UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Understanding Mathematics Classroom Instruction Through Students and Teachers

Permalink

https://escholarship.org/uc/item/45g815wc

Author

Schenke, Katerina

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Understanding Mathematics Classroom Instruction Through Students and Teachers

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Katerina Schenke

Dissertation Committee:
Distinguished Professor Jacquelynne S. Eccles, Chair
Associate Professor Stephanie Reich
Professor John R. Hipp

2015

© 2015 Katerina Schenke

DEDICATION

To:

My father Wolfgang

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	v
ACKNOWLEDGEMENTS	viii
CURRICULUM VITAE	ix
ABSTRACT OF THE DISSERTATION	xvi
CHAPTER 1: Introduction	1
CHAPTER 2: Describing profiles of teachers' instructional practices	12
CHAPTER 3: Understanding effective instruction: What do we observe? When do we observe? How do we set up our analyses?	45
CHAPTER 4: Mediating the association between instructional practices and student achievement: the role of student perceptions	95
CHAPTER 5: Key Findings and Future Research	146

LIST OF FIGURES

Figure 2.1	CLASS profiles for video 1	43
Figure 2.2	MQI profiles for video 1	44
Figure 4.1	Multilevel mediation model	134
Figure 4.2	Mediation model with middle school students' perceptions of interest (3,190 students in 190 classrooms)	137
Figure 4.3	Mediation model with middle school students' perceptions of emotional support (3,190 students in 190 classrooms)	138
Figure 4.4	Mediation model with middle school students' perceptions of instructional support (3,190 students in 190 classrooms)	139
Figure 4.5	Mediation model with middle school students' perceptions of academic press (3,190 students in 190 classrooms)	140
Figure 4.6	Mediation model with elementary school students' perceptions of emotional support (3,269 students in 322 classrooms)	141
Figure 4.7	Mediation model with elementary school students' perceptions of instructional support (3,269 students in 322 classrooms)	142
Figure 4.8	Mediation model with elementary school students' perceptions of academic press (3,269 students in 322 classrooms)	143

LIST OF TABLES

Table 2.1	Standardized factor loadings and model fit of the three dimensions of the CLASS	37
Table 2.2	Means, standard deviations, and correlations for observations of the CLASS and MQI	38
Table 2.3	Profile enumeration for CLASS	39
Table 2.4	Estimated means, proportions, and standard errors of profiles for CLASS	40
Table 2.5	Profile enumeration for MQI	41
Table 2.6	Estimated means, proportions, and standard errors of profiles for MQI	42
Table 3.1	Correlations for classroom achievement and classroom characteristics	78
Table 3.2	Associations between CLASS and achievement on state standardized test using a variable-centered approach (N=527)	79
Table 3.3	Associations between CLASS and achievement on BAM using a variable-centered approach (N=527)	80
Table 3.4	Associations between CLASS and standard deviation of classroom achievement using a variable-centered approach (N=527)	81
Table 3.5	Associations between CLASS and standard deviation of classroom BAM achievement using a variable-centered approach (N=527)	82
Table 3.6	Associations between MQI and achievement on state standardized test using a variable-centered approach (N=527)	83
Table 3.7	Associations between MQI and achievement on BAM using a variable-centered approach (N=527)	84
Table 3.8	Associations between MQI and standard deviation	85

	of classroom achievement using a variable-centered approach (N=527)	
Table 3.9	Associations between MQI and standard deviation of achievement on BAM using a variable-centered approach (N=527)	86
Table 3.10	Associations between CLASS and achievement on state standardized test using a pattern-centered approach (N=527)	87
Table 3.11	Associations between CLASS and achievement on BAM using a pattern-centered approach (N=527)	88
Table 3.12	Associations between CLASS and standard deviation of classroom achievement using a pattern-centered approach (N=527)	89
Table 3.13	Associations between CLASS and standard deviation of classroom BAM achievement using a pattern-centered approach (N=527)	90
Table 3.14	Associations between MQI and achievement on state standardized test using a pattern-centered approach (N=527)	91
Table 3.15	Associations between MQI and achievement on BAM using a pattern-centered approach (N=527)	92
Table 3.16	Associations between MQI and standard deviation of classroom achievement using a pattern-centered approach (N=527)	93
Table 3.17	Associations between MQI and standard deviation of achievement on BAM using a pattern-centered approach (N=527)	94
Table 4.1	Standardized factor loadings of student survey items for the middle school sample (n=3,190)	132
Table 4.2	Standardized factor loadings of student survey items for the elementary school sample (n=3,507)	133
Table 4.3	Means, standard deviations, and correlation for middle	135

school sample (n=3,190)

Table 4.4 Means, standard deviations, and correlation for elementary school sample (n=3,507) 136

Table 4.5 Unstandardized coefficients of variable predicting to Standardized mathematics achievement for middle school (n=3,190) 144

Table 4.6 Unstandardized coefficients of variable predicting to Standardized mathematics achievement for elementary School (n=3,507) 145

ACKNOWLEDGEMENTS

First and foremost I would like to thank my close colleagues and friends Arena Lam, Cathy Tran, Anny Jin Kyoung Hwang, Ken Lee, Teya Rutherford, and Erik Ruzek. There is no doubt that without your support, advice, and collaboration I would not have made it though this program. I am especially grateful to Arena and Erik who have proven to be my most valued collaborators. I have thoroughly enjoyed our brainstorming and believe that we have unique contributions to offer to the field. To Cathy, our collaborative writing sessions have made me become a better writer.

I would like to acknowledge and thank my mother, Danuta, for her unwavering support. Throughout my life, she has taught me never to quit and has supported me unconditionally. And to Bruce, for putting up with a new stepdaughter during this time of transition and growth. Thank you to my childhood friends Nazita Lajevardi and Liz Kim. I'm glad we could embark on this journey together.

Thank you to Tyler Watts for being my rock during both the rewarding and hectic times throughout graduate school. You are my true love and I learn so much from you everyday. Thank you to Ana Auger for introducing us and for becoming a close friend. Your unconditional support has instilled great confidence in myself.

I want to thank the many advisors I have had during my time at UCI. To Lindsey Richland, thank you for being the first to take a chance on me and to teach me in the program. I continue to look up to you as a role model. To Michael Martinez, thank you for being the most supportive advisor a student could ask for. I was devastated by your loss and hope to continue helping others as you have helped and supported me. To AnneMarie Conley, for agreeing to advise me during a difficult time. I know that Mike's loss was particularly difficult for you and I thank you for stepping up and advising me when I needed it the most. And finally, to Jacque Eccles, you really took a chance on me and mentored me during the last and most crucial stage of the Ph.D. Without you, I would not be here and I truly thank you for that.

I would like to thank my other mentors: Dan Hickey and Stuart Karabenick. At times, you supported me like I was your student and I have learned so much from you. The weekly meetings with Dan have shaped me into the scholar I am today. You have introduced me to other ways of thinking about learning and instruction that I truly appreciate. Stuart, thank you for always being there for me. I thank you for your support and for encouraging me to continue persevering during this process. We have established a beautiful collaboration that I hope will continue into the future. I would also like to thank Drew. I have been waiting for years to learn from someone with your expertise and work ethic. And although we did not meet until last year, I am excited about the work we will do in the future.

Finally, I would like to thank my dissertation committee Jacque, Stephanie, and John. It truly means a lot to me that you have agreed to serve on my committee and I appreciate all of the valuable feedback you have given me.

Katerina Schenke

CURRICULUM VITAE

School of Education
University of California, Irvine
3200 Education, Irvine, CA 92697

Email: kschenke@uci.edu
Phone: (949)500-3821
Website: katerinaschenke.com

Education

Ph.D. **Education**, specializing in Learning Cognition, and Development
University of California, Irvine, 2010-2015

B.A. **Psychology, German, minor in French and Francophone Studies**
University of California, Los Angeles, 2005-2009
Spring 2008 Universität Göttingen
Fall 2007 Paris III (Nouvelle Sorbonne)

Research Areas

Classroom climate ■ Classroom instructional practices ■ Motivation in different contexts (classrooms, e-learning environments, badge ecosystems, & digital games) ■ Cognition and instruction in STEM learning ■ Assessment ■ Help seeking

Honors & Awards

2015 UC Irvine School of Education Graduate Student Service Award
2014 AERA Division C Graduate Student Seminar participant
2013 Paul R. Pintrich Motivation in Education SIG Best Paper Award
2013 EARLI/JURE Best Paper Award

Published Journal Articles

Lam, A. C., Ruzek, E., **Schenke, K**., Conley, A. M., & Karabenick, S. (in press). Student Perceptions of Classroom Achievement Goal Structure: Is it Appropriate to Aggregate? *Journal of Educational Psychology.*

O'Byrne, W. I., **Schenke, K.,** Willis III, J. E., & Hickey, D. T. (2015). Digital badges: Recognizing, assessing, and motivating learners in and out of school contexts. *Journal of Adolescent & Adult Literacy, 58*(6), 451-454.

Schenke, K., Lam, A. C., Conley, A. M, & Karabenick, S. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology, 41,*133-146. doi:10.1016/j.cedpsych.2015.01.003

Schenke, K, Rutherford, T., & Farkas, G. (2014). Alignment of game design features and state mathematics standards: Do results reflect intentions? *Computers and Education, 76,* 215-224.
* Meets What Works Clearinghouse evidence standards

Hickey, D. T., Itow, R. C., Rehak, A., **Schenke, K.,** & Tran, C. (2013). Speaking personally—with Erin Knight. *American Journal of Distance Education*, *27*(2), 134-138.

Published Conference Proceedings

mentored undergraduate students listed in italics

Jiang, S., Williams, A., **Schenke, K.,** Warschauer, M., & O'Dowd, D. (2014, July). Predicting MOOC performance with week 1 behavior. Proceedings of the 7th International Conference Educational Data Mining, London.

Han, M., Vu, M., Bediones, C., Tran, C., & **Schenke, K**. (2014, June) How kids inform the development of a science game. Proceedings of the Games, Learning, and Society 10.0 Conference. Madison, WI.

Tran, C., **Schenke, K.,** & Hickey, D. T. (2014, June). Design Principles for motivating learning with digital badges: Consideration of contextual factors of recognition and assessment. Proceedings of the International Conference of the Learning Sciences, Boulder, CO.

Vattel, L., Higgin, T., & **Schenke, K**. (2013, June). Gaming on the ground: Assessment at the Playmaker School. Paper presented as part of a symposium entitled "Beyond Badges & Points: Gameful Assessment Systems for Engagement in Formal Education". Proceedings of the Games, Learning, and Society 9.0 Conference. Madison, WI.

Bediones, C., Macalinao, C., McDowd, B., **Schenke, K**., Tran, C., & Conley, A. M. (2013, June) Are we washing poop?: Unintended consequences in educational game design. Proceedings of the Games, Learning, and Society 9.0 Conference. Madison, WI.

Publications Under Review

mentored undergraduate students listed in italics

Richland, L. E. & **Schenke, K.** Preservice teachers' use of contrasting cases in mathematics instruction. *Under revision.*

Reimer, L., **Schenke, K.,** Nguyen, T., O'Dowd, D., Domina, T. & Warschauer, M. Evaluating promising practices in undergraduate STEM lecture courses. *Under revision.*

Schenke, K. Lam, A. C., Rutherford, T., & Bailey, D. H. Construct confounding among predictors of mathematics achievement. *Under review.*

Casasola, T. S., **Schenke, K.,** Nguyen, T., & Warschauer, M. Can flipping the classroom work? Evidence from undergraduate chemistry. *Under review.*

Manuscripts in Preparation

Schenke, K. Mediating the association between instructional practices and student achievement: The role of student perceptions.

Rutherford, T., **Schenke, K.**, Lam, A. C., Lee, D. S., Leyrer, J., Young, N. S., Tran, C., & Graham, J. D. Integrating cognitive and motivation trainings for a "Brain Boost."

Reports

Hickey, D., Itow, R., **Schenke, K.,** Tran, C., Otto, N., & Chow, C. (2013). *Badge Design Principles Documentation Project Interim Report*. Indiana University. Retrieved from http://iudpd.indiana.edu/InterimReport

Peer-reviewed Presentations

mentored undergraduate students listed in italics

Schenke, K., Tran, C., Nguyen, T., Reimer, L. & Domina, T. (2015, April). I thought I was going to get an A! Understanding the role of self-efficacy and calibration on student achievement and help seeking. Paper to be presented at the annual meeting of the American Education Research Association, Chicago, IL.

Schenke, K., Lam, A. C., Ruzek, E., Conley, A. M., & Karabenick, S. (2015, April). Heterogeneity of student perceptions of the classroom climate: A latent profile approach. Paper to be presented as part of a symposium entitled "Student Heterogeneity in the Classroom: Precursors and Links to Outcomes." Annual meeting of the American Education Research Association, Chicago, IL.

Schenke, K., Rutherford, T., Lam, A. C., & Lee, D. S. (2015, March). Working memory and self-concept as determinants of achievement in elementary mathematics: Reciprocal relations across three years. Poster to be presented at the biennial meeting of the Society for Research on Child Development, Philadelphia, PA.

Chang, A., Ruzek, E., **Schenke, K**., Conley, A. M., & Karabenick, S. (2014, April). Using multilevel confirmatory factor analysis to understand students' perceptions of the classroom goal structure. Paper presented at the annual meeting of the American Education Research Association, Philadelphia, PA.

Schenke, K., Rutherford, T., & Farkas, G. (2014, April). Linking educational technology to standardized assessments: Game content and features. Paper presented as part of a symposium entitled "Facets of large-scale evaluation illustrated with STMath: Examining outcomes, mediators and moderators." Annual meeting of the American Education Research Association, Philadelphia, PA.

Hickey, D.T. Itow, R., Rehak, A. M. **Schenke, K**., & Tran, C. (2014, April). Design principles and relevant resources for recognizing, assessing, motivating, and studying learning with digital badges. Paper presented as part of a symposium entitled "Innovating education practice through digital badges: Recent research, current practices, and future directions." Annual meeting of the American Education Research Association, Philadelphia, PA.

Ruzek, E., Downer, J. & **Schenke, K.** (2014, March). Why won't my teacher help me? Exploring variability in teachers' responsiveness to student help seeking within classrooms. Poster presented at the biennial meeting of Society for Research on Adolescence, Austin, TX.

Schenke, K., Tran, C., Young, N., & Conley, A. M. (2013, August). Down with food: The journey of developing a digital game that addresses science misconceptions. Paper presented as part of a symposium entitled "Designing effective multimedia digital games to promote learning." Annual meeting of the European Association for Research on Learning and Instruction. Munich, Germany.

Schenke, K., Tran, C., Young, N., *Bediones, C.*, *Chung, K., Dinh, C., Fernandez, K., Yam, M.,* & Conley, A.M. (2013, August). Down with food: An iPad game that addresses science misconceptions. ICT demonstration presented at the annual meeting of the European Association for Research on Learning and Instruction. Munich, Germany.

Hickey, D., Itow, R., Rehak, A., & **Schenke, K.** (2013, March). Digital Badges Design Principles Documentation Project. Poster presented at the annual meeting of the Digital Media and Learning Conference, Chicago, Il.

Schenke, K., Chang, A., Conley, A. M., & Karabenick, S. (2013, April). Perceived classroom context influences on students' help-seeking behavior: What reason and from whom. Paper presented as part of a symposium entitled "The Role of Motivation in Help-Seeking: Peers, Processes, and Classroom Perceptions." Annual meeting of the American Education Research Association, San Francisco, CA.

Rutherford, T., Lee, D. S., **Schenke, K.,** Chang, A., Tran, C., Young, N. S., Conley, A. M., Graham, J.D., Leyrer, J., & Martinez, M. E. (2013, April). Brain Boost: Randomized trial of a program to enhance intelligence in elementary and middle school. Poster presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Chang, A., **Schenke, K.,** & Conley, A. M. (2013, April). Mathematics achievement, help seeking, and classroom goal structures. Poster presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Tran, C., **Schenke, K.,** Siddiqi, F. & Conley, A. M. (2012, September). Exploring how motivation in game-based learning influences conceptual change. Paper presented to the European Association for Research on Learning and Instruction SIG Conceptual Change, Trier, Germany.

Rutherford, T., **Schenke, K**., Conley, A. M., & Martinez, M. E. (2012, August) The association between math test scores, math expectancy, and cognitive abilities. Poster presented to the annual meeting of the American Psychological Association, Orlando, FL.

Schenke, K., Chang, A., Rutherford, T., Lee, D. S., Tran, C., Young, N. S., Leyrer, J., Graham, J., Martinez, M. E. (2012, August). Brain Boost: A model for enhancing cognitive ability in an after school environment. Poster presented at the annual meeting of the American Psychological Association, Orlando, FL.

Richland, L., & **Schenke, K.** (2012, April). Teacher epistemologies of comparing and contrasting. Paper presented as part of a symposium entitled "On the Design, Implementation, and Outcomes of Using Contrasts in Learning." Annual meeting of the American Education Research Association, Vancouver, Canada.

Schenke, K. (2011, May) Student perspectives on the implementation of the knowing and learning in mathematics and science course at UC Irvine. Round table presentation at the annual U-Teach institute-NMSI conference, Austin, TX.

Invited Talks

Schenke, K. (2013, December). Alignment of game design features and state mathematics standards: Do results reflect intentions? MIND Institute, Irvine, CA.

Hickey, D. T., Itow, R., Tran, C., & **Schenke, K.** (2013, April). Badge Design Principles Documentation Project. Presented at the National Science Foundation special meeting on Badges-Based STEM Assessment: Current Terrain and the Road Ahead. NSF Headquarters, Washington, D.C.

Schenke, K. & Lam, M. (2012, November). Teach Aerodynamics with Aero! Presented at the Annual CUE LA Conference, Los Angeles, CA.

Schenke, K. (2012, September). Unpacking the effect of ST Math: Which skills does it affect? Presented at the Annual Principal's Kickoff Meeting, Irvine, CA.

Funding

2014- present — Co-PI
Multidisciplinary Design Grant.
Funding: University of California Irvine. ($3,000)

Research Positions

2014-present — Graduate Student Researcher
Achievement Research Lab UCI.
PI: Jacquelynne Eccles.
Funding: National Science Foundation

2012-present	Project Researcher *Documenting Design Principles from the Badges for Lifelong Learning* PI: Daniel T. Hickey. Funding: MacArthur Foundation
2012-2014	Graduate Student Researcher *Documenting Instructional Practices in STEM Lecture Courses* PI: Mark Warschauer, Co-PI Thurston Domina. Funding: National Science Foundation
2010-2014	Graduate Student Researcher *Spatial Temporal Mathematics at Scale: An Innovative and Fully Developed Paradigm to Boost Math Achievement Among All Learners* PI: George Farkas. Funding: Institute for Education Sciences

Professional Experience

Research Consultant, GameDesk Institute, 2012- 2013
Research and development of educational games, grant proposal writing, and school observations for an educational non-profit that designs games and learning environments for middle school students.

Teaching Experience

Teaching Assistant	EDUC 55, Knowing and Learning in Math and Science, School of Education, University of California, Irvine, 2010. Professor: Michael E. Martinez
Teaching Assistant	EDUC 170, Child Development in Education, School of Education, University of California, Irvine, 2010. Professor: M. Jung
Teaching Assistant	EDUC 170, Cognition and Learning in Educational Settings, School of Education, University of California, Irvine, 2010. Professor: Michael E. Martinez
Guest Presentation	Stats'n'Snacks, "Introduction to Mplus" Department of Psychology and Social Behavior, University of California, Irvine, 2013
Guest Lecture	"Contextual Influences of Motivation" EDUC 185 School of Education, University of California, Irvine, 2014
Guest Lecture	"Motivation in Education" EDUC 173/Psy Beh 192T School of Education, University of California, Irvine, 2014

National and International Service

Ad Hoc Reviewer
Journal of Educational Psychology, 2015
Computers and Education, 2014, 2015
Learning and Instruction, 2013
Journal of Early Adolescence, 2013

Conference Proposal Reviewer
American Educational Research Association Annual Meeting, Division C, 2014

Featured Participant at the Southwest Consortium of Innovative Psychology in Education Biannual conference, 2013

University Service

Student Representative for the Ph.D. in Education Professional Development Workshops, 2011- 2013
Student Representative for the School of Education Ad Hoc Software Committee, 2013
Session Chair for Undergraduate Research Opportunities Program Symposium, 2013

Professional Affiliations

American Educational Research Association (AERA)
Division C, Learning and Instruction
Division D, Measurement and Research Methodology
Special Interest Group on Motivation
Special Interest Group on the Learning Sciences
European Association for Research on Learning and Instruction (EARLI)

Statistics Skills and Training

2014 Applied Latent Class Analysis and Finite Mixture Modeling, Lawrence, KS.
2013 UCLA Summer Methods Institute – Hierarchical Linear Modeling, Los Angeles, CA.
2013 UTeach Institute SEDL UTOP Training & Workshop, Austin, TX.

- Experience with the following software: Stata, SPSS, Mplus, E-prime, FlexMIRT
- Expertise in the following methods: OLS and Logistic Regression, Multilevel Modeling, Structural Equation Modeling, Item Response Theory, Latent Class/Profile Analysis

Other

Fluent in oral and written German and French

ABSTRACT OF THE DISSERTATION

Understanding Mathematics Classroom Instruction Through Students and Teachers

By

Katerina Schenke

Doctor of Philosophy in Education

University of California, Irvine, 2015

Distinguished Professor Jacquelynne S. Eccles, Chair

High quality instruction is necessary for students of all ages to develop a deep understanding of mathematics. Value-added models, a common approach used to describe teachers and classroom practices, are defined by the student standardized achievement gains teachers elicit. They may, however, fail to account for the complexity of mathematics instruction as it actually occurs in the classroom. To truly understand both a teacher's impact on his/her students and how best to improve student learning, a richer picture of the classroom experience than can be provided from test scores is required.

This dissertation considers broader conceptualizations of classroom practices to understand effective mathematics instruction. The first study aims to characterize instructional practices in terms of patterns of practices and examines the correlations among teachers' ratings of instructional practices across the four time points of observed instruction. Findings from this study indicate that observed instructional practices are highly variable during the academic year. The second study uses patterns obtained in the first study and relates them to classroom level achievement gains using both a domain-general and a domain-specific observation protocol. Findings from this study suggest that a domain-general observation protocol is better at explaining variance in classroom-level achievement gains. The third study tests whether student

perceptions of the classroom environment are a mechanism to explain the association between observed instructional practices. I find that different aspects of observed instruction related to different aspects of student perceptions and that these perceptions explained the association between observed instruction and student outcomes.

Despite the identification of certain best practices in mathematics instruction, researchers and policymakers fail to understand how several aspects of mathematics instruction work together to yield gains in student achievement. Combining instructional practices and situating them within a classroom context allows researchers, practitioners, and policymakers to gain a better understanding of the complex phenomena surrounding classroom instruction.

Chapter 1: Introduction

Introduction

In the U.S., there is growing concern about the achievement of students on measures of mathematics proficiency (National Advisory Panel for Mathematics, 2008). This concern is exacerbated by findings that students' mathematics skills at school entry are predictive of their later achievement and that grave differences in students' incoming mathematics knowledge exist (Claessens & Engel, 2013; Duncan et al., 2007). Efforts around improving mathematics instruction in the U.S. point toward the role that schools and classrooms play in helping students learn math. This task is daunting and requires detailed conceptualizations and measurement of classroom instruction. Fortunately, many individuals have a stake in this issue of improving classroom instruction. Practitioners are interested in measuring teaching to improve their practice and obtain formative feedback; policymakers are interested in finding a reliable way to measure instructional quality for administrative purposes; and educational psychologists are interested in measuring teaching as a way to understand how students learn and how teaching can be improved. Because of these various perspectives, the research on teaching is highly saturated, yet several questions still remain in terms of how to conceptualize instruction (as a collection of unique variables or a pattern of variables) and how exactly instruction influences students' behavior and achievement.

