the Wider Benefits of Learning, Institute of Education, University College London.

Schultz, T. (1961). Investment in human capital. *American Economic Review*, 51(1), 1-17.

Schweig, J. (2014). Cross-Level Measurement Invariance in School and Classroom Environment Surveys Implications for Policy and Practice. *Educational Evaluation and Policy Analysis*, *36*(3), 259-280.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, *85*(4), 571.

Spillane, J. P. (2005, June). Distributed leadership. *The Educational Forum*, *69*(2). 143-150.

Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D. F., ... & Stecher, B. M. (2011). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT). *Society for Research on Educational Effectiveness*.

Stallings, J. (1977). *Learning to look: A handbook on classroom observation and teaching models*. Belmont, CA: Wadsworth Publishing Company.

Strayhorn, T. L. (2014). What role does grit play in the academic success of Black male collegians at predominantly White institutions? *Journal of African American Studies*, *18*(1), 1-10.

Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of education*, 126-141.

Swanlund, A. (2011). Identifying working conditions that enhance teacher effectiveness: The psychometric evaluation of the Teacher Working Conditions Survey. Chicago. IL: American Institutes for Research.

Sztompka, P. (1998) Trust, Distrust, and Two Paradoxes of Democracy. European Journal of Social Theory 1(1): 19-32.

Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.

Taylor, E.S., & Tyler, J.H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-51.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research 83*(3), 357-385.

Thomas, D. S. (1929). *Some new techniques for studying social behavior* (Child Development Monographs, No. 1). New York, NY: Teachers College, Columbia University.

Thomas, E. (2010). Why we must fire bad teachers. *Newsweek.* http://www.newsweek.com/why-we-must-fire-bad-teachers-69467

Tripod Project. (2016). Become a Tripod School. http://tripoded.com/school-leaders/

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*(1), 55-73.

United Teachers Los Angeles (2012). Teacher Development and Evaluation Framework. Work in Progress, submitted to the Board of Directors, March 21, 2012.

U.S. Department of Education (2002). No Child Left Behind Act. Public Law 107–110—Jan. 8, 2002. http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf ; http://www2.ed.gov/policy/elsec/leg/esea02/index.html

U.S. Department of Education (2009). Race to the Top Fund Executive Summary. American Recovery and Reinvestment Act. http://www2.ed.gov/programs/racetothetop/executive-summary.pdf; http://www2.ed.gov/policy/gen/leg/recovery/statutory/stabilization-fund.pdf

U.S. Department of Education (2014). Improving Teacher Preparation: Building on Innovation. http://www.ed.gov/teacherprep; http://www.ed.gov/news/press-releases/us-department-education-proposes-plan-strengthen-teacher-preparation

U.S. Department of Education (2013). National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Teacher Data File," 1987-88 through 2011-12; "Private School Teacher Data File," 1987-88 through 2011-12; and "Charter School Teacher Data File," 1999-2000. (This table was prepared July 2013.) http://nces.ed.gov/programs/digest/d13/tables/dt13_209.10.asp?referrer=report

Vergara v. California (2014). Superior Court of the State of California, County of Los Angeles. http://studentsmatter.org/wp-content/uploads/2014/06/Tenative-Decision.pdf

Vygotsky, L. S. (1979). Consciousness as a problem in the psychology of behaviour. *Soviet Psychology, 17*(4), 3–35.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What Can Student Perception Surveys Tell Us About Teaching? Empirically Testing the Underlying Structure of the Tripod Student Perception Survey. *American Educational Research Journal*, *53*(6), 1-35. DOI: 10.3102/0002831216671864

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*.

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of educational psychology*, *89*(3), 411.

Williams, T., Kirst, M., Haertel, E., Rosin, M., Perry, M., Webman, B., & Woodward, K. M. (2010). Gaining ground in the middle grades: Why some schools do better. *Mountain View, CA: EdSource*.

Worrell, F. C., & Kuterbach, L. D. (2001). The use of student ratings of teacher behaviors with academically talented high school students. *Journal of Secondary Gifted Education*, *12*(4), 236-247.

Yecke, C. P. (2005). Mayhem in the Middle. *Washington, DC: Fordham Institute*.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models problems and solutions. *Organizational Research Methods*, *12*(4), 695-719. http://orm.sagepub.com/content/12/4/695.full.pdf