In my dissertation I examine the association between observed instructional practices and student achievement using various methods and examine a mechanism through which instructional practices could have their effect. The specific research questions answered in this dissertation are:

(1) What patterns of instructional practices emerge in mathematics classrooms?

(2) *How stable are a teacher's instructional practices over the academic year?*

(3) *Does a domain-general or domain-specific observation protocol better predict to student outcomes?*

(4) *Is there a difference in using just one videotaped observation in analyses versus an average of four videotaped observations?*

(5) *Does a person-centered approach have more predictive power than a variable-centered approach in explaining classroom-level achievement?*

(6) *To what extent do specific aspects of observed instruction differentially predict student perceptions of the classroom?*

(7) *To what extent do student perceptions of instructional practices mediate the effects of instruction on student learning and achievement?*

Background

In the following paragraphs, I will briefly review major areas of programmatic research as they pertain to the study of instruction. This is by no means an exhaustive list of how researchers have come to study teaching, but rather provides some background on where we came from and where we are. I have specifically focused on the research on observing teaching rather than on other issues such as teacher credentialing and standards. Hill, Rowan, & Ball (2005), for example, give an excellent review of how the research on teacher preparation and teacher credentials emerged to explain some of the reasons why teachers differ in their effectiveness and how these perspectives grew in response to the process-product literature. In my brief summary, I start with an introduction to the process-product literature which has paved the way for all of the later efforts in research on teaching, then I discuss the Trends in International Mathematics and Science Study (TIMSS) studies which were an important marker

in thinking not just about teaching the U.S., but also around the world, and end with a section on value-added models with the specific focus on the lead up to the data that I am using for my dissertation—the Measures of Effective Teaching study (Bill and Melinda Gates Foundation, 2013)—which was entirely shaped by individuals within the theoretical lens of investigating value-added models of teaching.

Process-product research. Process-product describes a large set of studies spanning from before the 1960's to the late1980's. The goal of process-product research was to link teacher behavior (process) with student outcomes (product). Several reviews of this literature exist (Brophy & Good, 1986; Dunkin & Biddle, 1974; Rosenshine, 1970). Within this line of work, researchers observed instruction in all different kinds of content, age groups, and different aspects of teaching beyond just teacher personality characteristics, which had been the emphasis in earlier efforts of understanding effective teaching. Researchers were quickly developing their own observation protocols for understanding instruction and these observation protocols were catalogued (Simon & Boyer, 1967, 1970). During these decades of focused research, several important findings have come out such as the work on classroom management (Brown, 1952), teacher expectations (Rosenthal & Rubin, 1978; Jussim, 1989), and student questioning (Flanders, 1970).

At the same time, several important criticisms of the process-product literature came to light. For example, the process-product literature did not seem to focus on a specific content area and rather looked at domain-general specifications of teaching (Shulman, 1986). Additionally, the process-product literature had been criticized for ignoring teacher's intentions and the context of the teacher's behavior—what came right before and right after whatever it is that the teacher did (see Gage & Needles, 1989). In spite of these criticisms, the process-product literature

provides a broad base of evidence for particular instructional practices that researchers can draw on.

TIMSS studies. Another major effort in research on teaching and learning was through cross-cultural studies called the Trends in International Mathematics and Science Study (TIMSS) from the mid 1990's that came from the perspective that teaching is a cultural activity. These studies pointed toward intentions to improve the quality of teaching in U.S. classrooms by contrasting observed instructional practices in the U.S. with those found in other countries such as Japan and Hong Kong, as these nations produced students with higher scores on international assessments of mathematics achievement. This research did not intend to link specific instructional practices to a student's test score, but rather describe instructional practices on average.

Examinations of the teaching practices in eighth grade mathematics classrooms found that, in the U.S., mathematics lessons were characterized as unchallenging and procedurally oriented (Hiebert et al., 2003). In addition, mathematics instructors spent more time teaching procedures and reviewing old material rather than conceptual knowledge (seeing connections between concepts and possessing flexible knowledge that can be applied across contexts; Rittle-Johnson, Siegler, & Alibali, 2001; Wong & Evans, 2007). Whereas this cultural approach is useful for understanding what successful teachers may do, the usefulness of this approach for teacher preparation programs in the U.S. is limited because it is not easy to "import" the practices observed in Japan back to the U.S. (Stigler & Hiebert, 1997).

Value-added models. Currently, research on teaching from the policy side as taken to calculating Value Added Models (VAMs) of teaching (Raudenbush, 2004; Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti & Miller, 2002). VAMs compute a score for every teacher denoting

the average effect that teacher has on student learning—usually measured by state standardized test scores. These are calculated from the residual error terms from multilevel models predicting to a students' test score that include student demographic characteristics, previous achievement and a host of other child and school-level characteristics. The point of this research is to most accurately compute these VAM scores and use those to make important policy decisions such as hiring, firing, or promotion of teachers.

However, the evidence suggesting VAMs are poor indicators of classroom instructional quality is beginning to accumulate. Hill and colleagues (2011) compared teacher's value-added scores with observations of their classroom instruction in mathematics and found that some teachers had high value-added scores but scored poorly on observations of mathematics instructional quality when rated by external observers. The authors present case studies of two teachers who have high value-added scores, but fail to implement best practices in mathematics. In another study, Bitler and colleagues (in prep) compared teacher effects on student achievement with their effects on students' height (a variable that cannot possibly be an effect of the teacher) and found similar effect sizes for both. Both lines of research, although extremely different, suggest that value-added models do not provide researchers and policy-makers with a full understanding of identifying effective mathematics teachers; however, I still mention this approach since it is policy-relevant to understanding classroom instruction. The Measures of Effective Teaching study (where the data for this dissertation come from) sought to understand how other measures of teacher effectiveness such as surveys of student perceptions and observations of teachers' instructional practices can be combined with VAM scores to get the most accurate prediction of a teacher's effectiveness.

Data

The Measures of Effective Teaching (MET) Data

To answer my research questions, I will use the Measures for Effective Teaching (MET) dataset, a study that collected a variety of indicators of teaching quality on a sample of fourth through ninth grade teachers (Bill & Melinda Gates Foundation, 2013). MET is a two-year study (2009 and 2010 school years) on over 2,000 teachers and 100,000 students from 317 schools in six large school districts across the United States. In the second year of the study, students were randomly assigned to teachers, which was purposefully done to allow researchers to obtain better estimates of the effects of instruction on student learning without the confound of students selecting into classrooms.

In the second year of the study, video-recorded observational data from four mathematics lessons during each school year were collected for each teacher. These videos were coded on multiple protocols of teacher quality, including the Classroom Assessment Scoring System (CLASS; Pianta, Lo Paro, & Hamre, 2008) and the Mathematical Quality of Instruction (MQI; Hill, Blunk, Charalambous, Lewis, Phelps, Sleep, & Ball, 2008). MET researchers also collected student-level information on demographic characteristics (grade, age, gender, ethnicity, gifted status, English Language learner status, eligibility for free or reduced lunch,), state standardized test scores, scores on the balanced assessment in mathematics (BAM), and surveys of student perceptions of the classroom (Tripod; Ferguson, 2010).

Overview of the studies

Study 1: Describing profiles of teachers' instructional practices. The first study takes an approach of studying classroom instruction as one that involves conceptualizing instruction as dynamic and holistic. In this study, I aim to use pattern-centered approaches to describe

instructional practices on two protocols—one domain-general and one domain-specific. In doing so, I also aim to see how stable instructional practices are during the academic year.

Study 2: Understanding effective instruction: What do we observe? When do we observe? How do we set up our analyses?

For this study, I use the patterns obtained in Study 1, and link those with classroom standardized achievement to understand whether or not a pattern-centered approach is better able to predict to changes in classroom-level outcomes than a variable-centered approach. In answering this question, I also predict to classroom performance on an alternative assessment of mathematics knowledge, as well as to the standard deviation of classroom achievement to see if instructional practices are associated with how students' scores are distributed within a classroom.

Study 3: Mediating the association between instructional practices and student achievement: The role of student perceptions. The third study examines the role of student perceptions of mathematics instruction by exploring the pathway from observed instructional practice to student outcomes and takes an individual approach to understanding student learning that suggests that not all students experience instruction in the same way. Additionally, I seek to understand if various aspects of observed instruction such as teachers encouraging their students to ask questions in class are perceived as such by students.

Significance

Although, recent calls have pushed for using other measures in addition to teacher value-added scores to understand the effects of instruction on student outcomes (American Statistical Association, 2014), doing so will not achievement the goal of studying teaching is to make recommendations to improve a teacher's practice. How we theoretically conceptualize

teaching—as a collection of unique variables or as patterns within individuals—and identifying mechanisms for *how* instruction influences student achievement are important steps towards allowing researchers to have a more narrowed focus on designing professional development opportunities. If a teacher's practice is indeed best represented by using a pattern-centered approach, then future research must look at variables that explain why a teacher exhibits a specific pattern of instructional practice. Likewise, if student perceptions mediate the association between observed instructional practices and student achievement, understanding the specific links between observed instruction and student perceptions could steer professional development programs to emphasize certain aspects of instruction. Indeed, the research on teaching is saturated and at times duplicative, however, it has not yet answered critical questions about teaching such as the stability of instructional practices.

References

American Statistical Association. (2014). ASA statement on using value-added models for educational assessment. *Alexandria, VA: Author. Retrieved from https://www. amstat. org/policy/pdfs/ASA_VAM_Statement. pdf*.

Bill & Melinda Gates Foundation. (2013). *Measures of Effective Teaching (MET) project*: *Final report*. Seattle, WA: Author.

Bitler, M., Corcoran, S.P., Domina, T., & Penner, E. (In prep). Teacher effects on student achievement and height: A cautionary tail.

Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328-375). New York: Macmillan.

Brown, E. (1952). *Managing the classroom: The teacher's part in school administration.* New York: Ronald Press.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*,*115*, 1-29.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart & Winston.

Ferguson, R. (2010, October 14). *Student perceptions of teaching effectiveness*. Discussion brief from the National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, Cambridge, MA.

Flanders, N. A. (1970). Analyzing teaching behavior. Reading, MA: Addison-Wesley.

Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *The Elementary School Journal*, 89(3), 253-300.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. NCES 2003-013. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430-511.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831.

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

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 83*(4), 518-527.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

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

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

Rittle-Johnson, B., Siegler, R., & Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346–362.

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

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectations and children's academic achievements: A literature review. *Merrill-Palmer Quarterly, 29*, 1-23.

Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research*, *40*(2), 279-300.

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

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

Simon, A., & Boyer, E. G. (1967). *Mirrors for behavior: An anthology of classroom observation instruments.* Philadelphia, PA: Research for Better Schools.

Simon, A., & Boyer, E. G. (1970). *Mirrors for behavior: An anthology of classroom observation instruments.* Philadelphia, PA: Research for Better Schools.

Stigler, J., & Hiebert, J. (1997). Understanding and Improving Classroom Mathematics Instruction: An Overview of the TIMSS Video Study. *INSIGHTS FROM TIMSS*, 52.

Wong, M., & Evans, D. (2007). Improving basic multiplication fact recall for primary school students. *Mathematics Education Research Journal*, *19*(1), 89-106.

Chapter 2: Describing profiles of teachers' instructional practices

How do we conceptualize a teacher's instructional practice? The interactions that individuals have within the classroom are complex and involve various players (the individual, the teacher, peers) and involve interactions of various types (student-teacher relationships, peer relationships, and instructional practices). Researchers who study instructional practices in mathematics classrooms give homage to the complexity of these contexts (see Brophy, 1986). Henningsen and Stein (1997), for example, suggest that orchestrating classroom activity is rooted in how classroom norms are set up, the classroom management practices exhibited by teachers, and students' personal motivation. Although these ideas of complex interactions have been rooted in the theoretical literature, empirical articles have not yet taken the approach of characterizing classrooms as composed of many different components simultaneously despite the call from researchers (see Ames, 1992; Blumenfeld, 1992; Eccles, Wigfield, & Schiefele, 1998). As such, the purpose of this paper is to use a pattern-centered approach as an alternative way to describe mathematics teachers' instructional practices. I also examine the stability of teacher's instructional practices across four time points during the school year.

Early work on teaching focused on the use of observation measures in understanding a teacher's instructional practices (Brophy & Good, 1986). This body of work, called *process-product* research, sought to identify links between specific instructional practices (or processes) and student achievement outcomes (or products). Spanning over two decades, this body of work has given rise to a long list of instructional practices that are correlated with student achievement such as classroom management, teacher questioning strategies, and active teaching (see Brophy, 1986 for a review), as well as spawned many prominent critiques (see Gage & Needles, 1989 for a review). In the subsequent decades of research on this work, however, more emphasis has been

placed on understanding the link between teacher qualifications and student achievement (Darling-Hammond, 2000) and linking teacher beliefs about teaching with student achievement (see Fang, 2000 for a review). An emphasis on observing instruction has come back into favor (Hill, Blunk, Charalambous, Lewis, Phelphs, Sleep, & Ball 2008; Palardy & Rumberger, 2008; Pianta, LaParo, & Hamre, 2007). Yet, the question still remains as to how we use these observed instructional practices in analyses of student achievement and how stable they are during the academic year.

Methods used to study the impact of instructional practices on student achievement

Those who have studied instructional practices have used a variety of measures including teacher surveys of self-reported instructional practices (e.g., Engel, Claessens, & Finch, 2013), surveys of teacher's beliefs of instructional practices (e.g., Stipek, Givvin, Salmon, & MacGyvers, 2001), and observations of instructional practices (e.g., Allen, Gregory, Mikami, Lun, Hamre, & Pianta, 2013; Hamre & Pianta, 2005; Mashburn et al., 2008). Even in studies using observations of classroom instruction, different statistical treatments have been used and the implications for teacher preparation programs vary. Palardy and Rumberger (2008), for example, observed the frequency of use of math worksheets, frequency of work on problems with calendars, and frequency of use of geometric manipulations and found marginally significant effects of the frequency of use of math worksheets on student achievement as measured by a standardized measure of mathematics achievement, significant effects of the frequency of work on problems with the calendars, and significant negative effects of frequency of use of geometric manipulations. An important feature of the design of the analysis is that all instructional practices were inputted into the analysis at the same time thereby changing the interpretation of the effects of the various instructional practices. For example, it might not be

that the frequency of use of geometric manipulations is significantly negatively associated with mathematics achievement, but rather, the frequency of use of geometric manipulations holding constant the frequency of use of math worksheets and the frequency of work on problems with calendar is negatively associated with student achievement outcomes. Unfortunately, a zero order correlation table was not provided to the reader.

There are alternative ways to understand the association between instructional practices and student achievement. One alternative is to come up with different domains of mathematics achievement that are measured or observed and then to combine these into a general measure. In the case of content knowledge of teaching mathematics, Hill, Rowan, and Ball (2005) developed a general score for a teacher's content knowledge for teaching mathematics. This composite was then used to relate to student achievement outcomes. A similar approach was used by Bourke (1986) who measured things like interactions between teachers and students, teacher probes after a question, student questions, and so on and developed a composite measure of teacher instruction.

General measures capture more of the variance in student achievement outcomes—they typically have coefficients larger in magnitude because they are more general, but do not provide specific implications for practice. Though isolating variables of instructional practices are beneficial for designing specific teacher preparation programs, these analyses leave more to be desired in terms of how these practices are contextualized. Brophy, 1986, made this point very clear "one must then reconsider the behaviors as parts of patterns occurring within particular contexts in order to interpret the identified relationships" (p.1074; Brophy, 1986). For example, when developing guidelines about how long teachers should wait after asking a question, researchers should think about the nature of the question, whether students will profit form this

additional time and whether the length of the wait time has implications for the flow of the lesson (Brophy, 1986). This makes things especially complicated and although it is not feasible to measure all aspects of the classroom—especially in large-scale studies—we can focus on a few aspects of instruction and whether they work together to yield gains in student achievement.

Variable centered approaches are limited in their ability to describe instructional practices as they occur within an individual. For example, simple correlations among instructional practices might yield moderate to high correlations among variables of interest such as teacher questioning strategies, teacher use of student responses in modifying his or her lesson, and teacher emotional support. What correlations fail to give the researcher is an understanding of the extent to which these practices vary within the individual. For example, it could be that (in Palardy and Rumberger 2008) teachers who used mathematics worksheets frequently also used problems that connected to calendars more frequently as well as other beneficial instructional practices. It may be that the teachers who relied more heavily on geometric manipulations were less likely to engage in the more beneficial instructional practices. Because a correlation table was not provided in the paper, we cannot answer this. Likewise, the implications for practice if this is the case are unclear if teachers who do one thing also do the other, than if you take this perspective of looking for unique effects of highly covarying practices. A solution to this issue is to use a pattern-centered approach to understanding the influences of teacher's instructional practices.

Though there is theoretical support for conducting pattern-centered approaches in studying instruction (e.g., Brophy, 1986), few empirical studies exist on the topic. One exception to the variable-centered approaches to understanding instruction is work by Spillane and Zeuli (1999) who qualitatively analyzed patterns of 25 teachers and identified three patterns of practice

in response to policy changes towards mathematics reform. They wanted to identify practices that were changed by the reform movement and those that were not. As part of this analysis, teachers were administered self-report measures of their instructional practices such as "*How often do you ask students to explain the reasoning behind an idea"*. These patterns varied on dimensions of how mathematical tasks were set up and the extent to which procedural knowledge was emphasized. For example, some teachers were categorized as having conceptually oriented tasks and procedurally oriented discourse with their students. Understanding instruction as occurring in patterns could help researchers understand the different but functionally equivalent paths to comparable achievement and to see if patterns of instructional practices cluster within an individual.

Pattern-centered analyses

A promising method for investigating effective instructional practices is conducting a *pattern-centered analysis*. A pattern-centered approach looks at combinations, or patterns of variables, as they exist within an individual or groups of individuals. In contrast, a variable-centered approach assumes that the relationship among all variables is the same within each individual; the population is homogenous with respect to how the relations between variables hold (Bergman, Magnusson & El-Khouri, 2006). For more information on the history and specifications of pattern-centered analyses, please refer to Masyn (2013).

Research in other areas in psychology is beginning to use pattern-centered approaches. For example, in a study looking at sports participation in adolescence and youth development, researchers found that it was the patterns of youth participation in sports that were predictive of youth functioning, not individual variables, such that the combination of the specific types of afterschool activities were related to the outcome of interest (Zarett, Fay, Li, Carrano, Phelps, &

Lerner, 2009). For example, Zarett and colleagues (2009) found that participation in sports had a negative effect on positive youth development, but had a positive effect when paired with other out of school activities such as religious programs. In another example, Conley (2012) examined patterns of motivational profiles of middle and high school students and found that there are multiple motivational profiles that emerge as adaptive for student achievement outcomes. As such, using pattern-centered approaches can provide researchers with a valuable tool to understanding multiple paths to student achievement.

There are several reasons why pattern-centered approaches are important to investigate when looking at instructional practices. First, it could allow researchers to understand the multiple paths through which instructional practices might influence student achievement. For example, it may not be that one specific practice is important for achievement, but rather that specific practice occurs in combination with each other. Second, these patterns could serve as a more practical way to give feedback to teachers because it is much easier to communicate patterns of practice to teachers than mean-level coefficients of variables. Third, the interpretation of results gleaned from pattern-centered versus variable-centered approaches may yield different implications. For example, it might not be that a specific variable has negative effects on student achievement, but rather the combinations of variables that are driving these negative effects.

Stability of instructional practices

In addition to describing a teacher's instructional practices as occurring in patterns, I will also look at the stability of teacher's instructional practices across four time points over the academic year. Surprisingly, there is very limited research on the stability of the quality of instructional practices (none to the author's knowledge). This is unfortunate because knowing about the stability of instructional practices has implications for how observations of

instructional practices are used. If teachers are stable in their instructional practices over the school year, it might not matter when an external observer comes to conduct the observation. Likewise, if instructional practices are stable across the academic school year, analyses could include just one observation when looking at the association between instructional practices and student achievement. If however, teachers vary greatly in the quality of their teaching, it might mean that teachers vary their instruction during the academic year and as such, conclusions based on results from only one observation might not be valid. It might also mean that weak effects obtained from linking instructional practices to student achievement reflect this mobility rather than the unimportance of specific instructional practices.

The Present Study

My purpose is to use pattern-centered approaches to describe mathematics instructional practices. I specifically conduct these analyses on two observation protocols: the Mathematical Quality Index (MQI: Hill, et al., 2008) and the Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008). I will describe these instructional practices using a pattern-centered approach across four time points. I will then use correlations to describe the stability of these practices. The subsequent chapter (Chapter 3) aims to link these patterns with student outcomes. Of interest in the current paper is contrasting domain-general observation protocol (which look way more like the protocols used in the process-product literature), with domain-specific protocol (which was developed by people who were interested in mathematical content knowledge for teaching).

Method

Participants

Data for this study come from year two of the Measures of Effective Teaching study. In the second year of the study, students were randomly assigned to teachers, which allows researchers to obtain better estimates of the effects of instruction on student learning without the confound of students selecting into classrooms. Additionally, teachers in year 2 had four days of videotaped instruction. A total of 527 fourth through eight grade mathematics teachers were in the current sample. Only mathematics classrooms were used in the analyses. The total sample resulted in 10,994 students, and 527 teachers, in six school districts. The number of students per classroom ranged from 7 to 34 with a mean of 21 and a standard deviation of 5. There were 152 fourth grade classrooms, 179 fifth grade classrooms, 102 sixth grade classrooms, 79 seventh classrooms, and 15 eighth grade classrooms. There were fewer eighth grade classrooms because Algebra 1 teachers were excluded from the analysis because a different observation protocol was used. Classrooms had to have valid observation data for the two protocols that were used.

Procedure

Video-recorded observational data from four mathematics lessons during each school year were collected for each teacher. These videos were coded on multiple protocols of teacher quality, including the Classroom Assessment Scoring System (CLASS; Pianta, LoParo, & Hamre, 2008) and the Mathematical Quality of Instruction (MQI; Hill et al., 2008). Only the first thirty minutes of each observed lesson were scored and four lessons for each teacher were recorded. Classroom instruction was videotaped, and then coded for both protocols. On both protocols, research assistants were trained and certified prior to conducting the coding of the videos and also completed a daily calibration assessment. Scoring on the videos began at the 00:00 time mark. Eighteen percent of the videos were double coded by observers. I randomly picked a video from those that were randomly coded using a random number generator in *Stata*

13 (StataCorp, 2012). I then ordered the videos based on the dates that they were captured. The capture dates from year 2 of the study ranged from October 2010 to May 2011. Videos were sorted based on when they were captured and ordered by year, month and day so that they could be labeled as videos 1 through 4. Because there was no consistent timeframe when the videos were recorded, I also present results from averaging scores on the dimensions from the four observations, consistent with the way other researchers use data from multiple observations (e.g., Allen et al., 2013).

Measures

Classroom Assessment Scoring System. CLASS is an observational protocol that provides a measure of the global classroom quality on the following three domains: emotional support, classroom organization, and instructional support. These three domains are hypothesized to serve as latent variables describing three elements of the classroom.

Videos coded with the CLASS were divided into three ten-minute segments where observers would watch the lesson and take notes on the specific behaviors they observe related to each of the CLASS dimensions and repeated this for the duration of the lesson. After each segment, observers would give a rating on each of the indicators of the class. Raters indicated their scores on a 7-point scale ranging from low (1-2), middle (3-5), and high (6-7). Scores for each coded segment were then averaged across the entire lesson to obtain scores representing the average rating of a particular lesson. Eleven observed indicators corresponded to each of the three broader categories of the CLASS (emotional support, instructional support, and classroom organization). For the purposes of the Latent profile analysis (LPA), I first averaged items corresponding to each dimension. I opted to use raw averages, because it would make the mean

levels of each variable interpretable in my LPA instead of trying to interpret profiles where all indicators have a mean of 0 and a standard deviation of 1.

To assess the reliability of the CLASS measure, I conducted a confirmatory factor analysis (CFA) on observed items of the CLASS and mapped items of the class onto the three dimensions (emotional support, instructional support, and classroom organization). Table 1 presents the reliabilities and fit indices for the constructs at each time point. Positive climate, negative climate (reverse coded), sensitivity, and regard for student perspective had high consistent standardized factor loadings between .41 and .86 at all four time points. Model fit was also acceptable (CFI ranged from .98 to .99; TLI ranged from .95 to .98; RMSEA ranged from .02 to .06; and SRMR was .02 at all four time points; Hu & Bentler, 1999). For the classroom organization dimension, behavior management, productivity, and instructional learning formats had high consistent standardized factor loadings on the latent variable that ranged from .44 to .97 at all time points. Model fit was also acceptable (CFI ranged from .97 to .998; TLI ranged from .91 to .995; RMSEA ranged from .04. to .018; and SRMR ranged from .01 to .03 across all time points). I am unable to report the results of the model for of classroom organization for the first video.[1] Concept development, quality of feedback, and language modeling had high consistent factor loadings on the instructional supports dimension ranging from .63 to .88 across all time points. Model fit was also acceptable with CFI ranging from .97 to .998; TLI ranged from .91 to .995; RMSEA ranged from .04 to .18; and SRMR ranged from .01 to .03. It is important to note that due to the relatively small sample size of 527 teachers, some of the fit indices obtained from the four time points may be seen as fitting poorly.

[1] I encountered a non-positive definite issue for video 1 of the classroom organization dimension. I tried a number of estimation strategies and tweaks to the model to no avail.

Mathematical Quality of Instruction. Developed by Heather Hill and colleagues (2008), the MQI focuses on observing the mathematical procedures at work as well as the "rigor and richness of the mathematics lesson" (Hill et al., 2008, p. 431) and purposefully does not include indicators of classroom climate or generic instructional practices. A shorter version of the MQI called the MQI Lite was used in the study to save on time. The MQI Lite includes six dimensions corresponding to (1) classroom work connected to mathematics (dichotomous), (2) richness of mathematics (1 to 3), (3) working with students and mathematics (1 to 3), (4) errors and imprecision (1 to 3), (5) student participation in the meaning making and reasoning (1 to 3), (6) and explicitness and thoroughness. Classroom work connected to mathematics was not included in these analyses because of low variance associated with that variable (most teachers were doing work connected to math), and explicitness and thoroughness was not included because it was only included for Algebra 1 students who were not in the current study sample. Errors and imprecision captures the extent to which teachers solved problems and defined terms incorrectly, equating two non-identical mathematical terms, or forgetting key concepts or definitions. Errors and imprecision was rated on a three-point scale such that higher values means more errors and imprecision. I decided to keep this scoring of the dimension as is to make the interpretation in the latent profile analysis more understandable. Richness of mathematics captures the extent to which teachers capture the depth of mathematics to students. This has to do with explicitly making links or connections among different representations or mathematical ideas and procedures. Working with students and mathematics captures whether teachers respond to student remarks about mathematics. Student participation in meaning-making and reasoning dimension captures the extent to which students "do" mathematics. This can be in the form of student explanations. Scoring of the MQI was done by dividing each video (30 minutes total)

into four 7.5-minute segments, which were coded separately. In addition to coding each segment, raters gave a holistic rating of the MQI, however, because I wanted to make a more direct comparison with the CLASS and the MQI, I averaged the ratings from each 7.5 minute segment across the four segments to obtain an average score of each dimension of the MQI to characterize the average quality of instruction for the lesson.

Overview of Analytic Models and Methods

I used latent profile analysis (LPA; McLachlan & Peel, 2000; Vermunt & Magidson, 2002), which is a type of mixture modeling to create my profiles. Mixture modeling is a technique that identifies homogenous subgroups out of an assumed heterogeneous distribution of individuals. In LPA, indicators (items) are related to a categorical latent variable that explain the unobserved heterogeneous groups.

The procedure by which to identify these unobserved subgroups is subjective and typically involves testing multiple models using a set of fit indices and an inspection of the interpretability of the profiles. The number of groups must be well defined and differentiated (Masyn, 2013). The most commonly used fit indices for examining model fit in a latent profile analysis are the BIC, CAIC, AWE and the Vuong-Lo-Mendell Rubin likelihood test, which compares whether or not that class specification is significantly better than a classification with one less class. In addition to evaluating the relative fit indices, I also evaluated the interpretability of each model, the percentage of participants classified in each profile, and because I had longitudinal data from four time points, I compared the fit indices of each time point. I made the decision to keep the profile enumeration constant across time points, thus yielding comparable types of clusters, so that I could test profile membership across time more cleanly than had I based my profile enumeration solely on the use of fit indices. It is also

recommended to keep the same measurement model over time for ease of interpretation (Nylund, 2007). I made the deliberate decision not to include covariates in my model specification of the profiles because it is usually better not to let the covariates define the profiles, and instead use the covariates as covariates in models where classes predict to a later outcome—as is what is done in Chapter 3 in the dissertation (Lubke & Muthén, 2005).

To investigate stability of instructional practices over the four time points, I present simple correlations among dimensions of each protocol (CLASS and MQI) at each time point.

Results

Latent Profile Analysis was conducted in Mplus 7.12 (Muthén & Muthén 1998-2015) and profile enumeration for CLASS and MQI measures were conducted separately. Means were free to vary across the classes. One of the assumptions I made in this analysis was the assumption of independence such that indicators within groups are uncorrelated with each other and the associations that occur among indicators are explained by the grouping variables or profiles (e.g., McLachlan & Peel, 2000; Muthén, 2001; Uebersax, 1999). Some procedures call for splitting the sample into random halves and running the LPAs on those random halves to see if the profiles replicate (see Conley, 2012). I did not do this because the sample was so small and I had four time points on which to run the profiles. Also, in conducting preliminary analyses on these data, I used Year 1 observations (a larger group of teachers than in year 2) to run profiles. Profiles from the preliminary analyses yielded identical cluster types as those obtained using year 2 of the data. Means, standard deviations and correlations of variables of observed instruction are in Table 2. I will highlight interesting correlations between constructs at the four time points in the results section when I talk about stability of instructional practices over time.

Profile enumeration procedure

Profile enumeration procedure was done for all four videos separately and four scores on all four videos averaged to understand if average teaching during the year is a better indicator teaching. For all analyses, if the loglikelihood could not be replicated when running the LPAs, I increased the number of random starts.

Description of Profiles

CLASS. I will start with a description of the profile enumeration of the CLASS. Table 3 shows the fit indices (BIC, CAIC, AWE and the Vuong-Lo-Mendell Rubin likelihood test), loglikelihood, and number of parameters for each profile enumeration as well as the entropy for the 3-profile model. As mentioned earlier, I opted to choose a profile solution that could be applied to all time points for parsimony's sake as well as to make the interpretation of looking at the stability of profile membership over time more interpretable. Though there are slight differences in the fit statistics of the profile number across all of the time points, the means of each construct in each profile and percent of participants who fall into each profile is roughly the same. I found that the Vuong-Lo-Mendell Rubin likelihood test favored the three-class solution in three of the four time points. Table 4 shows the means for each of the profiles and time points and Figure 1 shows a bar chart depicting these profiles.

High. Across time points, the high profile (profile 3) represented 27-41% of the teachers. It is characterized by having a high mean on each of the dimensions of the CLASS (ranging from 5.03 to 5.26 for emotional support, 5.61 to 5.77 on the classroom organization dimension, and 3.96 to 4.41 on the instructional support dimension on a 7-point scale). Means for dimensions were consistently lower at the last time point than in the previous ones.

Medium. Across all time points, the medium profile (profile 2) represented 50-55% of the teachers. It is characterized by having a medium mean on each of the dimensions of the CLASS

(ranging from 4.38 to 4.62 for emotional support, 5.12 to 5.32 on the classroom organization dimension, and 3.16 to 3.34 on the instructional support dimension on a 7-point scale).

Low. Across all time points, the low profile (profile 1) represented 4-23% of the teachers. It is characterized by having a medium mean on each of the dimensions of the CLASS (ranging from 2.99 to 3.89 for emotional support, 3.72 to 4.63 on the classroom organization dimension, and 2.14 to 2.61 on the instructional support dimension on a 7-point scale).

MQI. I will start with a description of the profile enumeration of the CLASS. Table 5 shows the fit indices (BIC, CAIC, AWE and the Lo-Mendell Rubin likelihood test), loglikelihood, and number of parameters for each profile enumeration as well as the entropy for the 3-profile model. Profile enumeration for the MQI was not as straight forward as the CLASS. As seen in Table 5, fit indices of analyses conducted at two time points favor the 4 profile solution, one the two profile solution and one the three profile solution. Though there are slight differences in the fit statistics of the profile number across all of the time points, the means of each construct in each profile and percent of participants who fall into each profile is roughly the same. I found that the Lo-Mendell Rubin likelihood test favored the three-class solution in three of the four time points. Table 6 shows the means and profile membership for each of the four time points and Figure 2 shows a bar chart depicting these profiles.

High Error and low opportunity for student interactions. Across all time points, the high error and low opportunity for student interactions (profile 1) represented 8-11% of the teachers. This profile was characterized by statistically higher means on the error and imprecision dimension than the other profiles and lower means of working with students and mathematics, richness of mathematics, and student participation in the meaning making and reasoning. Means ranged from 1.97 to 2.19 for error and imprecision, from 1.09 to 1.26 for working with students

and mathematics, 1.10 to 1.18 for richness of mathematics, and 1.11 to 1.13 for student participation in the meaning making and reasoning across the four time points on a scale from 1 to 3.

Low error and low opportunity for student interactions. Across all time points, the low error and low opportunity for student interactions (profile 2) represented 72-76% of the teachers. This profile was characterized by lower levels of errors and imprecision and low levels of working with students and mathematics, richness of mathematics, and student participation in the meaning making and reasoning. Means ranged from 1.13 to 1.17 for the errors and imprecision, 1.21 to 1.22 for working with students and mathematics, 1.21 to 1.24 for richness of mathematics, and 1.09 to 1.10 for student participation in the meaning making and reasoning.

Low error and high opportunity for student interactions. Across all time points, the low error and high opportunity for student interactions (profile 3) represented 13-20% of the teachers. This profile was characterized by lower levels of errors and imprecision (similar to that of the low error and low opportunity for student interactions profile) and higher levels of working with students and mathematics, richness of mathematics, and student participation in the meaning making and reasoning. Means ranged from 1.12 to 1.17 for errors and imprecision, 1.55 to 1.65 for working with students and mathematics, 1.49 to 1.55 for richness of mathematics, and 1.29 to 1.78 for student participation in the meaning making and reasoning.

Stability of instructional practices

Results from the correlational analysis. Table 2 shows the correlations of instructional practices for each dimension of the CLASS and MQI across four time points. Correlations for the emotional support dimension of the CLASS ranged from .30 to .39 across the four time points. Correlations of the instructional support dimension of the CLASS ranged from .29 to .39 across

all four time points. Correlations of the classroom organization dimension of the CLASS ranged from .30 to .35 across four time points. Correlations among variables in the MQI are much smaller than those found among the CLASS. For the errors and imprecision dimension, correlations ranged from .01 to .10. Correlations for the student participation in meaning making and reasoning ranged from .06 to .13. Correlations for richness of instruction ranged from .03 to .14. Finally correlations for the working with students on mathematics ranged from -.04 to .09. Correlations among the dimensions of the CLASS can be described as moderate whereas correlations among the MQI can be described as small (Cohen, 1988).

Discussion

The aim of this study was to determine whether or not patterns of instructional practices could be extracted from two protocols of mathematics instruction (the CLASS and the MQI). I found that whereas profiles obtained from the CLASS described mean level differences in variables, the profiles obtained from the MQI revealed qualitatively interesting profiles that would not have been revealed had only mean levels been investigated. Additionally, I have described the extent to which there is stability in observed instructional practices over four time points during the academic year. These correlations revealed low stability of instructional practices across four time points for both protocols.

One interesting finding from this descriptive study is the differences in profiles that emerged that can be attributed to the nature and focus of the two observation protocols. The MQI protocol focuses on observing specific instructional practices and was developed specifically for use in mathematics instruction. The principles from the MQI stem out of the reform oriented mathematics literature in which a focus on student interaction with each other and with the instructor became more apparent. The second observation protocol (CLASS) evolved out the

literature on parent emotional support from the field of developmental psychology. It gives a nod to the importance of specific instruction by containing indicators of instructional support such as analysis and problem solving, instructional dialogue, quality of feedback, and content understanding. What the MQI categorizes as student participation and meaning making in mathematics and working with students and mathematics may be captured in the emotional support dimension of the CLASS. This is certainly a longer discussion and can be investigated by looking at the correlations between a teacher's score on the two observation protocols (see Blazar, Braslow, & Charalambous, 2015 for preliminary work on this).

Descriptively, the more interesting profiles were obtained from the MQI. The profiles obtained from the CLASS are less interesting and might even indicate that teachers' instructional practices as measured by the CLASS are not heterogeneous, and that in fact, teachers don't seem to differ on the qualities as measured by the CLASS. This is also indicated in the same time correlations among indicators of the CLASS which ranged from $r = .55$ to .75. This is further evidenced by the high correlations among the three indicators of the CLASS as well as the low entropy obtained from the CLASS compared with the entropy obtained from the MQI. Because of these less interesting profiles, they may not capture anything that a linear representation of the variables would not be able to capture. And in fact, some might even argue that creating profiles of these creates discrete categories from continuous data. Because of this, I chose to investigate the stability of instructional practices over time use a simple correlation matrix.

Findings from the four time-point correlations reveal strikingly low stability of instructional practices over the school year. While part of this might be due to measurement error, future research aiming to link instructional practices to student outcomes may consider incorporating multiple observations in analyses. Additionally, future research should consider

documenting the extent to which teachers vary their instructional practices during the academic year and if these changes are in response to their students, or correspond to the content of what is being taught changing. Further evidence of the instability of instructional practices is provided by looking at the profile analysis of the MQI when observations from the four time points were averaged during the year. The standard deviation becomes significantly smaller and the high errors profile is not apparent in the three-profile enumeration. This could mean that might not be consistently making high errors but that these high errors might be associated with the particular mathematics content that is being taught.

References

Ainsworth, M. D., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: Assessed in the strange situation and at home*. Earlbaum Associates, Hillsdale: N.J.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system-secondary. *School Psychology Review*, *42*(1), 76-97.

Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J.A. Bovaird, & N.A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255-284). Mahwah, NJ: Erlbaum.

Bergman, L. R., Magnusson, D., & El-Khouri, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach.* Mahwah, NY: Erlbaum.

Bill & Melinda Gates Foundation. (2013). *Measures of Effective Teaching (MET) project*: *Final report*. Seattle, WA: Author.

Blazar, D., Braslow, D., & Charalambous, C. Y. (2015, March). *Dimensionality of upper elementary mathematics instruction: Exploring factors across two observational instruments.* Paper presented at the Society for Research on Educational Effectiveness, Washington D.C. Paper retrieved from https://www.sree.org/conferences/2015s/program/downloads/abstracts/1438_1.pdf

Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology*, *84*(3), 272-281.

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal*, *23*(4), 558-571.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, *41*(10), 1069- 1077.

Brophy, J., & Good, T. L.(1986). Teacher behavior and student achievement. In Wittrock, M.C. (Ed.) *Handbook of research on teaching* (pp. 328-375). London: Collier Macmillan Publishers.

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology*,*104*(1), 32-47.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, *8*(1), 1-44.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (series ed.) and N. Eisenberg (volume ed.), Handbook of Child Psychology, 5th ed., Col 3, New York: Wiley.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157-178.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*(1), 47-65.

Gage, N. L. & Needles, M. C. (1989). Process-product research on teaching: A review of criticisms. *Elementary School Journal, 89,* 253-300.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure?. *Child Development*, *76*(5), 949-967.

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments as developmental processes: Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of Research on Schools, Schooling and Human Development* (pp. 25-41). New York: Routledge.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*, 524-549.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. NCES 2003-013. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430-511.

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

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

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*,*10*(1), 21-39.

McLachlan, G., & Peel, D. (2000). Finite mixture models. New York: John R. Wiley & Sons.

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.) The Oxford handbook of quantitative methods in psychology (Vol. 2, pp. 551-611). New York, NY: Oxford University Press.

Murdock, T. B., & Miller, A. (2003). Teachers as sources of middle school students' motivational identity: Variable-centered and person-centered analytic approaches. *The Elementary School Journal*, 383-399.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In A. Sayers & L. Collins (Eds.), *New methods for the analysis of change.* Washington, DC: American Psychological Association.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nylund, K. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation). Retrieved from https://www.statmodel.com/download/nylunddis.pdf.

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

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System*. *Baltimore: Paul H. Brookes*.

Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research*, 40(2), 279-300.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68-78.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, *21*(1), 1-27.

StataCorp (2012). *Stata statistical software: Release 13*. College Station, TX: StataCorp LP.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*(2), 213-226.

Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: conditional independence/dependence models. *Applied Psychological Measurement*, *23*(4), 283-297.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis*, *11*, 89-106.

Zarrett, N., Fay, K., Li, Y., Carrano, J., Phelps, E., & Lerner, R. M. (2009). More than child's play: variable-and pattern-centered approaches for examining effects of sports participation on youth development. *Developmental Psychology*, *45*(2), 368-382

Table 2.1

Standardized factor loadings and model fit of the three dimensions of CLASS

	Video 1	Video 2	Video 3	Video 4
Emotional Support (α ranged from .77 to .78)				
Item	λ	λ	λ	λ
Positive climate	0.82	0.86	0.82	0.81
Negative climate	-0.41	-0.51	-0.46	-0.46
Sensitivity	0.8	0.75	0.75	0.78
Regard for student perspective	0.69	0.62	0.63	0.66
CFI	0.99	0.99	0.98	0.99
TLI	0.98	0.96	0.95	0.98
RMSEA	0.06	0.09	0.09	0.06
SRMR	0.02	0.02	0.02	0.02
Classroom Organization (α ranged from .71 to .74)				
Item	λ	λ	λ	λ
Behavior management	--	0.87	0.87	0.85
Productivity	--	0.8	0.77	0.81
Instructional learning formats	--	0.49	0.44	0.46
Instructional Support (α ranged from .85 to .87)				
Item	λ	λ	λ	λ
Content understanding	0.74	0.74	0.76	0.77
Analysis and problem solving	0.69	0.63	0.68	0.71
Quality of feedback	0.84	0.87	0.88	0.84
Instructional Dialogue	0.86	0.82	0.83	0.82
CFI	0.98	0.99	0.97	0.998
TLI	0.95	0.97	0.91	0.995
RMSEA	0.13	0.09	0.18	0.04
SRMR	0.02	0.02	0.03	0.01

Note. The models for classroom organization were just identified, therefore measures of fit were not available.

Table 2.2

Means, standard deviations, and correlations for observations of the CLASS and MQI at four time points.

	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1. T1 ES	4.64 (.63)	1.00																										
2. T1 Org	5.31 (.62)	**0.55**	1.00																									
3. T1 IS	3.52 (.75)	**0.75**	**0.55**	1.00																								
4. T1 E&I	1.24 (.36)	-0.05	0.01	-0.02	1.00																							
5. T1 SPMMR	1.19 (.29)	**0.15**	0.07	**0.19**	-0.07	1.00																						
6. T1 RI	1.25 (.34)	**0.10**	**0.15**	**0.14**	**-0.11**	**0.34**	1.00																					
7. T1 WWSM	1.25 (.32)	**0.12**	-0.01	**0.15**	**-0.14**	**0.38**	**0.27**	1.00																				
8. T2 ES	4.64 (.63)	**0.38**	**0.32**	**0.32**	-0.02	**0.15**	**0.11**	0.06	1.00																			
9. T2 Org	5.30 (.63)	**0.27**	**0.34**	**0.25**	-0.01	**0.16**	**0.13**	0.04	**0.64**	1.00																		
10. T2 IS	3.49 (.70)	**0.32**	**0.27**	**0.33**	0.01	**0.18**	**0.15**	0.08	**0.74**	**0.59**	1.00																	
11. T2 E&I	1.23 (.33)	-0.02	-0.00	-0.02	0.03	-0.02	-0.10	-0.02	-0.05	0.01	-0.03	1.00																
12. T2 SPMMR	1.18 (.28)	**0.14**	0.07	**0.17**	-0.07	**0.12**	0.04	**0.10**	**0.11**	**0.09**	**0.13**	-0.04	1.00															
13. T2 RI	1.27 (.35)	**0.11**	0.09	**0.17**	-0.01	0.01	0.08	0.07	**0.12**	**0.09**	**0.13**	**-0.09**	**0.27**	1.00														
14. T2 WWSM	1.27 (.33)	**0.10**	0.01	**0.12**	0.01	0.08	**0.09**	0.04	**0.09**	**0.09**	**0.09**	-0.03	**0.41**	**0.31**	1.00													
15. T3 ES	4.62 (.63)	**0.39**	**0.33**	**0.39**	-0.05	**0.14**	**0.10**	0.05	**0.35**	**0.30**	**0.35**	-0.05	**0.19**	**0.13**	0.07	1.00												
16. T3 Org	5.28 (.62)	**0.24**	**0.33**	**0.27**	0.02	**0.10**	**0.13**	0.00	**0.30**	**0.33**	**0.29**	-0.01	**0.12**	0.08	0.04	**0.60**	1.00											
17. T3 IS	3.51 (.77)	**0.29**	**0.28**	**0.37**	-0.00	**0.16**	**0.14**	0.00	**0.26**	**0.24**	**0.29**	-0.06	**0.18**	**0.18**	0.08	**0.72**	**0.57**	1.00										
18. T3 E&I	1.22 (.35)	-0.04	0.06	0.00	0.03	-0.01	-0.03	-0.02	**-0.10**	-0.03	-0.06	0.08	-0.01	0.07	-0.01	-0.02	-0.01	-0.03	1.00									
19. T3 SPMMR	1.22 (.31)	**0.19**	**0.09**	**0.21**	-0.01	0.06	0.06	-0.01	**0.10**	0.01	**0.14**	-0.07	**0.12**	0.01	0.07	0.19	0.12	**0.27**	-0.07	1.00								
20. T3 RI	1.26 (.34)	**0.10**	**0.10**	**0.12**	0.04	0.05	**0.14**	-0.00	**0.10**	0.06	**0.11**	-0.08	0.05	**0.09**	**0.15**	**0.15**	**0.09**	**0.21**	**-0.10**	**0.37**	1.00							
21. T3 WWSM	1.29 (.34)	**0.11**	0.05	**0.11**	-0.04	-0.03	0.01	0.01	0.07	0.05	**0.09**	-0.03	0.03	-0.01	0.03	**0.11**	0.04	**0.12**	-0.08	**0.50**	**0.30*** **	1.00						
22. T4 ES	4.64 (.64)	**0.30**	**0.24**	**0.29**	0.01	**0.10**	0.06	**0.11**	**0.34**	**0.34**	**0.33**	-0.01	**0.11**	0.05	0.04	**0.35**	**0.28**	**0.30**	-0.08	**0.13**	**0.09***	0.04	1.00					
23. T4 Org	5.29 (.64)	**0.23**	**0.30**	**0.22**	0.07	0.05	**0.10**	-0.01	0.27	0.41	0.25	-0.02	0.08	0.06	0.04	**0.28**	**0.35**	**0.24**	-0.01	0.05	0.07	0.05	**0.62*** **	1.00				
24. T4 IS	3.48 (.75)	**0.28**	**0.19**	**0.28**	0.05	**0.09**	0.04	0.07	0.27	0.25	0.27	0.01	**0.12**	0.08	0.05	**0.29**	**0.22**	**0.28**	-0.04	**0.14**	**0.10***	0.05	**0.75*** **	**0.53*** **	1.00			
25. T4 E&I	1.23 (.36)	0.00	-0.02	-0.04	0.01	0.03	0.08	**0.09**	-0.03	0.01	-0.02	**0.10**	-0.03	-0.05	-0.05	0.02	0.07	0.05	0.07	-0.00	0.02	-0.04	0.04	0.01	0.02	1.00		
26. T4 SPMMR	1.19 (.30)	0.07	0.05	**0.14**	-0.03	**0.12**	**0.09**	0.00	0.06	0.02	0.08	-0.08	**0.10**	0.06	0.03	**0.16**	**0.09**	**0.20**	-0.06	**0.13**	**0.09***	**0.09***	0.07	**0.13** **	**0.16*** **	-0.02	1.00	
27. T4 RI	1.26 (.33)	0.08	0.08	**0.15**	-0.00	0.04	0.03	0.00	0.07	**0.10**	**0.11**	-0.08	0.01	0.06	-0.06	**0.13**	**0.13**	**0.17**	0.02	0.03	0.07	0.01	**0.10***	0.04	**0.11***	-0.07	**0.26*** **	1.00
28. T4 WWSM	1.26 (.32)	0.06	0.03	0.08	-0.05	0.05	-0.06	-0.04	0.06	0.00	0.06	-0.02	0.07	0.01	0.02	**0.10**	0.02	**0.11**	0.02	**0.13**	0.08	**0.09***	**0.09***	0.08	**0.11***	**-0.13** **	**0.44*** **	**0.21*** **

Note. ** p<.05* *** p< .01* **** p<.001.* T1 is video 1 (time 1). T2 is video 2 (time 2). T3 is video 3 (time 3). T4 is video 4 (time 4). ES is emotional support, Org is classroom organization, and IS is instructional support dimensions of the CLASS. E&I is error and imprecision, SPPMR is student participation in meaning-making and reasoning, RI is richness of instruction, and WWSM is working with students on mathematics from the MQI.

Table 2.3

Profile enumeration for CLASS

Video 1 Entropy = 0.79

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
one-class	-1592.14	6	3221.89	3227.88	3277.49	
two-class	-1397.84	10	2858.36	2868.35	2951.02	<0.01
three-class	-1256.85	14	2626.51	2615.44	2731.18	<0.01
four-class	-2530.415	18	5183.65	5191.64	5340.45	0.12
five-class	-1245.11	22	2628.09	2650.10	2831.98	0.27

Video 2 Entropy = 0.82

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
one-class	-1568.51	6	3174.62	3180.62	3230.23	
two-class	-1384.82	10	2832.32	2842.31	2924.98	<0.01
three-class	-1260.43	14	2608.59	2622.60	2738.34	0.05
four-class	-1194.35	18	2501.51	2519.51	2668.32	0.16
five-class	-1166.57	22	2471.02	2493.02	2674.90	<0.01

Video 3 Entropy=0.75

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
one-class	-1612.81	6	3263.23	3269.22	3318.83	
two-class	-1399.74	10	2862.15	2872.15	2954.82	<0.01
three-class	-1325.61	14	2738.96	2752.96	2868.70	<0.01
four-class	-1289.36	18	2691.53	2709.53	2858.34	0.14
five-class	-1271.98	22	2681.84	2703.84	2885.72	0.07

Video 4 Entropy= 0.78

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
one-class	-1620.147	6	3277.9	3283.90	3333.50	
two-class	-1424.58	10	2911.84	2921.83	3004.50	0.02
three-class	-1325.47	14	2738.69	2752.68	2868.42	0.15
four-class	-1278.41	18	2669.62	2687.63	2836.44	0.11
five-class	-1247.1	22	2632.08	2654.08	2835.96	0.06

Average of videos Entropy=0.84

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
one-class	-1050.08	6	2137.76	2143.76	2193.37	
two-class	-803.06	10	1668.79	1678.79	1761.46	0.04
three-class	-633.17	14	1354.08	1368.08	1483.82	0.03
four-class	-563.98	18	1240.76	1258.77	1407.58	0.02
five-class	-531.47	22	1200.81	1222.82	1404.70	0.31

Note. LL is log likelihood, npar is number of parameters estimated.

Table 2.4

Estimated means, proportions, and standard errors of profiles for CLASS

	Profile 1 (13%)		Profile 2 (55%)		Profile 3 (32%)	
Video 1	Mean	SE	Mean	SE	Mean	SE
Emotional support	3.65	0.08	4.53	0.06	5.24	0.07
Classroom organization	4.33	0.15	5.31	0.04	5.72	0.06
Instructional support	2.45	0.08	3.34	0.07	4.28	0.09
	Profile 1 (4%)		Profile 2 (55%)		Profile 3 (41%)	
Video 2	Mean	SE	Mean	SE	Mean	SE
Emotional support	2.99	0.20	4.38	0.06	5.16	0.06
Classroom organization	3.72	0.42	5.12	0.06	5.69	0.04
Instructional support	2.14	0.14	3.16	0.06	4.07	0.08
	Profile 1 (23%)		Profile 2 (50%)		Profile 3 (27%)	
Video 3	Mean	SE	Mean	SE	Mean	SE
Emotional support	3.89	0.08	4.62	0.07	5.26	0.08
Classroom organization	4.63	0.09	5.32	0.07	5.77	0.05
Instructional support	2.59	0.08	3.44	0.10	4.41	0.10
	Profile 1 (11%)		Profile 2 (51%)		Profile 3 (38%)	
Video 4	Mean	SE	Mean	SE	Mean	SE
Emotional support	3.55	0.26	4.46	0.14	5.19	0.08
Classroom organization	4.24	0.36	5.21	0.12	5.71	0.05
Instructional support	2.42	0.14	3.20	0.15	4.16	0.13
	Profile 1 (9%)		Profile 2 (52%)		Profile 3 (40%)	
Average of videos	Mean	SE	Mean	SE	Mean	SE
Emotional support	3.75	0.13	4.49	0.05	5.03	0.04
Classroom organization	4.41	0.17	5.21	0.04	5.61	0.03
Instructional support	2.61	0.09	3.30	0.05	3.96	0.05

Note. N=527. Percentages indicate the proportion of the sample who fall into the profiles.

Table 2.5

Profile enumeration for MQI

Video 1 Entropy =0.91

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
One class	-625.56	8	1301.26	1309.26	1375.40	
Two class	-464.96	13	1011.39	1024.39	1131.87	0.05
Three class	-364.12	18	641.04	859.05	1007.86	0.03
Four class	-291.36	23	726.86	749.87	940.01	<0.01
Five class	-245.22	28	665.93	693.92	925.40	0.45

Video 2 Entropy=0.93

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
One class	-607.96	8	1266.06	1274.06	1340.20	
Two class	-412.02	13	905.51	923.81	1040.92	<0.01
Three class	-334.52	18	781.86	807.18	969.34	0.04
Four class	-295.45	23	735.05	767.41	974.61	0.25
Five class	-277.88	28	731.25	770.65	1022.89	0.26

Video 3 Entropy= 0.90

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
One class	-684.37	8	1418.87	1426.88	1493.02	
Two class	-486.03	13	1053.53	1071.83	1188.94	<0.01
Three class	-395.29	18	903.39	928.72	1090.88	0.28
Four class	-302.3	23	748.74	781.11	988.31	0.15
Five class	--	--	--	--	--	--

Video 4 Entropy=0.94

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
One class	-637.44	8	1325.02	1333.02	1399.16	
Two class	-455.6	13	992.68	1005.67	1113.15	<0.01
Three class	-348.35	18	809.51	827.51	976.32	<0.01
Four class	-289.67	23	723.48	746.49	936.63	<0.01
Five class	-240.48	28	656.45	684.44	915.92	0.31

Average of videos Entropy=0.79

Model	*LL*	*npar*	BIC	CAIC	AWE	Adj LMR-LRT p value Vuong-Lo-Mendell_Rubin
One class	632.73	8	-1215.32	-1207.32	-1141.18	
Two class	764.89	13	-1448.31	-1435.31	-1327.83	<0.01
Three class	794.638	18	-1476.47	-1458.47	-1309.66	0.02
Four class	821.398	23	-1498.65	-1475.65	-1285.50	0.07
Five class	838.69	28	-1501.89	-1473.90	-1242.42	0.20

Note. LL is log likelihood; npar is number of parameters estimated. The loglikehood for the average of videos is positive. This happens. The likelihood is the product of the density evaluated at the observations. Usually the density takes values that are smaller than one, so its logarithm will be negative. However, this is not true for every distribution.

Table 2.6

Estimated means, proportions, and standard errors of profiles for MQI

Video 1						
	Profile 1 (8%)		Profile 2 (76%)		Profile 3 (16%)	
	Mean	SE	Mean	SE	Mean	SE
Errors and imprecision	2.19	0.18	1.17	0.02	1.12	0.02
Working with students and mathematics	1.09	0.04	1.21	0.01	1.55	0.07
Richness of mathematics	1.1	0.04	1.22	0.02	1.51	0.06
Student participation in the meaning making and reasoning	1.13	0.04	1.09	0.01	1.72	0.05
Video 2						
	Profile 1 (11%)		Profile 2 (76%)		Profile 3 (13%)	
	Mean	SE	Mean	SE	Mean	SE
Errors and imprecision	1.97	0.06	1.14	0.01	1.15	0.04
Working with students and mathematics	1.26	0.05	1.21	0.01	1.64	0.06
Richness of mathematics	1.18	0.04	1.24	0.02	1.55	0.07
Student participation in the meaning making and reasoning	1.13	0.04	1.09	0.01	1.78	0.05
Video 3						
	Profile 1 (8%)		Profile 2 (72%)		Profile 3 (20%)	
	Mean	SE	Mean	SE	Mean	SE
Errors and imprecision	2.12	0.24	1.14	0.04	1.13	0.02
Working with students and mathematics	1.16	0.04	1.22	0.02	1.63	0.06
Richness of mathematics	1.15	0.05	1.21	0.02	1.5	0.05
Student participation in the meaning making and reasoning	1.11	0.04	1.09	0.02	1.73	0.04
Video 4						
	Profile 1 (11%)		Profile 2 (75%)		Profile 3 (14%)	
	Mean	SE	Mean	SE	Mean	SE
Errors and imprecision	2.05	0.07	1.13	0.01	1.17	0.03
Working with students and mathematics	1.11	0.03	1.21	0.02	1.65	0.06
Richness of mathematics	1.17	0.04	1.23	0.02	1.49	0.05
Student participation in the meaning making and reasoning	1.13	0.03	1.1	0.01	1.29	0.05
Average of videos						
	Profile 1 (11%)		Profile 2 (69%)		Profile 3 (20%)	
	Mean	SE	Mean	SE	Mean	SE
Errors and imprecision	1.59	0.05	1.19	0.02	1.16	0.02
Working with students and mathematics	1.24	0.03	1.22	0.01	1.46	0.02
Richness of mathematics	1.21	0.03	1.22	0.01	1.42	0.03
Student participation in the meaning making and reasoning	1.17	0.03	1.14	0.01	1.42	0.03

Note. N=527. Percentages indicate the proportion of the sample who fall into the profiles.

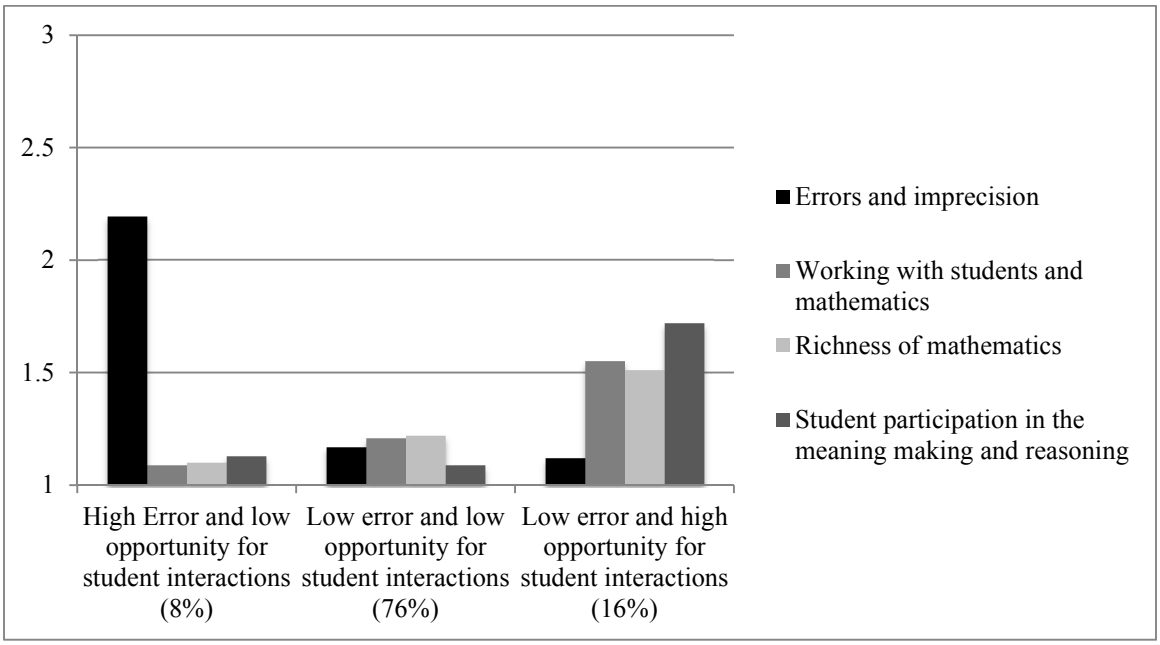


Figure 2.1. CLASS profiles for video1.

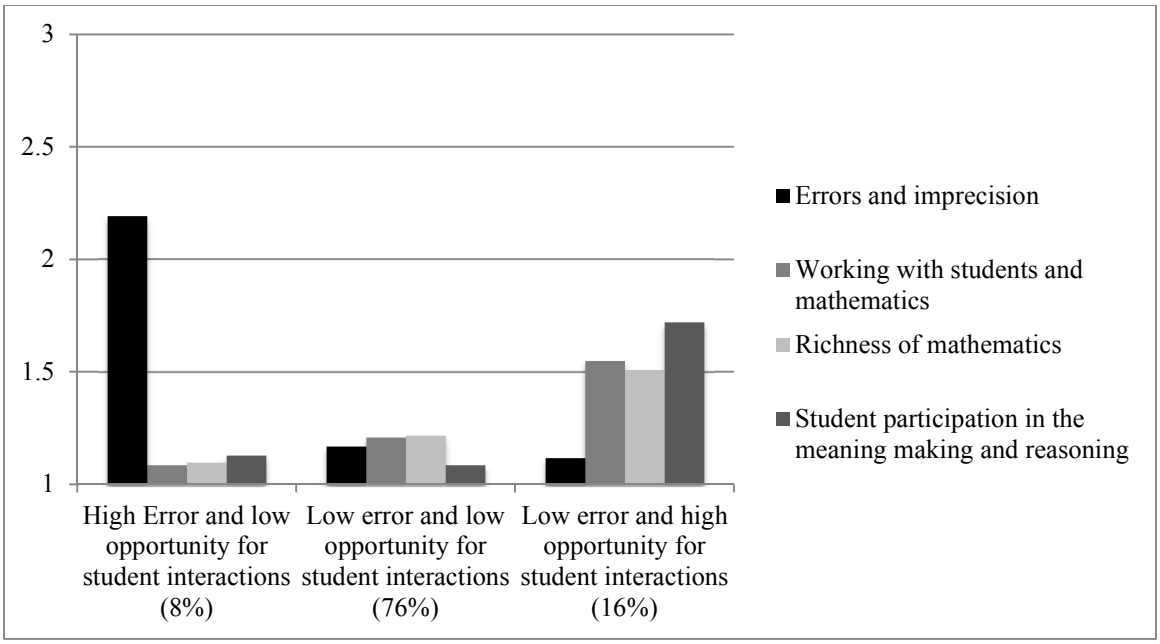


Figure 2.2. MQI profiles for video1.

Chapter 3: Understanding effective instruction: What do we observe? When do we observe? How do we set up our analyses?

Educational researchers and district personnel are challenged to develop valid and reliable measures of instructional quality. This has been an increasing concern in the age of teacher accountability because we want to know who the effective teachers are; what elements of instruction matter for student learning; and how best to prepare teachers. Several methods have been used to accomplish these goals ranging from the use of statistical models to calculate a teacher's yearly average gains on student achievement (Raudenbush, 2004; Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti & Miller, 2002), to asking teachers to report what they taught and linking mathematics content or time on mathematics content to student achievement gains (Engel, Claessens, & Finch, 2013). In the midst of these discussions, there is a re-emerging interest in developing large-scale observation protocols to identify teacher quality (see Kane, Kerr, & Pianta, 2014). The task of observing and evaluating instructional practices is not easy. It requires making numerous decisions about what to observe, how often to observe, and, once these data are gathered, which statistical approach to use. The purpose of this study is to understand differences in claims that can be made about effective instruction based on what you observe, when you observe, and how you model these instructional practices.

What do we observe?

There has been a preponderance of research around understanding and measuring instructional practices. Comprehensive reviews on the research on classroom teaching date back to the 1970's and 1980's (Brophy & Good, 1986; Moos, 1979; Porter & Brophy, 1988). Much of this research on teaching has primarily come from a domain-

general approach (Brophy, 1986). That is, practices that are not tied to specific content such as English language arts classrooms or mathematics, but rather are indicative of good teaching overall. For example, research from the process-product literature focused on aspects of the classroom such as classroom management, whether or not teachers engage in active teaching, small group instruction, and how students are questioned (Brophy, 1986; Brophy & Good, 1986). Indeed, much of this domain-general research was conducted by psychologists or educational psychologists who looked at instructional practices in a variety of grades and content domains such as third grade science students (Wright & Nuthall, 1970), early childcare environments (Stallings, Cory, Fairweather, & Needles, 1978), and junior high English classrooms (Emmer, Evertson, & Brophy, 1979). This domain-general research came from psychologists and educational psychologists who were interested in how to teach and where not interested in what to teach or the content that students should be learning (Brophy, 1986).

In the subsequent decades, researchers from mathematics education (e.g., Hiebert, et al., 2003; Hiebert & Grouws, 2007; Schoenfeld, 1988; Stipek, Givvin, Salmon, & MacGyvers, 2001) focused on understanding domain-specific instructional practice in mathematics. These scholars critiqued the process-product literature for its focus on domain-general conceptualizations of instruction (Shulman, 1986) and focused instead on understanding specific practices that underlie high quality mathematics instruction. Examples of these specific practices are the use of contrasting cases and multiple representations when teaching mathematics (e.g., Carpenter, Fennema, Franke, Empson & Levi, 1999; Rittle- Johnson & Star, 2007, 2009; Schwartz, Chase, Oppezzo, & Chin, 2011); the content of instruction (e.g., Gamoran, Porter, Smithson, & White, 1997);

teacher's mathematical knowledge for teaching (e.g., Hill, Rowen, & Ball 2005); and how teachers execute their instruction such as how teachers can facilitate productive mathematical discussion with their students (e.g., Stein, Engle, Smith, & Hughes, 2008). Research from this perspective typically relied on small-scale qualitative studies (e.g., Carpenter et al., 1996). However, the need for domain-specific observation protocols for mathematics instruction is now clear and such protocols are being developed (see the Mathematical Quality Index Hill et al., 2008; and UTeach Observation Protocol Walkington et al, 2010; the Reformed Teaching Observation Protocol Piburn & Sawada, 2000). For example Walkington and colleagues (2010) designed the UTeach observation protocol to specifically evaluate graduates of the UTeach program (a programed designed for undergraduate STEM majors who wish to earn a teaching credential along with their Bachelor's of Science degree).

Though much effort now has gone into establishing the key domain-specific aspects of teaching, there are certain challenges that arise when trying to assess them. Practical considerations such as the time and expertise it takes for district personnel to be trained on domain-specific observation protocols is much higher than if personnel are able to use one domain-general protocol to conduct observations of all classes. Likewise, the use of discipline-specific protocols has further siloed areas of research, because of the tendency to publish primarily in discipline-specific journals as is often the case when observing instruction in higher education. It is believed that discipline-specific observation protocols can both produce more specific recommendations for practitioners and show stronger links with students' mathematics achievement than domain-general observation protocols. However, it remains unclear if domain-specific protocols are

indeed better predictors of student achievement than are domain-general protocols. As such, one purpose is to test whether or not the domain-specific protocol is better at predicting to student outcomes than the domain-general protocol.

When do we observe?

When we observe is also critical to the claims we can make about the association between instructional practices and student achievement. The number of lessons or days observed varies across existing studies and how many times a researcher or district personnel observes a teacher is related to the assumptions made about the stability of instruction. For example, when a researcher observes a teacher's instruction once during the school year and links that observation with student's end of the year achievement, the researcher is operating on the assumption that a teacher's instructional practice does not vary greatly during the school year and that conducting one observation is sufficient. Indeed, many studies either explicitly or implicitly make this assumption (e.g., Allen, Gregory, Mikami, Lun, Hamre, & Pianta, 2013). In other studies instruction is observed over multiple days are averaged in order to get a more reliable measure of teaching (Hill, Kapitula & Umland 2011; Reyes, Brackett, Rivers, White, & Salovey, 2012). In these studies, researchers assume that instructional practices are relatively stable across the school year but that it is important to conduct multiple observations to eliminate measurement error. Decisions regarding how often and when to observe must take into account both the measurement issues and the stability issues. Concerns about the reliability of the measurement are now being examined using the large-scale data available through the Measures of Effective Teaching study (Bill & Melinda Gates Foundation, 2013). In contrast, little attention has been paid to the stability of

instructional practices and evidence from the previous chapter suggests that this is the case. To fill this gap, another purpose of this study is to use observations conducted at four different time points of the school year, as well as an instructional quality score averaged over those four time points, to predict to student achievement. Thus, I will be able to compare associations between observations at just one time point to scores averaged over four time points.

How do we set up our analyses?

My final goal is to understand differences in the ways in which researchers can set up their statistical analyses of instructional practices. Variable centered approaches are based on the assumption that variables have unique and additive effects on the outcome of interest whereas pattern centered approaches rely on understanding the constellation of variables within an individual that are associated with the outcome. Following Bauer and Shanahan (2007) and Murdock and Miller (2003), I will contrast pattern centered and variable centered approaches to understanding instruction. More information on pattern-centered approaches is given in chapter 2 of this dissertation.

The Present Study

My aim in the present study is three fold: (1) to understand whether or not domain-general or domain-specific protocols of instructional practices better predict to student outcomes (2) to understand differences in using just one observation or using an average of four observations, (3) to understand whether pattern-centered approaches to instruction yield different associations than variable-centered approaches. In investigating these three aims, I look at four achievement outcomes: (1) average classroom achievement on the state standardized mathematics test; (2) average classroom

achievement on the Balanced Assessment for Mathematics (BAM)—an alternative mathematics assessment to measure students' conceptual understanding of mathematics; (3) the standard deviation of average classroom achievement on the state standardized mathematics test; (4) and the standard deviation of average classroom achievement on the BAM. Mean-level classroom achievement only captures one metric of what we might think successful classrooms look like and it is also important to consider whether instruction is beneficial for decreasing the variance of student achievement in the classroom. For example, it could be that high quality instructional practices are only beneficial for already high achieving students and that these practices do not help the lowest achievers in the classroom become higher achieving. It could be argued that narrowing the variance of classroom achievement and causing low achievers to perform better is the ultimate goal of researchers and practitioners.

Method

Participants

Data for this study come from year two of the Measures of Effective Teaching study (MET; Bill & Melinda Gates Foundation, 2013). In the second year of the study, students were randomly assigned to teachers, which allows researchers to obtain better estimates of the effects of instruction on student learning without the confound of students selecting into classrooms. Additionally, teachers in year 2 had four days of videotaped instruction. A total of 527 fourth through eight grade mathematics teachers were in the current sample. Only mathematics classrooms were used in the analyses. The total sample resulted in 10,994 students, and 527 teachers, in six school districts. The number of students per classroom ranged from 7 to 34 with a mean of 21 and a standard

deviation of 5. There were 152 fourth grade classrooms, 179 fifth grade classrooms, 102 sixth grade classrooms, 79 seventh classrooms, and 15 eighth grade classrooms. I excluded Algebra 1 teachers from the analysis in this paper because a different observation protocol was used in these classrooms. As a result, I have included fewer eighth grade classrooms in my analyses. Classrooms had to have valid observation data for the two protocols that were used in the analyses.

Procedure

Video-recorded observational data from four mathematics lessons during each school year were collected for each teacher. These videos were coded on multiple protocols of teacher quality, including the Classroom Assessment Scoring System (CLASS; Pianta, LoParo, & Hamre, 2008) and the Mathematical Quality of Instruction (MQI; Hill, Blunk, Charalambous, Lewis, Phelphs, Sleep, & Ball 2008). Only the first thirty minutes of each observed lesson were scored and four lessons for each teacher were recorded. Classroom instruction was videotaped, and was subsequently coded for both protocols. On both protocols, research assistants were trained and certified prior to coding of videos; they also completed a daily calibration assessment. Scoring on the videos began at the 00:00 time mark. Eighteen percent of the videos were double coded. I randomly picked a video from those that were randomly coded using a random number generator in *Stata13* (StataCorp, 2012). I then ordered the videos based on the dates that they were captured. The capture dates from year 2 of the study ranged from October 2010 to May 2011. Videos were sorted based on when they were captured and ordered by year, month and day such so that they could be labeled as videos 1 through 4. I also present results from averaging scores on the dimensions across the four observations because this

is the aggregation method used in other studies when there are observational data from multiple observations (e.g., Allen et al., 2013).

Measures

Classroom Assessment Scoring System. CLASS is an observational protocol that provides a measure of the global classroom quality on the following three domains: emotional support, classroom organization, and instructional support. These three domains are hypothesized to serve as latent variables describing three elements of the classroom.

Videos coded with the CLASS were divided into three ten-minute segments where observers would watch the lesson and take notes on the specific behaviors they observe related to each of the CLASS dimensions and then repeat this for each 10 minute segment duration of the lesson. After each segment, observers gave a rating on each of the indicators of the class. Raters indicated their scores on a 7-point scale ranging from low (1-2), middle (3-5), and high (6-7). Scores for each coded segment were then averaged across the entire lesson. Eleven observed indicators corresponded to each of the three broader categories of the CLASS (emotional support, instructional support, and classroom organization) and averages from these eleven observed indicators were used to create scores for each of the three domains (emotional support, instructional support, and classroom organization). Results from a confirmatory factor analysis are presented in chapter 2 of this dissertation.

Mathematical Quality of Instruction. Developed by Heather Hill and Colleagues (2008), the MQI focuses on observing the mathematical procedures at work as well as the "rigor and richness of the mathematics lesson" (Hill et al., 2008, p. 431); it

purposefully does not include indicators of classroom climate or generic instructional practices and is the domain-specific protocol in this analysis. A shorter version of the MQI called the MQI Lite was used in the MET to save time. In the MQI Lite, six dimensions of instruction were observed: (1) Classroom work connected to mathematics, (2) richness of mathematics, (3) working with students and mathematics, (4) errors and imprecision, (5) student participation in the meaning making and reasoning, and (6) explicitness and thoroughness. I did not include classroom work connected to mathematics in these analyses because there was low variance on this variable and explicitness and thoroughness because this was only rated for Algebra 1 classrooms, which were excluded from my analyses. Errors and imprecision captures the extent to which teachers solved problems and defined terms incorrectly, equating two non-identical mathematical terms, or forgetting key concepts or definitions. Errors and imprecision was rated on a three-point scale such that higher values means more errors and imprecision. I decided to keep this scoring of the dimension as is to make the interpretation in the latent profile analysis more understandable. Richness of mathematics captures the extent to which teachers capture the depth of mathematics to students. This has to do with explicitly making links or connections among different representations or mathematical ideas and procedures. Working with students and mathematics captures whether teachers respond to student remarks about mathematics. Student participation in meaning-making and reasoning dimension captures the extent to which students "do" mathematics, which can be in the form of student explanations. Scoring of the MQI was done by dividing each video (30 minutes total) into four 7.5-minute segments, which

were each coded separately. I averaged the ratings from each 7.5-minute segment across the four segments to obtain an average score of each dimension of the MQI.

State standardized mathematics tests. Districts provided MET researchers with data on students' standardized test scores. Data came from six districts in six different states, where students in each state took different standardized tests as states are allowed to choose which assessment should be administered. Scores were standardized within district and within grade and district dummies were used such that scores from the students in the same district are compared to each other. Scores from students in the same classroom were averaged and used in the analyses. The intra class correlation—a measure of how much variance in the students' achievement scores can be accounted for by the classroom—for the state standardized achievement was .34. Additionally, to understand whether or not instructional practices were associated with changes in the variance or standard deviation of instruction, I created variables associated with each classroom's standard deviation of achievement.

Balanced assessment for mathematics (BAM). The BAM was administered for the fourth through eight graders in the study years. The BAM was designed to measure higher order reasoning skills and conceptual understanding. It contains 40 items that ask students to solve a series of tasks and performance on those tasks was scored based on a scoring rubric on a scale from one to eight. BAM scores that were z-scored within grade and test form (students were given on of three test forms) were provided to me and I further z-scored scores within district to match the interpretation of the state standardized tests. One percent of the sample was missing on BAM scores. The intraclass correlation for the BAM was .36. Additionally, to understand whether or not instructional practices

were associated with changes in the variance or standard deviation of instruction, I created variables for each classroom's standard deviation of achievement.

Covariates. Data from districts was gathered for the following covariates: grade level (fourth, fifth, sixth, seventh, or eighth), district, ethnicity (white, black, Hispanic, other), English Language learner status, Free or reduced lunch status, and gender (whether male or female). Free or reduced lunch status was not available for all students in one district (25% of the sample). Because I am using district level dummies, this does not affect my interpretation of this covariate. Classroom-level variables were created to signify the percentage of students in the classroom of a particular demographic characteristic. In addition, classroom size and previous student achievement (using scores on the state standardized achievement tests from the previous year) were included as covariates in the models.

Analytic approaches

To determine the association between instructional practices and classroom-level achievement, I conducted eight sets of separate analyses. I look at two different outcomes (classroom-level achievement and classroom-level standardized achievement) using two different methods (variable-centered and pattern centered), and two different observation protocols (domain-general and domain-specific). These analyses are then conducted five times, using each of the four time points and a final analysis using an average of the four time points. The analyses for the variable-centered approach with classroom-level achievement take the following general form:

$$\text{Achievement}_j = \beta_{0j} + \beta_1 V1EmotionalSupport_j + \beta_2 V1ClassroomOrganization_j + \beta_3 V1InstructionalSupport_j + \lambda_1 Covariates_j + \varepsilon_j$$

Where achievement in classroom j is a function of an overall intercept (β_{0j}), the main effect for observed emotional support for the first video ($\beta_1 V1EmotionalSupport_j$), the main effect for classroom organization for the first video ($\beta_2 V1ClassroomOrganization_j$), the main effect for instructional support for the first video ($\beta_3 V1InstructionalSupport_j$), a vector of classroom-level covariates ($\lambda_1 Covariates_j$) including previous achievement, number of students in the class, percent male, percent white ethnicity, percent black ethnicity, percent Hispanic ethnicity, percent other ethnicity, percent English Language learners, percent free and reduced lunch, grade-level, and district-level dummy variables. [2]

To determine the association between instructional practices and standard deviation of achievement, I conduct analyses using the basic following form:

$$\begin{aligned}\text{SDAchievement}_j &= \beta_{0j} + \beta_1 V1EmotionalSupport_j + \beta_2 V1ClassroomOrganization_j \\ &+ \beta_3 V1InstructionalSupport_j + \beta_4 PrevSDAchievement_j \\ &+ \beta_5 Achievement_j + \lambda_1 Covariates_j + \varepsilon_j\end{aligned}$$

Where standard deviation of achievement in classroom j is a function of an overall intercept (β_{0j}), the main effect for observed emotional support for the first observation ($\beta_1 V1EmotionalSupport_j$), the main effect for classroom organization for the first observation ($\beta_2 V1ClassroomOrganization_j$), the main effect for instructional support for the first video ($\beta_3 V1InstructionalSupport_j$), the main effect for the previous standard deviation of achievement for students in the class ($\beta_4 PrevSDAchievement_j$), the main effect of the classroom's current achievement

[2] Because these dimensions are highly correlated, I also ran the analyses separately for the three dimensions of the CLASS.

($\beta_5 Achievement_j$), a vector of classroom-level covariates ($\lambda_1 Covariates_j$) including number of students in the class, percent male, percent white ethnicity, percent black ethnicity, percent Hispanic ethnicity, percent other ethnicity, percent English Language learners, percent free and reduced lunch, grade-level, and district-level dummy variables, and an error term (ε_j).

The analyses for the pattern-centered approach were conducted in two stages. The first stage is described in the previous chapter of this dissertation where profiles of instructional practices were first identified. In the second stage (this chapter), teacher's profile membership as well as profile probability were extracted and then put into a new analysis. The analyses for the pattern-centered approach will take the following general form:

$$\text{Achievement}_j = \beta_{0j} + \beta_1 V1Profile1_j + \beta_2 V1Profile3_j + \lambda_1 Covariates_j + \varepsilon_j$$

Where achievement in classroom j is a function of an overall intercept (β_{0j}), the main effect for a teacher being in profile 1 ($\beta_1 V1Profile1_j$), the main effect for a teacher being in profile 3 ($\beta_2 V1Profile3_j$), a vector of classroom-level covariates ($\lambda_1 Covariates_j$) including previous achievement, number of students in the class, percent male, percent white ethnicity, percent black ethnicity, percent Hispanic ethnicity, percent other ethnicity, percent English Language learners, percent free and reduced lunch, grade level, and district-level dummy variables. The regressions were weighted by probability of being in that profile using the p weight command in *Stata* as in Pastor, Barron, Miller and Davis, 2007. All continuous variables were standardized.

Results

I begin by examining the descriptive statistics for the classroom-level achievement variables and covariates (descriptive statistics for the observation protocols can be found in the previous chapter) and correlations of the variables of interest. Next, I detail the results from the variable-centered approach, looking at classroom achievement, and standard deviation of classroom achievement for both protocols (CLASS and MQI). Then, I detail the results from the pattern-centered approach where I look at achievement and the standard deviation of classroom achievement for both protocols (CLASS and MQI). Finally, I include additional analyses of interest to the reader.

Table 1 provides the means and bivariate correlations between the classroom-level covariates (percent ethnicity, percent male, percent English Language learners, percent free and reduced lunch, and number of students in the classroom), the state standardized achievement test, the standard deviation of the state standardized achievement test, the BAM, and the standard deviation of the BAM. Notable correlations are between a classroom's previous achievement and their subsequent achievement (r = .88, $p < .001$), the standard deviation of classroom achievement and classroom achievement (r = -.13, $p < .01$), and between a classroom's achievement on the state standardized test and the BAM (r = .72, $p < .001$). This means that the state standardized achievement assessment and the BAM may not be measuring very different things. Also, correlations between dimensions of the CLASS were statistically significant and ranged from .66 and .84.[3] Because of these high correlations, I inspected the variance inflation factors (VIF) for those variables to assess the severity of multicollinearity. VIFs for these classroom variables were well before the cut-off value of 10 at around 2 (Kutner,

[3] Correlations for instructional practices come from averaged ratings.

Nachtsheim, & Neter, 2004) suggesting that multicolinearity among these variables was not an issue. Because many of these correlations among dimensions of the CLASS were high, I conducted additional analyses using just one dimension of the CLASS as a predictor at a time to eliminate the possibility of suppression effects (Mackinnon, Krull, & Lockwood, 2000). Correlations between dimensions of the MQI ranged from -.12 to .49 with negative correlations existing for the errors and imprecision dimension and other dimensions of the MQI and the highest correlation for students participating in mathematics and working with students on mathematics.

Variable Centered Approach

CLASS.

State standardized test. Table 2 shows the results from the CLASS predicting to the state standardized achievement. For videos 1, 2 and the average of the videos only the classroom organization dimension of the CLASS was significantly associated with classroom achievement ($\beta = .09$, $p = .002$; $\beta = .13$, $p < .001$, and $\beta = .15$, $p < .001$, respectively) controlling for the covariates. This means that a 1 standard deviation increase in classroom organization results in a .09 standard deviation change in a classroom's average achievement.

BAM. Table 3 shows the results from the CLASS predicting to the BAM. For videos 2 and the average of the videos, only the classroom organization dimension of the CLASS was significantly associated with classroom achievement ($\beta = .07$, $p = .03$; $\beta = .09$, $p = .02$, respectively).

Standard deviation of state standardized test. Table 4 shows the results from the CLASS predicting to the standard deviation of the state standardized assessment.

Significant associations between emotional support and the standard deviation of achievement (β = - .02, p = .04) and between instructional support and the standard deviation of achievement (β = .02, p = .03) were found when using observations from video 4. This suggests that the higher classrooms were rated on emotional support, the smaller the standard deviation of achievement became controlling for the previous year's standard deviation of achievement. Additionally, classrooms that were rated higher in instructional support were associated with classrooms that changed to have a larger standard deviation of achievement over the year suggesting that instructional support may either lead the higher achieving students to achieve higher or the lower achieving students to achieve lower.

Standard deviation on BAM Table 5 shows the results from the CLASS predicting to the standard deviation of the BAM assessment. There was a significant association between emotional support and the standard deviation of achievement for video 4 (β = - .02, p = .03). This suggests that the higher classrooms were rated on emotional support, the smaller the standard deviation of achievement became controlling for the previous year's standard deviation of achievement.

MQI.

State standardized test. Table 6 shows the results from the MQI predicting to the state standardized achievement. No dimensions of the MQI were significantly related to state standardized achievement.

BAM. Table 7 shows the results from the MQI predicting to the BAM. No dimensions of the MQI were significantly related to a classroom's average BAM score.

Standard deviation of state standardized test. Table 8 shows the results from the CLASS protocol predicting to the standard deviation of the state standardized test. No dimensions of the MQI were significantly related to the standard deviation of the state standardized test.

Standard deviation of the BAM. Table 9 shows the results from the CLASS protocol predicting to the standard deviation of the BAM. No dimensions of the MQI were significantly related to the standard deviation of the BAM.

Pattern-Centered Approach

CLASS.

State standardized test. Table 10 shows the results from the CLASS predicting to the state standardized achievement. For video 1 and the average of the videos, being in profile 1 was negatively associated with state standardized achievement compared with being in profile 2 ($\beta = -.20$, $p = .008$; and $\beta = -.20$, $p = .02$, respectively) controlling for the covariates. This suggests that teachers characterized in profile 1 in video 1 were associated with significantly lower classroom achievement and classrooms that were characterized as profile 2 or profile 3.

BAM. Table 11 shows the results from the CLASS predicting to the BAM. For videos 2 and the average of the videos, only being in profile 1 was significantly associated with classroom achievement ($\beta = -.17$, $p = .01$; $\beta = -.29$, $p = .002$, respectively) compared with being in profile 2. For video 2, being in profile 3 was significantly associated with achievement on the BAM ($\beta = .14$, $p = .02$), compared with being in profile 2. On average, students in profile 3 classrooms did significantly better on the BAM assessment than did students in classrooms in profile 2.

Standard deviation of state standardized test Table 12 shows the results from the CLASS predicting to the standard deviation of the state standardized test. Profiles of the CLASS were not significantly related to the standard deviation of the state standardized test.

Standard deviation of the BAM. Table 13 shows the results from the CLASS predicting to the standard deviation of the BAM. Profiles of the CLASS were not significantly related to the standard deviation of the BAM.

Classroom achievement for MQI

State standardized test. Table 14 shows the results from the MQI predicting to the state standardized achievement. Profiles of the MQI were not significantly related to state standardized achievement.

BAM. Table 15 shows the results from the MQI predicting to the BAM. Profiles of the MQI were not significantly related to achievement on the BAM.

SD of classroom achievement

State standardized test. Table 16 shows the results from the MQI predicting to the standard deviation of the state standardized achievement test. Profiles of the MQI were not significantly related to the standard deviation of the state standardized achievement test.

BAM. Table 17 shows the results from the MQI predicting to the standard deviation of the BAM. Profiles of the MQI were not significantly related to the standard deviation of the BAM.

Additional Analyses

Grade level differences. In addition to the analyses in which grade levels were treated as covariates, I conducted the analyses using the averages of videos with samples disaggregated by grade level. In the above analyses, grade levels five through eight were compared to fourth grade. Using data aggregated by grade might mask some of the grade specific differences in the association between instructional practices and achievement. I conducted analyses using the CLASS protocol predicting to classroom achievement on the state standardized test and achievement on the BAM using the average of the videos. Using the variable-centered approach, the coefficients were not statistically significantly different from each other for fourth, fifth, and sixth grade and reflect the results obtained when aggregating grades. However, I did not find a significant association between dimensions of the CLASS and state standardized achievement as the outcome for seventh grade students.[4] Using the person centered approach, the coefficient for being in profile 1 versus profile 2 was not significant for the fourth and fifth grade students, but it was significant for the sixth and seventh grade classrooms suggesting that the negative effects of being in profile 1 compared to being in profile 2 were driven by the middle school sample when looking at state standardized achievement as the outcome. However, because the number of classrooms is dramatically decreased when looking at each grade separately, it is unclear whether or not these are particularly robust grade-level differences. When looking at the BAM, I found that the negative effect of being in profile 1 versus profile 2 was driven by the fourth grade classrooms; this negative association did not exist in the other grades.

[4] I did not conduct the grade level analysis with my eighth grade sample since there are too few eighth grade classrooms.

Analyses with the CLASS. To alleviate concerns about the high correlations among the dimensions of the CLASS, I reran the variable-centered analyses using each domain of the CLASS as a predictor in separate analyses. I did this because the interpretation of my findings for the variable-centered approach could be interpreted as *only* classroom organization mattering for student achievement and not emotional support or instructional support. When I enter these dimensions into separate regressions, I found that emotional support and instructional support had significant associations with classroom achievement, however, the magnitude of those coefficients was smaller than the magnitude of the coefficient for classroom organization ($\beta = .07$, $p = .01$, $\beta = .06$, $p = .03$, $\beta = .07$, $p = .01$, $\beta = .12$, $p < .001$, for emotional support, instructional support, and classroom organization, respectively).

In the previous chapter, I noted concern over using a pattern-centered approach to understand instructional practices using the CLASS. This was a concern because the patterns that emerged from the latent profile analysis using the CLASS could have been modeled using one continuous variable. To address this concern, I used a variable that weighted the three domains of the CLASS (emotional support, instructional support, and classroom organization) by their standardized factor loadings. Results obtained from this method yielded standardized regression coefficients that were still smaller in magnitude than those obtained from using the pattern-centered approach ($\beta = .07$, $p = .003$ for video 2 state standardized assessment; $\beta = .10$, $p < .001$ for video 2 BAM; $\beta = .05$, $p = .05$ for video 3 BAM; $\beta = .05$, $p = .02$ for video 4 state standardized assessment; $\beta = .09$, $p < .001$ for average of videos state standardized assessment; $\beta = .08$, $p = .01$ for average of videos BAM). This suggests one of two things: (1) the whole is greater than the sum of

its parts, or (2), that the CLASS is useful in differentiating the teachers in the lowest third of teaching quality from the teachers in the top two thirds of teaching quality. To address point two, I modeled the association of the CLASS and achievement using a quadratic term; however, this quadratic term did not reach statistical significance and it is unclear if I can conclude that the CLASS is useful in differentiating teachers in the lowest third of teaching quality.

Discussion

I examined to what extent observations of mathematics teachers using domain-specific and domain-general protocols related to over the year changes in classroom level achievement. To understand this, I used both person and variable centered methods and looked at outcomes on two different assessments (state standardized assessment and the BAM). I also looked at whether or not instructional practices were associated with changes in the standard deviation of achievement.

Across all analyses, the domain-general protocol, the CLASS, did a better job predicting classroom level outcomes than did the domain specific protocol the MQI. Additionally, this association was stronger when using observation measures that were averaged across the four observation points than when individual time points were used. Finally, results from the person-centered approach highlighted differences in instructional practices more so than did the variable-centered approach. This was also true when using a weighted composite of the elements of the CLASS.

Domain-specific versus domain-general observation protocols

I hypothesized that the domain-specific protocol would be more predictive for student achievement than the domain-general protocol. The protocols have different

theoretical underpinnings and developed out of different literatures and it is therefore not surprising that they yielded different associations with student achievement. Some could argue that the MQI was designed to capture conceptual knowledge that might be better measured by the BAM than the state standardized assessment. However, the high correlations between the BAM and the state standardized assessment could explain why differences in the effects of teacher quality were not captured with the different observation protocols. However, this is not to say that researchers should rule out domain-specific measures for observing mathematics instruction. In many ways, domain-specific measures have higher face validity—that is, they look like they would be more accurate in making judgments about instructional practices that would indeed help students learn math. One explanation for the fact that I did not find significant associations between the MQI and classroom achievement is that there has been considerably less work done on reliability and validity of the MQI as compared with the CLASS. To the author's knowledge, no published studies exist using the MQI protocol to predict to student outcomes in samples other than the MET.

An interesting finding from this study was the significant positive association between emotional support and standard deviation of achievement (as measured by both achievement measures) and the significant negative association between instructional support and standard deviation of achievement as measured by the state standardized test. Because studies do not look at the variance in classroom-level achievement I do not have previous literature to draw on to explain these findings and can only speculate. It might be that teachers who are higher on emotional support are able to support lower achieving students' in ways they were not supported before. For example, emotional support could

lead lower achieving students who might not otherwise engage in adaptive learning behaviors such as help seeking to do so (Schenke, Lam, Conley, & Karabenick, 2015). Additionally, the positive association of instructional support and variance in student achievement might be due to a difference in students' interpretation of those events. Chapter 4 of this dissertation provides some explanation for this finding. It is unclear why these relationships exist only when looking at the fourth observed lesson. It could be that video 4 is the time point that is closest to when students' took the assessments. Future research should be done looking at standard deviation of classroom achievement if the ultimate goal for researchers and practitioners is to understand the instructional practices that are beneficial for low achieving students.

When to observe?

Results from this study point towards using multiple time points of observed instruction because only using one time point could yield different results depending on when instruction is observed. In this study, I found that the significant association between the CLASS and student achievement was only found when using the average of the four videos, only appeared in the first two videos. Only using one observation might yield researchers to conclude false negatives in regards to the association between instructional practices and student achievement. This might be because observations at certain time points were more unreliable than observations at earlier time points. It could also be that researchers need to take into account lagged effects and the time it takes for instructional practices to influence student achievement. We do not know yet *how* instructional practices have their effect on student achievement but it could be that instructional practices influence how students study the material or the questions they ask

during class. We therefore cannot expect high instructional quality to have an immediate influence on student achievement. Further theory must be developed to better understand this process. In chapter 4 of this dissertation, I elaborate on possible models researchers can investigate. Although this study was not designed to test for differences in using one, two, three or even four time points in measuring instructional practices, future studies, especially those using simulated data, would be helpful for researchers in understanding the optimal number of time points instructional practices should be measured. Simulation studies would allow researchers to check the probability of some event under different theoretical assumptions and would allow us to have a more nuanced understanding of how many times researchers and district personnel should observe a teacher's instructional practice. We know from issues of reliability that more is better in terms of reliability but as some have noted (e.g., Hill, Kapitula, & Umland, 2010), however, a higher number of observations comes at a price. If observation systems are to be used large scale across school districts to make inferences about the quality of instructional practices, having more knowledge about the exact number of lessons that need to be observed is paramount.

Pattern centered versus variable centered

In comparing the pattern centered and variable centered approaches to understanding instruction, it is unclear how different these approaches are. The variable-centered approached demonstrates an additive interpretation such that teachers who are rated higher on classroom organization yield higher gains in their classroom's average standardized achievement. When using an average of the three dimensions of the CLASS to predict to student achievement outcomes, I found that this average yielded regression

coefficients larger in magnitude than looking at any one of the dimensions separately. However, the pattern-centered approach which could arguably be described as representing a continuous rating of the CLASS, yielded even higher coefficients in magnitude than when classrooms were characterized using a variable-centered approach. Furthermore, I only found that the profile 1 of the CLASS (the profile in which teachers were rated lowest in teacher quality) was significantly negatively associated with achievement. I further tested this by modeling the association between the CLASS and achievement using a curvilinear equation, however, I did not find that this quadratic term was statistically significant. Though domain-general protocols have been theorized as capturing as being better at discriminating at the bottom of the distribution (Brophy, 1986), I did not find definitive evidence of this. Future research on classroom quality could take the approach that parenting researchers have taken, which is to characterize parenting in terms of a threshold of "good enough" instruction (Scarr, 1992). It might be that after a certain point, an additional unit of teaching quality is not beneficial for student achievement.

Limitations

There was considerable variability in when the observation were taken even when they were characterized as the same time point in this analysis. The relatively small associations between observed instruction and student achievement might really be lower bound estimates of the true effect. It might be that there are certain times in the year that are critical for instructional practices and that observations conducted at these critical time points yield stronger effects than observations conducted at non critical time points.

However, it is important to consider the findings from the previous chapter, which found that observations during the year are highly variable.

Another limitation concerns the reliability of the measures. Using observations from multiple observers yields observations that are more reliable, however, since fewer than 20% of the observations were doubly scored, I was unable to make use of this double scoring (Bill & Melinda Gates Foundation, 2013). Therefore, results using scores averaged across the four videos might be more closer to the true effects instructional quality has on student achievement.

Conclusion

Creating valid and reliable observation protocols that researchers and school administers can use is an important step in figuring out who the effective teachers are and which instructional practices they exhibit that make them effective. The aim of this study was to contrast different observation protocols (domain-general versus domain-specific), different time points (observations at four times points versus using the average of scores from the four time points), and different methods (pattern-centered versus variable-centered) in examining the association between observed instruction and student achievement. Though findings from this study do not point to specific implications for the design of teacher preparation programs or definitive models for how researchers should model the association between instructional practices and student achievement, this study does uncover some interesting insights. Specifically, the domain-general protocol did better than the domain-specific protocol in predicting to achievement, and that certain aspects of the domain-general protocol were significantly associated with the

standard deviation of student achievement. I hope this study will provide researchers with alternative ways of conceptualizing and defining what effective instruction is.

References

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

Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J.A. Bovaird, & N.A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255-284). Mahwah, NJ: Erlbaum.

Bill & Melinda Gates Foundation. (2013). *Measures of Effective Teaching (MET) project*: *Final report*. Seattle, WA: Author.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, *41*(10), 1069-1077.

Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328-375). New York: Macmillan.

Carpenter, T.P., Fennema, E., Franke, M.L., Empson, S.B., & Levi, L.W. (1999). *Children's mathematics: Cognitively guided instruction*. Portsmouth, NH: Heinemann.

Emmer, E. T., Evertson, C. M., & Brophy, J. E. (1979). Stability of teacher effects in junior high classrooms. *American Educational Research Journal, 16*, 71-75.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157-178.

Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis*,*19*(4), 325-338.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, 267-271.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. NCES 2003-013. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Greenwich, CT: Information Age.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430-511.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831.

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

Kane, T., Kerr, K., & Pianta, R. (2014). Designing teacher evaluation systems: New guidance from the measures of effective teaching project. San Francisco, CA: Jossey-Bass.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression* (4th ed.). New York: McGraw-Hill.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181.

Moos, R. H. (1979). *Evaluating educational environments*. San Francisco: Jossey-Bass.

Murdock, T. B., & Miller, A. (2003). Teachers as sources of middle school students' motivational identity: Variable-centered and person-centered analytic approaches. *The Elementary School Journal*, *103*(4), 383-399.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, *32*(1), 8-47.

Pianta, R.C., La Paro, K.M., & Hamre, B. (2007). Classroom Assessment Scoring System—CLASS. Baltimore, M.D.: Brookes.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, *45*(8), 74-85.

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

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*(3), 700-712.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. Journal of Educational Psychology, 101(3), 529-544.

Rittle-Johnson, B, & Star, J.R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99,* 561-574.

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

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

Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development*, *63*(1), 1-19.

Schenke, K., Lam, A. C., Conley, A. M, & Karabenick, S. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology, 41,*133-146.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of 'well-taught" mathematics courses. *Educational Psychologist*, *23*(2), 145-166.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759-775.

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

Stallings, J. A., Cory, R., Fairweather, J., & Needles, M. (1978). A study of basic reading skills taught in secondary schools. Menlo Park, CA: Stanford Research Institute.

StataCorp (2012). *Stata statistical software: Release 13*. College Station, TX: StataCorp LP.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*(2), 213-226.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Helping teachers learn to better incorporate student thinking. *Mathematical Thinking and Learning, 10*(4), 313-340

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., & Marder, M. (2011). Development of the UTeach Observation Protocol: A classroom

observation instrument to evaluate mathematics and science teachers from the UTeach Preparation Program (UTeach Technical Report 2011-01). *UTeach Natural Sciences, University of Texas at Austin*.

Wright, C. J., & Nuthall, G. (1970). Relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. *American Educational Research Journal*, 7(4), 477-491.

Table 3.1

Correlations for classroom achievement and classroom characteristics (N=527)

	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Class achievement state	.10 (.56)	1.00												
2. Class achievement BAM	.01 (.61)	0.72***	1.00											
3. Prev. class achievement state	.06 (.53)	0.88***	0.73***	1.00										
4. SD class achievement state	.72 (.19)	-0.13**	-0.13**	-0.13**	1.00									
5. SD class achievement BAM	.76 (.15)	-0.09*	-0.06	-0.09*	0.52***	1.00								
6. Class size	21 (5)	0.21***	0.27***	0.22***	-0.01	0.09*	1.00							
7. Percent white	25%	0.43***	0.64***	0.48***	-0.12**	-0.09*	0.03	1.00						
8. Percent Hispanic	30%	-0.13**	-0.14***	-0.15***	-0.06	0.01	-0.09*	-0.36***	1.00					
9. Percent black	35%	-0.39***	-0.56***	-0.40***	0.15***	0.06	-0.07	-0.55***	-0.49***	1.00				
10. Percent other ethnicity	10%	0.35***	0.38***	0.33***	-0.03	-0.00	0.28***	0.01	0.00	-0.40***	1.00			
11. Percent reduced lunch	56%	-0.12**	-0.28***	-0.14**	-0.04	-0.03	-0.20***	-0.37***	0.64***	-0.23***	0.05	1.00		
12. Percent male	50%	-0.04	-0.09*	-0.02	0.02	0.01	-0.01	-0.07	0.05	0.01	0.03	0.04	1.00	
13. Percent ELL	14%	-0.34***	-0.25***	-0.37***	0.09*	0.11*	-0.10*	-0.26***	0.58***	-0.23***	-0.04	0.25***	0.13**	1.00
14. Grade 4	29%	-0.10*	0.01	-0.08	0.14**	0.07	0.00	-0.02	-0.09*	0.11**	-0.08	-0.13**	-0.06	0.05
15. Grade 5	34%	0.03	0.06	0.05	0.17***	0.17***	0.08	0.01	-0.14**	0.10*	0.00	-0.17***	-0.01	-0.05
16. Grade 6	19%	0.07	0.02	0.06	-0.18***	-0.15***	-0.11**	0.04	0.11*	-0.15***	0.07	0.14**	-0.01	-0.06
17. Grade 7	15%	0.06	-0.06	0.04	-0.22***	-0.13**	0.06	0.03	0.06	-0.09*	0.04	0.14***	0.06	0.03
18. Grade 8	3%	-0.08	-0.11*	-0.12**	0.02	-0.05	-0.10*	-0.14**	0.23***	-0.05	-0.05	0.21***	0.08	0.09*

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Achievement variables are measured in standard deviation units. BAM is balanced assessment of mathematics. ELL is percent English language learner. 402 classrooms had valid data on free or reduced status.

Table 3.2

Associations between CLASS and achievement on state standardized test using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Emotional support	0.00	0.01	0.00	0.00	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Instructional support	-0.02	-0.05	0.02	0.00	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Classroom organization	0.09**	0.13***	0.05	0.06	0.15***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
		Covariates			
Previous achievement	0.77***	0.77***	0.77***	0.76***	0.77***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class size	-0.03	-0.03	-0.02	-0.03	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Percent white	0.02	-0.02	0.01	0.01	0.00
	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)
Percent Hispanic	-0.01	-0.06	-0.02	-0.02	-0.03
	(0.26)	(0.26)	(0.26)	(0.26)	(0.26)
Percent black	-0.10	-0.15	-0.12	-0.14	-0.13
	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
Percent other	0.05	0.03	0.05	0.04	0.04
	(0.15)	(0.14)	(0.15)	(0.15)	(0.14)
Percent ELL	-0.07	-0.06	-0.06	-0.06	-0.06
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Percent reduced lunch	-0.05	-0.07	-0.05	-0.05	-0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
5th Grade	-0.03	-0.02	-0.02	-0.03	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
6th Grade	-0.05	-0.03	-0.06	-0.06	-0.04
	(0.08)	(0.07)	(0.08)	(0.08)	(0.07)
7th Grade	0.00	0.01	0.02	0.01	0.04
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
8th Grade	-0.05	-0.07	-0.04	-0.06	-0.03
	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)
Constant	0.02	0.04	0.00	0.02	0.05
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. R squared for the models ranged from .73 to .74. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here.

Table 3.3

Associations between CLASS and achievement on BAM using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Emotional support	0.00	0.05	0.04	0.02	0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Instructional support	-0.05	0.01	0.00	-0.03	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Classroom organization	0.06	0.07•	0.02	0.06	0.09•
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
			Covariates		
Previous achievement	0.68•••	0.67•••	0.68•••	0.67•••	0.67•••
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class size	0.04	0.04	0.05	0.05	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent white	-0.63•	-0.61•	-0.64•	-0.60	-0.65•
	(0.31)	(0.31)	(0.31)	(0.31)	(0.31)
Percent Hispanic	-0.72•	-0.70•	-0.72•	-0.70•	-0.74••
	(0.29)	(0.28)	(0.29)	(0.29)	(0.28)
Percent black	-0.98••	-0.93••	-0.98••	-0.95••	-0.99••
	(0.36)	(0.36)	(0.36)	(0.36)	(0.36)
Percent other	-0.34•	-0.33•	-0.34•	-0.33•	-0.35•
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Percent ELL	-0.06	-0.07	-0.07	-0.06	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent reduced lunch	-0.13•	-0.14••	-0.12•	-0.13•	-0.13•
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.05•	-0.05	-0.05•	-0.06•	-0.05•
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
5th Grade	-0.06	-0.05	-0.05	-0.06	-0.06
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
6th Grade	-0.08	-0.06	-0.08	-0.08	-0.07
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
7th Grade	-0.09	0.00	-0.03	-0.06	-0.03
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)
8th Grade	-0.15	-0.08	-0.09	-0.11	-0.09
	(0.20)	(0.20)	(0.19)	(0.19)	(0.20)
Constant	0.34••	0.21•	0.14	0.17•	0.19•
	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)

Note. • $p < .05$, •• $p < .01$, ••• $p < .001$. Standard errors are in parentheses. R squared for the models ranged from .69 to .70. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. Previous achievement was measured using state standardized test.

Table 3.4

Associations between CLASS and standard deviation of classroom achievement using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Emotional support	0.01	-0.01	-0.00	-0.02*	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Instructional support	-0.00	-0.00	0.01	0.02*	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Classroom organization	-0.01	0.01	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Covariates					
Previous standard deviation	0.61***	0.62***	0.62***	0.61***	0.61***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Classroom Achievement	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.14	-0.13	-0.13	-0.14*	-0.12
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Percent Hispanic	-0.14*	-0.14*	-0.14*	-0.15*	-0.13*
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)
Percent black	-0.05	-0.05	-0.05	-0.06	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent other	-0.15	-0.15	-0.15	-0.16*	-0.14
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Percent ELL	0.00	0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
6th Grade	-0.04*	-0.04*	-0.04*	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.05*	-0.05*	-0.05*	-0.05*	-0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.09*	-0.10*	-0.09*	-0.10*	-0.09
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Constant	0.24***	0.24***	0.24***	0.24***	0.24***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here.

Table 3.5

Associations between CLASS and standard deviation of classroom BAM achievement using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Emotional support	0.01	-0.01	0.00	-0.02*	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Instructional support	-0.01	0.01	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Classroom organization	-0.01	0.00	-0.01	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Covariates				
Previous standard deviation	0.37***	0.37***	0.37***	0.36***	0.37***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Classroom Achievement	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.07	-0.05	-0.06	-0.07	-0.05
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Percent Hispanic	-0.06	-0.05	-0.05	-0.06	-0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Percent black	-0.03	-0.02	-0.03	-0.03	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent other	-0.07	-0.05	-0.06	-0.07	-0.05
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Percent ELL	0.01	0.01	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.02	0.02	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
6th Grade	-0.01	-0.00	-0.00	-0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.01	-0.00	0.00	-0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.06	-0.05	-0.04	-0.05	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Constant	0.49***	0.49***	0.49***	0.49***	0.49***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here.

Table 3.6

Associations between MQI and classroom achievement using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Errors and imprecision	0.02	-0.03	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Student participation in meaning making and reasoning	-0.02	0.01	-0.02	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Richness of mathematics	0.05	-0.01	-0.01	0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Working with students and mathematics	0.01	0.01	0.04	0.00	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Covariates				
Previous achievement	0.77***	0.77***	0.77***	0.77***	0.76***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class size	-0.03	-0.02	-0.03	-0.03	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Percent white	0.01	0.03	0.00	0.03	0.06
	(0.28)	(0.29)	(0.29)	(0.29)	(0.29)
Percent Hispanic	-0.02	0.00	-0.03	0.00	0.02
	(0.26)	(0.27)	(0.27)	(0.27)	(0.27)
Percent black	-0.12	-0.11	-0.14	-0.11	-0.08
	(0.33)	(0.33)	(0.34)	(0.33)	(0.33)
Percent other	0.05	0.06	0.05	0.06	0.07
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
Percent ELL	-0.07	-0.06	-0.06	-0.06	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent reduced lunch	-0.06	-0.05	-0.05	-0.05	-0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
5th Grade	-0.01	-0.02	-0.02	-0.02	0.00
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
6th Grade	-0.06	-0.06	-0.07	-0.07	-0.06
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
7th Grade	-0.01	-0.02	-0.03	-0.03	-0.02
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
8th Grade	-0.08	-0.07	-0.08	-0.09	-0.07
	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)
Constant	0.00	0.00	0.01	0.00	-0.01
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. R squared for the models ranged from .73 to .74. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here.

Table 3.7

Associations between MQI and achievement on BAM using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Errors and imprecision	0.00	-0.04	0.00	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Student participation in meaning making and reasoning	-0.03	0.04	0.01	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Richness of mathematics	0.03	0.01	-0.05	-0.03	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Working with students and mathematics	-0.02	-0.04	0.00	0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	Covariates				
Previous achievement	0.68***	0.67***	0.68***	0.67***	0.67***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class size	0.05	0.05	0.05	0.05	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent white	-0.63*	-0.61*	-0.67*	-0.63*	-0.62*
	(0.31)	(0.31)	(0.31)	(0.31)	(0.31)
Percent Hispanic	-0.72*	-0.70*	-0.76**	-0.71*	-0.71*
	(0.29)	(0.29)	(0.29)	(0.29)	(0.29)
Percent black	-0.99**	-0.96	-1.03**	-0.98**	-0.97**
	(0.36)	(0.36)	(0.36)	(0.36)	(0.36)
Percent other	-0.34*	-0.33*	-0.36*	-0.34*	-0.34*
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Percent ELL	-0.13*	-0.12*	-0.12*	-0.12*	-0.12*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent reduced lunch	-0.06*	-0.06*	-0.06*	-0.07	-0.07
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Percent male	-0.07	-0.07	-0.06	-0.06*	-0.06*
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
5th Grade	-0.05	-0.06	-0.06	-0.07	-0.06
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)
6th Grade	-0.09	-0.09	-0.10	-0.09	-0.09
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
7th Grade	-0.07	-0.07	-0.09	-0.08	-0.08
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
8th Grade	-0.13	-0.12	-0.14	-0.15	-0.14
	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
Constant	0.15	0.15	0.15	0.15	0.14
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. R squared for the models ranged from .68 to .69. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. Previous achievement was measured using state standardized test.

Table 3.8

Associations between MQI and standard deviation of classroom achievement using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Errors and imprecision	-0.01	-0.00	-0.00	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student participation in meaning making and reasoning	-0.00	-0.01	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Richness of mathematics	-0.00	0.01	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working with students and mathematics	-0.00	0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Covariates				
Previous standard deviation	0.62***	0.61***	0.61***	0.61***	0.62***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Classroom Achievement	-0.00	-0.00	-0.00	-0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.14	-0.13	-0.14	-0.13	-0.13
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Percent Hispanic	-0.14*	-0.13*	-0.14*	-0.14*	-0.14*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
Percent black	-0.05	-0.05	-0.05	-0.05	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent other	-0.16	-0.14	-0.16	-0.15	-0.15
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Percent ELL	0.01	0.01	0.01	-0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.00	-0.00	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent male	-0.00	-0.00	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
6th Grade	-0.04*	-0.04*	-0.04*	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.05*	-0.05*	-0.05*	-0.05*	-0.05*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.09*	-0.09*	-0.09*	-0.09*	-0.09*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	0.24***	0.24***	0.24***	0.24***	0.24***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Note. $*p < .05$, $**p < .01$, $***\ p < .001$. Standard errors are in parentheses All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here.

Table 3.9

Associations between MQI and standard deviation of achievement on BAM using a variable-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Errors and imprecision	-0.00	-0.00	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student participation in meaning making and reasoning	-0.01	0.01	0.01	-0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Richness of mathematics	0.00	0.00	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Working with students and mathematics	-0.00	-0.01	-0.01	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Covariates				
Previous standard deviation	0.37***	0.37***	0.37***	0.37***	0.37***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Classroom Achievement	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.06	-0.06	-0.05	-0.06	-0.05
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Percent Hispanic	-0.06	-0.06	-0.05	-0.05	-0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Percent black	-0.03	-0.03	-0.05	-0.03	-0.03
	(0.04)	(0.04)	(0.08)	(0.04)	(0.04)
Percent other	-0.07	-0.06	-0.03	-0.06	-0.05
	(0.08)	(0.08)	(0.04)	(0.08)	(0.08)
Percent ELL	-0.00	-0.00	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	0.00	0.00	0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.02	0.02	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
6th Grade	-0.01	-0.00	-0.00	-0.00	-0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	0.00	0.00	0.00	0.00	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.05	-0.04	-0.05	-0.04	-0.04
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	0.49***	0.49***	0.49***	0.49***	0.49***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Note. $*p < .05$, $**p < .01$, $***$ $p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. Previous standard deviation of achievement was measured using state standardized test.

Table 3.10

Associations between CLASS and achievement on state standardized test using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Profile 1	-0.20**	-0.23	-0.06	-0.12	-0.20*
	(0.08)	(0.13)	(0.06)	(0.07)	(0.08)
Profile 3	0.04	0.07	0.05	0.04	0.04
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
			Covariates		
Previous achievement	0.76***	0.77***	0.76***	0.76***	0.76***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Class size	-0.02	-0.03	-0.02	-0.04	-0.02
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Percent white	0.12	-0.01	-0.13	0.12	0
	(0.34)	(0.34)	(0.38)	(0.34)	(0.35)
Percent Hispanic	0.09	-0.03	-0.13	0.1	-0.02
	(0.31)	(0.32)	(0.35)	(0.32)	(0.32)
Percent black	0	-0.13	-0.28	-0.01	-0.11
	(0.39)	(0.40)	(0.44)	(0.40)	(0.41)
Percent other	0.11	0.03	-0.01	0.1	0.05
	(0.17)	(0.18)	(0.19)	(0.18)	(0.18)
Percent ELL	-0.07*	-0.06	-0.06	-0.07*	-0.07
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Percent reduced lunch	-0.06	-0.06	-0.06	-0.05	-0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.02	-0.01	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
5th Grade	-0.02	-0.01	-0.01	0	0.01
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
6th Grade	-0.06	-0.05	-0.05	-0.05	-0.02
	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)
7th Grade	0.02	0.01	0.02	0.04	0.06
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
8th Grade	-0.04	-0.13	-0.04	-0.03	0.02
	(0.23)	(0.25)	(0.23)	(0.23)	(0.23)
Constant	0.04	0.02	-0.04	-0.01	-0.03
	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)

*Note. * $p < .05$, ** $p < .01$, *** $p < .001$.* Standard errors are in parentheses. R squared for the models ranged from .72 to .73. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. Profile 2is the reference profile. District dummies were included in the models but are not shown here.

Table 3.11

	Video 1	Video 2	Video 3	Video 4	Average
Profile 1	-0.13	-0.22	-0.17*	-0.16	-0.29**
	(0.09)	(0.13)	(0.07)	(0.09)	(0.09)
Profile 3	-0.10	0.14*	0.01	0.00	-0.04
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
			Covariates		
Previous achievement	0.67***	0.67***	0.68***	0.67***	0.66***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Class size	0.04	0.04	0.06	0.04	0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent white	-0.61	-0.55	-0.75	-0.61	-0.62
	(0.42)	(0.42)	(0.39)	(0.41)	(0.42)
Percent Hispanic	-0.69	-0.65	-0.82*	-0.69	-0.70
	(0.39)	(0.39)	(0.37)	(0.38)	(0.40)
Percent black	-0.96	-0.87	-1.10*	-0.95*	-0.96
	(0.49)	(0.49)	(0.46)	(0.48)	(0.50)
Percent other	-0.32	-0.30	-0.40*	-0.33	-0.34
	(0.21)	(0.21)	(0.20)	(0.21)	(0.22)
Percent ELL	-0.06	-0.07	-0.07	-0.07	-0.07
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Percent reduced lunch	-0.13*	-0.13*	-0.11*	-0.12*	-0.12*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.06*	-0.05	-0.05*	-0.06*	-0.05*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
5th Grade	-0.04	-0.03	-0.05	-0.03	-0.05
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
6th Grade	-0.06	-0.05	-0.08	-0.08	-0.09
	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
7th Grade	-0.07	-0.01	0.00	-0.02	-0.04
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)
8th Grade	-0.11	-0.13	-0.05	-0.10	-0.11
	(0.20)	(0.25)	(0.25)	(0.25)	(0.23)
Constant	0.19*	0.11	0.16	0.14	0.18
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)

*Note. * $p < .05$, ** $p < .01$, *** $p < .001$.* Standard errors are in parentheses. R squared for the models ranged from .67 to .68. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. Profile 2is the reference profile. District dummies were included in the models but are not shown here. Previous achievement was measured using state standardized test.

Table 3.12

Associations between CLASS and standard deviation of classroom achievement using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Profile 1	-0.01	0.00	-0.00	0.01	-0.00
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Profile 3	0.01	-0.00	-0.01	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	Covariates				
Previous standard deviation	0.61***	0.61***	0.61***	0.60***	0.62***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Classroom Achievement	-0.00	0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.14*	-0.16*	-0.13*	-0.14*	-0.14*
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Percent Hispanic	-0.15**	-0.16**	-0.14**	-0.15**	-0.15**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Percent black	-0.06	-0.06*	-0.05	-0.05	-0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Percent other	-0.16*	-0.17*	-0.15*	-0.16*	-0.17*
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Percent ELL	0.00	0.00	0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	-0.00	-0.00	-0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.01	0.01	0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
6th Grade	-0.04*	-0.05*	-0.04*	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.05*	-0.05*	-0.05*	-0.05*	-0.04
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.10	-0.09	-0.10	-0.11*	-0.10
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Constant	0.25***	0.25***	0.25***	0.25***	0.24***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. The reference profile is Profile 2.

Table 3.13

Associations between CLASS and standard deviation of classroom BAM achievement using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4	Average
Profile 1	0.03	0.02	0.01	0.03	0.02
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Profile 3	-0.01	-0.00	0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
			Covariates		
Previous standard deviation	0.37***	0.37***	0.38***	0.37***	0.35***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Classroom Achievement	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.07	-0.07	-0.07	-0.08	-0.06
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Percent Hispanic	-0.07	-0.07	-0.07	-0.07	-0.06
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Percent black	-0.04	-0.04	-0.04	-0.04	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Percent other	-0.08	-0.08	-0.08	-0.08	-0.07
	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)
Percent ELL	0.00	0.01	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.01	-0.00	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Male	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.02	0.02	0.02	0.02	0.02
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
6th Grade	-0.00	-0.00	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.01	-0.00	-0.00	-0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.06	-0.05	-0.06	-0.05	-0.04
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
Constant	0.50***	0.49***	0.48***	0.48***	0.49***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)

Note. $*p < .05$, $**p < .01$, $***p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. The reference profile is Profile 2.

Table 3.14

Associations between MQI and classroom achievement using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4
Profile 1	0.07	-0.07	-0.02	-0.10
	(0.11)	(0.07)	(0.09)	(0.08)
Profile 3	0.02	-0.05	-0.01	0.04
	(0.06)	(0.07)	(0.06)	(0.07)
Covariates				
Previous achievement	0.78***	0.77***	0.76***	0.76***
	(0.03)	(0.03)	(0.03)	(0.03)
Class size	-0.02	-0.02	-0.02	-0.02
	(0.03)	(0.03)	(0.04)	(0.03)
Percent white	-0.04	0.00	0.01	-0.05
	(0.35)	(0.37)	(0.35)	(0.37)
Percent Hispanic	-0.06	-0.02	-0.02	-0.08
	(0.32)	(0.34)	(0.32)	(0.35)
Percent black	-0.17	-0.14	-0.13	-0.22
	(0.41)	(0.43)	(0.41)	(0.44)
Percent other	0.03	0.05	0.05	0.02
	(0.18)	(0.19)	(0.18)	(0.19)
Percent ELL	-0.06	-0.06	-0.06	-0.07
	(0.04)	(0.03)	(0.03)	(0.04)
Percent reduced lunch	-0.05	-0.04	-0.06	-0.04
	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.03	-0.02	-0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)
5th Grade	-0.02	-0.02	-0.02	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)
6th Grade	-0.07	-0.06	-0.07	-0.08
	(0.07)	(0.07)	(0.07)	(0.07)
7th Grade	-0.01	-0.03	-0.02	-0.04
	(0.07)	(0.07)	(0.08)	(0.08)
8th Grade	-0.10	-0.08	-0.01	-0.10
	(0.24)	(0.22)	(0.24)	(0.22)
Constant	-0.01	0.00	0.01	0.00
	(0.09)	(0.08)	(0.08)	(0.08)

*Note. * $p < .05$, ** $p < .01$, *** $p < .001$.* Standard errors are in parentheses. R squared for the models was .73. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. Profile 2 is the reference profile. District dummies were included in the models but are not shown here.

Table 3. 15

Associations between MQI and achievement on BAM using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4
Profile 1	-0.01	-0.05	-0.02	0.02
	(0.13)	(0.07)	(0.09)	(0.09)
Profile 3	-0.05	-0.01	-0.01	0.08
	(0.07)	(0.08)	(0.06)	(0.09)
Covariates				
Previous achievement	0.68***	0.67***	0.68***	0.67***
	(0.04)	(0.04)	(0.04)	(0.04)
Class size	0.04	0.06	0.06	0.05
	(0.04)	(0.04)	(0.04)	(0.04)
Percent white	-0.68	-0.62	-0.46	-0.59
	(0.44)	(0.42)	(0.39)	(0.43)
Percent Hispanic	-0.77	-0.71	-0.56	-0.68
	(0.41)	(0.39)	(0.37)	(0.40)
Percent black	-1.04*	-0.97*	-0.77	-0.94
	(0.51)	(0.49)	(0.46)	(0.50)
Percent other	-0.36	-0.34	-0.26	-0.32
	(0.22)	(0.21)	(0.20)	(0.21)
Percent ELL	-0.07	-0.06	-0.06	-0.07
	(0.04)	(0.04)	(0.04)	(0.04)
Percent reduced lunch	-0.10*	-0.13*	-0.13*	-0.12*
	(0.05)	(0.05)	(0.05)	(0.05)
Percent male	-0.07**	-0.06*	-0.06*	-0.06*
	(0.03)	(0.02)	(0.03)	(0.03)
5th Grade	-0.05	-0.06	-0.03	-0.07
	(0.07)	(0.07)	(0.07)	(0.07)
6th Grade	-0.07	-0.09	-0.09	-0.09
	(0.08)	(0.08)	(0.08)	(0.08)
7th Grade	-0.05	-0.09	-0.07	-0.08
	(0.08)	(0.08)	(0.08)	(0.08)
8th Grade	-0.11	-0.14	-0.16	-0.11
	(0.25)	(0.25)	(0.25)	(0.25)
Constant	0.13	0.15	0.15	0.13
	(0.09)	(0.09)	(0.10)	(0.09)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. R squared for the models was 68. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. Profile 2 is the reference profile. District dummies were included in the models but are not shown here. Previous achievement was measured using state standardized test.

Table 3.16

Associations between MQI and standard deviation of classroom achievement using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4
Profile 1	-0.04	-0.03	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Profile 3	-0.01	-0.03	-0.00	0.01
	(0.01)	(0.02)	(0.01)	(0.02)
	Covariates			
Previous standard deviation	0.61***	0.61***	0.62***	0.61***
	(0.04)	(0.04)	(0.04)	(0.04)
Classroom Achievement	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.13*	-0.13*	-0.15*	-0.14*
	(0.06)	(0.06)	(0.06)	(0.06)
Percent Hispanic	-0.13*	-0.14**	-0.15**	-0.14**
	(0.05)	(0.05)	(0.06)	(0.05)
Percent black	-0.05	-0.05	-0.17*	-0.05
	(0.03)	(0.03)	(0.07)	(0.03)
Percent other	-0.14*	-0.16*	-0.06	-0.16*
	(0.07)	(0.07)	(0.03)	(0.07)
Percent ELL	0.01	0.01	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.00	-0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Male	-0.00	0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
6th Grade	-0.04*	-0.04*	-0.04*	-0.04*
	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.05*	-0.05*	-0.05*	-0.05*
	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.09	-0.09	-0.11*	-0.10
	(0.05)	(0.05)	(0.05)	(0.05)
Constant	0.25***	0.25***	0.25***	0.24***
	(0.04)	(0.04)	(0.04)	(0.04)

Note. $*p < .05$, $**p < .01$, $***\ p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. The reference profile is Profile 2.

Table 3.17

Associations between MQI and standard deviation of classroom BAM achievement using a pattern-centered approach (N=527).

	Video 1	Video 2	Video 3	Video 4
Profile 1	0.01	-0.04*	0.03	0.01
	(0.03)	(0.02)	(0.02)	(0.02)
Profile 3	-0.00	-0.00	0.03	0.00
	(0.01)	(0.02)	(0.01)	(0.02)
	Covariates			
Previous standard deviation	0.37***	0.37***	0.37***	0.37***
	(0.04)	(0.04)	(0.04)	(0.04)
Classroom Achievement	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Class size	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Percent white	-0.06	-0.06	-0.05	-0.06
	(0.06)	(0.06)	(0.08)	(0.06)
Percent Hispanic	-0.06	-0.05	-0.05	-0.06
	(0.06)	(0.06)	(0.07)	(0.06)
Percent black	-0.03	-0.03	-0.06	-0.03
	(0.03)	(0.03)	(0.09)	(0.03)
Percent other	-0.06	-0.06	-0.02	-0.06
	(0.07)	(0.07)	(0.04)	(0.07)
Percent ELL	-0.01	-0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Percent reduced lunch	-0.00	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Male	0.00	0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
5th Grade	0.02	0.02	0.02	0.02
	(0.01)	(0.02)	(0.01)	(0.01)
6th Grade	-0.01	-0.01	0.00	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
7th Grade	-0.00	-0.00	-0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
8th Grade	-0.05	-0.04	-0.04	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)
Constant	0.49***	0.49***	0.49***	0.49***
	(0.03)	(0.03)	(0.04)	(0.03)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Standard errors are in parentheses. All continuous variables are standardized. Fourth grade was the reference grade. ELL is percent English language learner. Percent other refers to percent other ethnicity. District dummies were included in the models but are not shown here. The reference profile is Profile 2.

Chapter 4: Mediating the association between instructional practices and student achievement: The role of student perceptions

A primary goal of educational research is to understand how and under what conditions students learn in classrooms. This is particularly relevant in the current climate of accountability where the use of value-added models to measure teacher and school effects on students' achievement is ubiquitous. Critics argue that value-added model approaches may not accurately account for the quality of classroom instruction (Hill, Kapitula, & Umland, 2011) and are not useful in giving meaningful feedback to teachers about their instruction (Hill & Grossman, 2013). Along with these critiques is the issue that these models do not include explanations as to how and why certain classroom environments are more or less adaptive for student learning. Truly understanding a teacher's impact on their students and how best to improve student learning requires a richer picture of the classroom experience—instructional practices *and* student experiences—than can be provided from test scores alone. Whereas effective instructional practices have been identified in a variety of studies, understanding students' role in their own learning as they relate to these instructional practices is less emphasized in the current discussions of teacher effectiveness (for exceptions, see Kunter, Tsai, Klusmann, Brunner, Kraus, & Baumert, 2008; Wagner, Göllner, Helmke, Trautwein, & Lüdtke, 2013). To address this gap in the current literature, my purpose is to use student perceptions of the classroom as a lens to better understand teaching and student learning. As such, I identify one such mechanism for student learning: student perceptions of the classroom.

Evidence suggesting that teacher value-added models are poor indicators of classroom instructional quality is beginning to accumulate. Hill and colleagues (2011) compared teacher's value-added scores with observations of their classroom instruction in mathematics and found that some teachers had high value-added scores but scored poorly on observations of mathematics instructional quality when rated by external observers. In another study, Bitler and colleagues (*in preparation*) compared teacher effects on student achievement with their effects on students' height (a variable that cannot possibly be affected by the teacher) and found similar effect sizes for both. Both lines of research, although oriented in different research traditions (one from an economics perspective and one from a teacher learning perspective), suggest that value-added models do not provide researchers and policymakers with a full understanding of identifying effective mathematics teachers. Whereas, much of the earlier research on teaching and learning focused on students as participants and agents of their own learning (e.g., Bandura, 1986; Berliner, 1976; Moos, 1979; Weinstein, 1982; Wittrock, 1986), student agency is less emphasized in current discussions of how students interpret and respond to elements of mathematics instruction. I argue that a critical element of how instruction "gets" into students is through their perceptions of that instruction. Students' interpretation of their environment is critical to their responses to that environment (Patrick, Anderman, Ryan, Edlein, & Midgley, 2001; Skinner & Belmont, 1993) and should be considered in models of teacher effectiveness.

Student Perceptions

There has been a long tradition of using student perceptions of the classroom (e.g., Ames & Archer, 1988; Brattesani, Weinstein, & Marshall, 1984; Feldlaufer,

Midgley, & Eccles, 1988; Trickett & Moos, 1973). There are several ways researchers have approached using student perceptions of the classroom. Student perceptions have been used as single or interraters of the classroom (Fauth, Decristan, Rieser, Klieme, & Buttner, 2014; Marsh, Lüdtke, Nagengast, Trautwein, Morin, Abduljabbar, & Köller, 2012). This means that student surveys are administered as the only or additional measure of classroom practices. Because students have daily experiences in the classroom, student reports are often used to measure dimensions of the classroom based on the assumption that external observations do not accurately capture students' long-term interactions within the classroom (Dorman, 2008; Wagner et al., 2013; Kane & Staiger, 2012). Thus, they are used to obtain a more reliable measure of what is happening in the classroom.

Student perceptions have also been considered as providing unique accounts of what happens in the classroom suggesting that students' previous experiences color how they perceive their current classroom environment (Ames & Archer, 1998). It distinguishes what is actually observed (*alpha press*) from participants' interpretation of what is observed (*beta press*; Murray, 1938). In classroom research, it suggests that multiple measures of the classroom should be combined, such as student surveys and external observations because they explain unique aspects of classroom practices (Turner & Meyer, 2000). Indeed, previous research has found that for certain dimensions of the classroom environment, student reports cannot reliably be aggregated to the classroom-level and that perceptions are best thought of as unique perspectives of the classroom (Lam, Ruzek, Schenke, Conley, & Karabenick, 2015; Miller & Murdock, 2007). This perspective is useful in explaining the variance components that predict to student achievement. For example, observations of classrooms are thought to explain less

variance in student outcomes than students' perception of their classroom (see Weinstein, 1982 for a review). What is not explained by this perspective, is to what extent these perceptions are influenced by an objective account of what is happening in the classroom. Such a path would allow researchers and practitioners to understand what aspects of the classroom can be purposefully designed and manipulated with the goal of increasing student achievement through influencing their perceptions.

Student perceptions of their environment are an integral part of several theories including social-cognitive theory (Bandura, 1986), expectancy-value theory (Eccles et al., 1983), and attribution theory (Weiner, 1985) and are thought to be critical mediators of students' experiences on motivation and behavior. Within this view, students form their perceptions of their classroom environment based on their interactions with and in that environment. Whereas many factors are thought to affect students' perceptions, such as their previous experiences and individual characteristics (Ames & Archer, 1988), understanding the elements of the classroom that students attend to is necessary for the design of classroom instruction. Despite the strong theoretical support for using student perceptions as mediators of environment and student outcomes, empirical evidence of this pathway is limited. As such, the purpose of this paper is to test the mediation model that the effect of instructional practices on student achievement outcomes can, in part, be explained by students' perceptions of those instructional practices.

Student perceptions of their classroom environments have some effects on their learning (see Urdan, 2010). Through teachers' use of instructional practices and policies within their classrooms, it is thought that teachers create classroom environments that directly and indirectly send messages to students about the goals of learning and

achievement (Ames, 1992; Kaplan, Middleton, Urdan, & Midgley, 2002; Urdan, 2010). The literature on student perceptions has identified several theories and aspects of student perceptions that matter for achievement, but in the present study, I focus on four dimensions of student perceptions—*emotional support*, *instructional support*, *interest*, and *academic press*—that can be placed in the framework of Self Determination Theory (SDT; Ryan & Deci, 2000). SDT posits that students will excel if their needs for competence, relatedness, and autonomy are met. Academic press and instructional support fall under providing students with feelings of competence, emotional support falls under feelings of relatedness, and value relates to students' feeling of autonomy. These dimensions are not meant to be exhaustive but present examples of constructs with which to measure student perceptions. I therefore do not go deep into a literature review on each construct but rather provide some rationale as to why these constructs matter.

Students' perceptions of emotional support refer to the teacher-student interactions that promote students' social and emotional functioning such as feelings of belongingness and positive peer relationships (Pianta, LaParo, & Hamre, 2007; Ryan & Deci, 2000). Emotional support affects child-related outcomes such as engagement with school, academic performance, increased student motivation, and positive behavioral outcomes (Pianta et al., 2012; Ryan & Deci, 2000). Instructional support (or instructional quality as referred to in Wagner et al., 2013) refers to the extent to which students perceive cognitive supports in classroom instruction. Academic press—the extent to which classrooms emphasize achievement-oriented values and goals—captures the demands that are placed on students (Middleton & Midgley, 2002). For example, classrooms that are high in academic press are those where the teacher emphasizes deep

understanding of the material, thorough thinking, placing full effort on work, and accuracy of answers (Henningson & Stein, 1997; Middleton & Midgley, 2002: Shouse, 1996). The last construct refers to environments that promote students' interest in the material. This relates to students' feelings of autonomy because they are acting from their values and interests (Ryan & Deci, 2000).

Observing instruction

Classrooms have been systematically observed for decades (see Rosenshine, 1970 for a review of early work; Simon & Boyer, 1974) and much work has gone into understanding the dimensions of classroom instruction that should be observed. Though much early work exists on the development of observation protocols, this knowledge is not referenced in current discussions of observing teaching. A recent focus in this area has been in developing domain-general protocols to observe instructional practices that are of practical use to teachers and administrators. Relevant to the present study, the Classroom Assessment Scoring System (CLASS; Pianta, Hamre, Hayes, Mintz, & LaParo, 2008; Pianta, LaParo, & Hamre, 2007) is one protocol that has gained a lot of traction in recent years. The CLASS is an observational protocol that provides a measure of the global classroom quality on the following three domains: emotional support, classroom organization, and instructional support. Please refer to Pianta and Hamre (2009) for a review of the theoretical underpinnings of the protocol.

Studies using the CLASS as a primary observation tool have been used in a variety of grades such as kindergarten and middle school (Mashburn et al., 2008; Allen, Gregory, Mikami, Lin, Hamre & Pianta, 2013) and in a variety of domains such as mathematics and English Language Arts. Notably, aspects of the CLASS (emotional

support, instructional support) have consistently predicted to achievement outcomes (Allen et al., 2013; Hamre & Painta, 2005; Mashburn et al., 2008). In addition, all CLASS dimensions—depending on how the dimensionality of the CLASS is treated, whether constructs are inputted into analyses together or separately—have been found to predict to student achievement on standardized measures (Allen et al., 2013; Hamre & Pianta, 2005; Mashburn et al., 2008) as well as students' cognitive self control, behavioral engagement, and time off task (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009).

Present Study

I seek to understand whether certain aspects of observed instruction correspond to students' perceptions. By and large, recent studies of perceptions of classroom climate have not included other dimensions of the classroom, such as quality of instruction as perceived by external observers (c.f. Downer et al., 2014; Feldlaufer, Midgley, & Eccles, 1998; Ruzek, Hafen, Allen, Gregory, Mikami & Pianta, *under review*; Turner, Midgley, Meyer, Gheen, Anderman, & Kang, 2002). By failing to focus on both observed instructional quality and student perceptions, recent work has overlooked critical elements of the classroom environment, which could provide more predictive power of instructional practices on student outcomes. As such, the present study incorporates conceptualizations of instructional quality with classroom climate (both observed and perceived by students) to better understand classroom practice.

The present study uses multilevel structural equation modeling to determine whether individual student perceptions of the classroom mediate the effect of observed instructional practices on student achievement outcomes. I ask to what extent do student

perceptions of instructional practices explain the effects of instruction on achievement. A critical feature of the present study is the ability to detect effects of instructional practices on student outcomes using a multi-level design that includes both student level (level 1) predictors and classroom level (level 2) observations of classroom practices. The use of data from two reporters of the classroom—external observers and students—provides additional evidence for the validity of these constructs. Furthermore, the present study's longitudinal design affords the necessary power to detect such effects using a large (*N*=527) sample of classrooms.

Method

Participants

Data for this study come from year 2 of the Bill & Melinda Gates Foundation Measures of Effective Teaching study (Bill & Melinda Gates Foundation, 2013). Elementary and middle school students were randomly assigned to teachers to more closely examine the association between instructional practices and student outcomes without the confound of students selecting into particular classrooms. The study took place in six urban schools districts in six states (New York City Department of Education, Charlotte-Mecklenburg Schools, Denver Public Schools, Memphis City Schools, Dallas Independent School District, and Hillsborough County Public Schools). Only teachers who were either generalists (taught all school subjects in a grade) or specialists (i.e. only taught mathematics) were included in the study. This included 152 fourth grade, 179 fifth grade, 102 sixth grade, 79 seventh grade, and 15 eighth grade

teachers, totaling 527 teachers.[5] Teachers were 84% female, 61% White, 32% African American, 5% Hispanic, and 2% other ethnicity. Thirty nine percent of the sample teachers had a Masters degree or higher and had a mean of 11 years teaching experience with a standard deviation of 9 years (range from 0 to 39 years of teaching experience).

Eligibility for the analysis sample consisted of two steps: eligible classrooms/teachers[6] and eligible students in those classrooms. To be an eligible classroom, classroom observations must have occurred at all four time points during the year, be coded on both the CLASS, and be a mathematics classroom (either specialist or generalist). Five hundred and twenty-seven classrooms consisting of 10,994 students were eligible in the present study. The average classroom size was 22 students (standard deviation of 5) with a range of 7 to 34 students. The sample of students was ethnically diverse with 35% African American, 25% White, 29% Hispanic, and 11% other ethnicity, was 50% male, and consisted of 14% English Language learners. One percent of the sample was missing information on all demographic characteristics. Students were eligible to be included in the analysis sample if they came from classrooms that were in the analysis sample, had valid standardized state achievement data, and responded to the student perception survey during the school year. Of the initial 10,994 students who were in the study sample of classrooms, only 6,930 (63% of the students in eligible classrooms) participated in the student survey. This included 3,548 elementary school students in 322 classrooms (147 fourth and 175 fifth grade classrooms) and 3,382 middle

[5] Algebra 1 teachers and students were not included in the study because the protocol that was used to observe Algebra 1 teachers was different than the protocol used to observe general mathematics instruction.
[6] Only one classroom was included for each teacher. Therefore, teachers and classrooms are synonymous.

school students in 190 classrooms (98 sixth, 78 seventh, and 14 eighth grade classrooms). The demographics for those who completed the student survey was similar to those in the larger sample with 50% male, 14% English Language Learner, 27% White, 30% African American, 31% Hispanic, and 12% other ethnicity. Academic achievement significantly differed between the initial sample and analysis sample for elementary school and middle school students. For elementary school the mean standardized achievement for the initial sample was .09, whereas the mean standardized achievement for elementary school students who were eligible to participate in the survey was .12 ($t = 2.01$, $p = .04$). For middle school the mean standardized achievement was .21 in the initial sample but the mean standardized achievement for middle school students eligible in the analysis sample was .27 ($t = 3.50$, $p < .001$). This statistically significant difference is not too concerning because the current study tests one mechanism by which observed might instructional practices influence student achievement, and not whether or not that relationship exists.

Measures and Procedure

I analyzed data from three sources: observed instructional practices, surveys taken by students on their perceptions of the classroom, and demographic and test score information obtained from each of the school districts. These sources are divided into level 1 (student-level) and level 2 (classroom-level) characteristics and are appropriately modeled as such. Students completed surveys once a year either online or in class about their mathematics teacher. The timing of when students took the survey varied. Thirty one percent of middle school students reporting taking the survey when they were in the class for less than four months, 9% reported taking the survey when being in their class for more than four months but not all year and 60% of students reported taking the survey

after having been in the class all year. Four percent of elementary school students reported taking the survey after having been in their class for less than two weeks, 23% of elementary school students reported taking the survey when they were in the class for three or four weeks, and 73% of students reported taking the survey when they were in the class for more than two months. Elementary and middle school schools students were given different versions of the survey; as such, all analyses are conducted separately for elementary and middle school students to take differences in the measures into account.

Classroom-level data.

Classroom Assessment Scoring System (CLASS). Video-recorded observational data from one lesson was used. Research assistants were trained and certified on the CLASS protocol and completed a daily calibration exercise. Coding consisted of watching the first thirty minutes of each mathematics lesson and was conducted in 15-minute increments on the dimensions of emotional support, instructional support, and classroom organization. Observers watched the video and took notes on specific observed behaviors and subsequently code according to the CLASS protocol. Raters indicated their scores on a 7-point scale ranging from low (1-2), middle (3-5), and high (6-7). Scores for each coded segment were then averaged across the entire lesson. I used observation scores from videos taken at the first time point.

Level-2 covariates. Classroom grade level (fourth and fifth in the elementary school sample; and sixth, seventh, and eighth in the middle school sample) and district dummy variables were used as level-2 covariates.

Student-level data.

Student surveys. Students completed either the elementary school version or the middle school version of the Tripod assessments (Ferguson, 2010). Surveys were completed once during the school year either during mathematics class via paper and pencil (36% of elementary school students and 43% for middle school students) or at home through an online survey (64% for elementary school students and 57% for middle school students).

The Tripod was developed to measure student perceptions of seven "C's" of the classroom: caring, captivating, conferring, controlling, clarifying, challenging, and consolidating. Example questions are: "*My teacher gives us time to explain our ideas*" and "*Our class stays busy and doesn't waste time*". Students had to rate their agreement on each statement on a scale of one (totally untrue) to five (totally true). Students responded to a total of 61 items in elementary school and 69 items in middle school. Recent work looking at the factor structure of the Tripod has not found evidence of adequate factor structure for the seven proposed dimensions of the Tripod (Schweig, 2014; Wallace, Kelcey, & Ruzek, *under review*). Therefore, I used my own conceptual categorization of the items on the Tripod survey to serve as my scales of interest that were aligned with the Self Determination Theory framework. Because surveys and items differed by elementary school and middle school, I conducted all analyses (including Confirmatory Factor Analyses [CFAs] and scale construction) separately by middle and elementary school. Below, I describe the items and model fit indices of the four dimensions. Two constructs (emotional support and instructional support) were directly aligned with the CLASS whereas the other two constructs (interest and academic press)

were not. The complete list of items and constructs are listed in Table 1 for middle school students and Table 2 for elementary school students.

Interest. Items indicating the extent to which students felt boredom, interest, or value in their mathematics classroom were characterized as being indicators of a latent variable called interest (e.g., "This class does not keep my attention—I get bored", and "My teacher makes learning enjoyable"). It is important to note that there was an insufficient number of interest items for the elementary school sample and thus, interest was only investigated in the middle school sample using three items. Standardized factor leadings ranged from .61 to .88.

Emotional Support. Items that characterized the extent to which students perceived their teachers as emotionally supportive asked about identifying with students' thoughts and feelings and the extent to which teachers cared for their students (e.g., "My teacher in this class makes me feel that s/he really cares about me", and "My teacher respects my ideas and suggestions"). Three items were used to characterize emotional support for middle school students and seven items were used to characterize emotional support for elementary school students. Results from the CFA revealed excellent fit of the data for elementary school sample (CFI = .95; TLI = .93; SRMR = .03; RMSEA = .07; Hu & Bentler, 1999). Standardized factor loadings ranged from .75 to .78 for the middle school sample and .45 to .78 for the elementary school sample.

Instructional Support. Items that characterized the extent to which students reported the teacher made efforts to support instruction (e.g., "We get helpful comments to help us know what we did wrong on assignments" and "My teacher gives us time to explain our ideas") were characterized as instructional support. Eleven items were used to

characterize instructional support in the middle school sample and 12 items were used to characterize instructional support in the elementary school sample. CFAs revealed excellent fit for both the middle school (CFI = .98; TLI = .98; SRMR = .02; RMSEA = .04) and elementary school (CFI = .97: TLI = .96; SRMR = .03; RMSEA = .04) samples. Standardized factor loadings ranged from .56 to .76 in the middle school sample and from .47 to .66 in the elementary school sample.

Academic Press. Items that characterized the extent to which students perceived teachers pushing students to do their best (e.g., "My teacher pushes everyone to work hard", and "The teacher in this class encourages me to do my best") were used to create a measure of academic press. Four items were used to characterize academic press in the middle school and elementary school samples. CFAs revealed excellent fit for both the middle school (CFI = 1.00; TLI = 1.00; RMSEA = .00; SRMR = .00) and elementary school samples (CFI = .97; TLI = .96; SRMR = .03; RMSEA = .04). Standardized factor loadings ranged from .54 to .67 in the middle school sample and .42 to .57 in the elementary school sample.

Mathematics achievement. Data from state standardized achievement tests were used as the measure of mathematics achievement. Assessments differed by state, and neither raw scores nor item-level data were available for students. Instead, students' standardized scores by grade and district were provided to the researcher such that standardized scores were computed using all students in the district (including those not in the current study sample). I controlled for grade-level and district as level-2 covariates. Therefore, coefficients are interpreted for students in the same grade and school district.

Level-1 covariates. Students' gender (i.e. male or female), ethnicity (i.e., white, African American, Hispanic, or other), and English Language learner status were used as covariates in my models. Additionally, students' standardized value on the mathematics assessment from the previous year was used to test for changes in students' mathematics achievement as a result of observed instructional practices and student perceptions.

Overview of the Analytic Models and Methods

To determine whether or not student perceptions of classroom climate mediated the effects of instructional practices on student achievement, I used multilevel structural equation modeling (MSEM; Preacher, Zyphur, & Zhang, 2010). An important feature of MSEM is the ability to parse out the between-level variance from within-level variance, that is the effect of classroom-level or level 2 variables on student-level or level-1 effects. Because my model is looking at level-1 variables (student perceptions) mediating the effect of a level-2 variable (observed instruction) on a level-1 outcome (students' mathematics achievement), I used Preacher and colleagues' (2010) 2-1-1 method for mediation. This is appropriate because all students in the same class received the same observed instruction. A limitation to this model is that I cannot investigate whether or not the effect of instructional practices on student achievement varies within classrooms; I can only look at if the effect of instructional practices on student achievement varies between classrooms. Results of the 2-1-1 model are interpreted as the average classroom practices influencing classroom average student perceptions and classroom average achievement, such that the classroom-level variable does not have effects on the individual level but can only explain variance at the classroom level. This is important to note because true 2-1-1 level effects cannot be evaluated in this model since a classroom-

level variable can only have effects on classroom-level phenomena. Therefore, level-1outcomes are actually aggregated to the classroom level, where the measurement model for the student perception latent variable was specified to be the same across both levels (L1 and L2). As a result, the interpretation of the mediation model is whether, and to what degree, *classroom-level variability* in student perceptions serves as a mediator of the classroom-level effect of instructional practices on classroom-level component of student outcomes.

In total, seven mediation models were tested: four models for middle school students and three models for elementary school students. Figure 1 depicts the basic form of the mediation models. Mediation is implied if instructional practices are significantly related to the mediator (student perceptions) and student perceptions are significantly associated with the outcome (student achievement), which is accomplished by multiplying those paths and seeing if they are significant (representing the indirect effect). Path multiplication was done directly in the *Mplus 7.2* (Muthén & Muthén, 1998-2012) software. In the MSEM approach, multiple indicators are used to measure the latent variables of student perceptions, observed emotional support, classroom organization, and instructional support. All models were specified with random intercepts and fixed slopes. Though methods testing mediation call for using bootstrapped confidence intervals (Preacher & Hayes, 2008), I was unable to do so because of constraints with the software.[7]

Results

[7] *Mplus 7.2* (Muthén & Muthén, 1998-2012) does not allow for bootstrapped confidence intervals in an MSEM framework.

I begin by examining the descriptive statistics and correlations between the variables of interest. Means and bivariate correlations of variables are depicted in Tables 3 and 4 and were conducted separately for middle and elementary school students. The dimension of observed classroom organization had the highest means (5.16 in middle school, and 5.41 in elementary school on a 7-point scale). In contrast, observed instructional support had a mean of 3.29 in middle school and 3.70 in elementary school, and observed emotional support had the lowest means (3.20 in middle school and 3.41 in elementary school). The three dimensions of the CLASS were moderately to highly positively correlated with one another (*r* ranged from .33 and .71 in middle school and .35 and .69 in elementary school). Correlations between the observed dimension of emotional support and the classroom-level student perception of emotional support were .15 in the middle school sample and .04 in the elementary school sample. It is important to note correlations between dimensions of student perceptions and dimensions of observed instruction were low or non-statistically significant among elementary school students whereas were statistically significantly correlated among middle school students. Tables 5 and 6 provide the unstandardized coefficients and *p*-values for all the variables in the model for middle school and elementary school students separately and show the association between level-1 and level-2 variables and mathematics achievement. These tables are not further discussed, instead I present results in the form of path diagrams.

Multilevel mediation models

Below, results for the mediation models are discussed separately for the elementary school and middle school samples.

Middle school sample. The following sections are labeled by the hypothesized mediating student perceptions.

Interest. Figure 2 depicts the mediated path for interest. The model fit was good (Chi-square 590.44, *df* 201, $p < .001$; RMSEA = .03; CFI = .94; TLI = .93; $SRMR_W = .02$ and $SRMR_B = 0.13$). The within effect of students' perceptions of interest and mathematics achievement was 0.05 ($p = .001$; 95% CI [.03, .07]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. This can be interpreted a one-point increase in an individual's self-reported interest in the mathematics classroom is associated with a .05 standardized increase in an individual's standardized achievement controlling for students' previous mathematics achievement, and demographic characteristics. The between effect from students' perceptions of interest and achievement was 0.08 ($p = .06$; 95% CI [0.01, 0.15]). The paths from observed instructional practices to average classroom interest were .45 ($p = .03$; 95% CI [.12, .79]) for emotional support, .11 ($p = .11$; 95% CI [-.01, .23]), for classroom organization and -.35 ($p = .17$; 95% CI [.79, .08]) for instructional support. The correlation between observed classroom organization and emotional support was .18 ($p = .001$), between observed classroom organization and instructional support was .13 ($p = .001$), and between instructional support and emotional support was .36 ($p < .001$). The indirect effect of observed classroom practices and student achievement through interest was not significant for each of the tested paths. The direct paths from instructional practice to student achievement were not significant accounting for the indirect effect through interest ($b = -.03$, $p = .75$; $b = .04$, $p = .21$, and $b = .04$, $p = .69$, for observed emotional support, classroom organization, and instructional support, respectively)

suggesting that student perceptions account for the association found in the previous chapter.

Emotional Support. Figure 3 depicts the mediated path for affect. Model fit was acceptable using certain fit indices (Chi-square 1172.25, *df* 203, and $p < .001$ (RMSEA = .04; CFI = .85; TLI = .82; $SRMR_W = .04$ and $SRMR_B = 0.14$). The within effect of students' perceptions of emotional support and mathematics achievement was 0.02 ($p = .06$; 95% CI [.00, .03]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. The between effect from students' perceptions of emotional support and achievement was 0.06 ($p = .19$; 95% CI [-0.02, 0.15]). The paths from observed instructional practices to average classroom emotional support were .52 ($p = .02$; 95% CI [.15, .88]) for emotional support, .13 ($p = .09$; 95% CI [.00, .26]), for classroom organization and -.42 ($p = .15$; 95% CI [-.90, .07]) for instructional support. The correlation between classroom organization and emotional support was .18 ($p = .001$), between classroom organization and instructional support was .13 (p = .001), and between instructional support and emotional support was .36 ($p < .001$). The indirect effect of observed classroom practices and student achievement through emotional support were not statistically significant. The direct paths from instructional practice to student achievement were not significant after accounting for perceptions of emotional support ($b = -.02$, $p = .83$; $b = .04$, $p = .21$, and $b = .04$, $p = .74$, for observed emotional support, classroom organization, and instructional support respectively).

Instructional Support. Figure 4 depicts the path diagram for the mediation model. Model fit was good (Chi-square 1448.20, *df* 505, and $p < .001$; RMSEA = .02; CFI = .95;

TLI = .94; $SRMR_W$ = .02 and $SRMR_B$ = 0.11). The within effect of students' perceptions of instructional support on mathematics achievement was 0.05 (p = .001; 95% CI [.02, .07]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. The between effect from students' perceptions of instructional support and achievement was 0.19 (p = .001; 95% CI [0.10, 0.29]). The paths from observed instructional practices to average classroom instructional support were .35 (p = .03; 95% CI [.08, .63]) for emotional support, .11 (p = .74; 95% CI [-.44, .66]), for classroom organization and -.21 (p = .42; 95% CI [-.65, .22]) for instructional support. The correlation between classroom organization and emotional support was .18 (p = .001), between classroom organization and instructional support was .13 (p = .001), and between instructional support and emotional support was .36 (p < .001). The indirect effect of observed classroom practices and student achievement through students' perceive instructional practices were not statistically significant. The direct paths from instructional practice to student achievement were not significant after accounting for the indirect effect through perceived instructional support (b = -.04, p =. 59; b =.03, p =.35, and b = .05, p = .66, for observed emotional support, classroom organization, and instructional support respectively).

Academic Press. Figure 5 depicts the path diagram for the mediation model. Model fit was good (Chi-square 659.93, *df* 232, p <.001, RMSEA = .02; CFI = .92; TLI = .90; $SRMR_W$ = .02 and $SRMR_B$ = 0.14). The within effect of students' perceptions of academic press on mathematics achievement was 0.02 (p = .20; 95% CI [-.01, .06]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. The between effect from students' perceptions of academic

press on achievement was 0.37 ($p < .001$; 95% CI [.21, .53]). The paths from observed instructional practices to average classroom academic press were .26 ($p = .06$; 95% CI [.21, .53]) for emotional support, .08 ($p = .07$; 95% CI [.01, .15]), for classroom organization and -.13 ($p = .44$; 95% CI [-.42, .15]) for instructional support. The correlation between classroom organization and emotional support was .18 ($p = .001$), between classroom organization and instructional support was .13 ($p = .001$), and between instructional support and emotional support was .36 ($p < .001$). The correlation between classroom organization and emotional support was .18 ($p = .001$), between classroom organization and instructional support was .13 ($p = .001$), and between instructional support and emotional support was .36 ($p < .001$). The indirect effect of observed classroom practices and student achievement through students' perceptions of academic press were not significant. The direct paths from instructional practice to student achievement were not significant after accounting for the indirect effect through perceived teacher expectations ($b = -.06$, $p = .39$; $b = .03$, $p = .42$, and $b = .05$, $p = .62$, for observed emotional support, classroom organization, and instructional support respectively).

Elementary school sample.

Emotional Support. Figure 6 depicts the path diagram for the mediation model. Model fit was acceptable (Chi-square 1252.79, *df* 249, $p < .001$, RMSEA = .04; CFI = .91; TLI = .89; $SRMR_W = .04$ and $SRMR_B = 0.10$). The within effect of students' perceptions of emotional support and mathematics achievement was 0.04 ($p = .15$; 95% CI [-.01, .07]) controlling for students' previous mathematics achievement, English Language learner statues, gender, and ethnicity. The between effect from students'

perceptions of emotional support and achievement is 0.08 (p = .35; 95% CI [-0.06, 0.23]). The paths from observed instructional practices to average classroom emotional support were .15 (p = .11; 95% CI [.00, .30]) for emotional support, .02 (p = .63; 95% CI [-.05, .10]), for classroom organization and -.14 (p = .11; 95% CI [-.27, .00]) for instructional support. The correlation between classroom organization and emotional support was .20 (p = .001), between classroom organization and instructional support was .20 (p = .001), and between instructional support and emotional support was .33 (p < .001). The indirect effects of observed classroom practices and student achievement through emotional support were not statistically significant. The direct paths from instructional practice and student achievement were not significant after accounting for the indirect effect through emotional support (b = -.04, p =.67; b =.13, p =.19, and b =.02, p =.85, for emotional support, classroom organization, and instructional support respectively).

Instructional Support. Figure 7 depicts the path diagram for the mediation model. Model fit was not idea (Chi-square 3373.15, *df* 530, and p < .001, RMSEA = .04; CFI = .82; TLI = .80; $SRMR_W$ = .09 and $SRMR_B$ = 0.17). The within effect of students' perceptions of instructional support and mathematics achievement was 0.01 (p =.77; 95% CI [-.03, .05]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. The between effect from students' perceptions of instructional support and achievement is 0.15 (p = .20; 95% CI [-0.04, 0.36]). The paths from observed instructional practices to average classroom instructional support were .13 (p < .12; 95% CI [-.01, .27]) for emotional support, .03 (p = .44; 95% CI [-.04, .10]), for classroom organization and -.16 (p = .04; 95% CI [-.29, -.03]) for instructional support. The correlation between classroom organization and emotional

support was .20 (p = .001), between classroom organization and instructional support was .20 (p = .001), and between instructional support and emotional support was .33 (p < .001). The indirect effect of observed classroom practices and student achievement through instructional support was not statistically significant. The direct paths from instructional practice and student achievement were not significant after accounting for the indirect effect through instructional support (b = -.04, p = .67; b=.13, p =.28, and b =.02, p =.83, for emotional support, classroom organization, and instructional support respectively).

Academic Press. Figure 8 depicts the path diagram for the mediation model. Model fit was acceptable (Chi-square 807.35, *df* 217, p < .001, RMSEA = .03; CFI = .90; TLI = .88; $SRMR_W$ = .02 and $SRMR_B$ = 0.11). The within effect of students' perceptions of academic press and mathematics achievement was 0.08 (p = .04; 95% CI [.02, .14]) controlling for students' previous mathematics achievement, English Language learner status, gender, and ethnicity. The between effect from students' perceptions of academic press and achievement is 0.22 (p = .20; 95% CI [-.06, .50]). The paths from observed instructional practices to average classroom academic press were .14 (p < .06; 95% CI [.02, .26]) for emotional support, .06 (p = .17; 95% CI [-.01, .13]), for classroom organization and -.15 (p = .02; 95% CI [-.15, -.02]) for instructional support. The correlation between classroom organization and emotional support was .20 (p = .001), between classroom organization and instructional support was .20 (p = .001), and between instructional support and emotional support was .33 (p < .001). The indirect effect of observed classroom practices and student achievement through academic press were not significant. The direct paths from instructional practice and student achievement

were not significant after accounting for the indirect effect through teacher expectations (b = -.05, p = .61; b =.12, p =.13, and b =.03, p =.76, for emotional support, classroom organization, and instructional support respectively).

Discussion

The present study proposed and tested a model for understanding whether student perceptions of the classroom (i.e., interest, emotional support, instructional support, and academic press) mediated the association of observed instructional practices (i.e., emotional support, classroom organization, instructional support) on student mathematics achievement. Though I did not find statistically significant mediation, I did find that student perceptions accounted for the paths from observed instruction to student achievement. I also found that different aspects of the classroom were related to different dimensions of student perceptions. Specifically, observed emotional support was significantly associated with student perceptions of interest, emotional support and instructional support in the middle school sample. In the elementary school sample, observed instructional support was significantly negatively related to students' perceptions of academic press and instructional support.

By understanding what aspects of instruction students attend to, researchers and practitioners can design learning environments with those constructs in mind. In the present study, I found that observed dimensions did not always align with what students perceived in the classroom, and in one instance, was significantly inversely aligned. Specifically, elementary school students' perceptions of instructional support were significantly negatively associated with observer's ratings of instructional support. This could mean that elementary school students' appraisals of what is observed to be high

quality instructional support might be interpreted as controlling behavior (Flink, Boggiano, & Barrett, 1990) and that there might be instances where a teacher's intentions or behaviors do not align with students' perceptions of it (Weinstein, 1982). Future research using cognitive interviewing techniques (Koskey, Karabenick, Woolley, Bonney, & Dever, 2010) could reveal the different interpretations students of different age groups have of what researchers might interpret as instructionally supportive.

There are several explanations as to why I did not find significant mediation. It could be an issue of measurement. Not all hypothesized meditational constructs that were tested were directly aligned with the dimensions of the CLASS (emotional support, classroom organization, and instructional support), and those that were did not have items that were directly aligned with the observed indicators. An important way to move forward in incorporating student perceptions into the discussion of effective teaching practices would be to design measures from multiple raters (trained observers, teachers, and students) that are directly aligned with one another. I am not aware of any studies that have used observation measures that are directly aligned with measures of student perceptions with the exception of the use of the CLASS-Student Report (Downer et al., 2014). However, their work only represents one way to conceptualize classroom instruction and student perceptions and I would be interested to see more measures created that ask students to respond to specific moments during instruction.

Another explanation for not finding significant mediation is that other stronger mediators such as student engagement might explain more of the association between instructional practices and student achievement. Studies have identified this link as measured by self-reported (Reyes, Brackett, Rivers, White, & Salovey, 2012; Skinner &

Belmont, 1993) or directly observed engagement (Ing, Webb, Franke, Turrou, Wong, Shin, & Fernandez, 2014). Specifically, Ing and colleagues (2014) used observations of student participation (conceptualized as explanations and engagement with others' ideas) as mediators of the association between teacher support of student participation and student achievement. Though the sample size was small (71 students in six classrooms), the authors still found that students' participation explained the association between teacher support of student participation and student achievement. While Skinner and Belmont (1993) provide the most compelling evidence for this meditational model future research should replicate and extend this work by looking at more than two time points during the year and observations occurring over a shorter time frame to more carefully model the reciprocal process.

Though I could not directly test for developmental differences in the elementary and middle school sample because of differences in the student perception surveys that were administered, there are still some interesting points to mention. Whereas observed emotional support was significantly related to middle school students' perceptions of interest and emotional support, observer's ratings of instructional support were inversely related to elementary school students' perceptions of academic press and instructional support. This could indicate developmental differences in students' perceptions. Previous research has shown mean-level differences in students' perceptions of different grades (Schenke, Lam, Conley, & Karabenick, 2015), however research in this area is still scant. Answering this question might require more research to understand how these perceptions emerge. Preliminary evidence from a sample of seventh grade students indicated that students' experiences in their previous classrooms were consistent

predictors of their perceptions of the current classroom (Schenke, Lam, Ruzek, Conley, Karabenick, & Conley, 2015). Although in the present study I could not test for developmental differences in how strongly students' experiences in their previous classrooms or personality characteristics influence their perceptions by age, it might be that as students get older, they take in more cues from their environment and might even have perceptions that are more aligned with what is observed such as found in the present study. This further provides evidence that student perceptions should be treated as a psychological construct (Maehr & Midgley, 1991) rather than as interraters of the classroom.

Limitations

Several limitations underlie this work. First, the timeframes of the observations of instructional practices and when students took the survey varied greatly. Though the observation always took place before students responded to the survey, this large timeframe might have limited the strength of the association found in this paper. Additionally, the correlations between constructs of student perceptions (interest, emotional support, instructional support, and academic press) suggest a halo effect. This has been noted in other studies using data from student perceptions (e.g., Schenke et al., 2015) and it is still unclear if students perceive these constructs as similar to each other or if classrooms that are high in emotional support are also high in interest.

Conclusion

Understanding the links between observed instructional practices and student perceptions could unlock a key component in understanding how instruction "gets" into students and therefore support better opportunities for the design of professional

development and teacher practice. While this paper failed to provide definitive evidence of student perceptions mediating the association of observed instruction and student achievement, student perceptions did account for the relationship. Student perceptions have long been used in educational psychology and I hope this paper illuminates the importance of them in discussions of educational policy and models of teacher effectiveness.

References

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

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260-267.

Bandura, A. (1986) *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Berliner, D. C. (1976). Impediments to the study of teacher effectiveness. *Journal of Teacher Education, 27,* 5-13.

Bill & Melinda Gates Foundation. (2013). *Measures of Effective Teaching (MET) project*: *Final report*. Seattle, WA: Author.

Bitler, M., Corcoran, S.P., Domina, T., & Penner, E. (*in preparation*). Teacher effects on student achievement and height: A cautionary tail.

Brattesani, K. A., Weinstein, R. S., & Marshall, H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology*, *76*(2), 236-247.

Dorman, J. P. (2008). The effect of clustering on statistical tests: an illustration using classroom environment data. *Educational Psychology*, *28*(5), 583–595. doi:10.1080/01443410801954201

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2014). Measuring effective teacher-student interactions from a student perspective a multi-level analysis. *The Journal of Early Adolescence*, vol page

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meese, J. L. & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), *Achievement and achievement motives* (pp.75-146). San Francisco, CA: Freeman.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1-9.

Feldlaufer, H., Midgley, C., & Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high school. *The Journal of Early Adolescence*, *8*(2), 133-156.

Ferguson, R. (2010, October 14). *Student perceptions of teaching effectiveness*. Discussion brief from the National Center for Teacher Effectiveness and the Achievement Gap Initiative, Harvard University, Cambridge, MA.

Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology*, *59*(5), 916.

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments as developmental processes: Conceptualization and measurement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of Research on Schools, Schooling and Human Development* (pp. 25-41). New York: Routledge.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*(3), 794-831.

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.

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

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2014). How student participation mediates the relationship between teacher practices and student achievement. Paper presented at the annual meeting of the American Educational Research, Philadelphia, PA.

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*.

Kaplan, A., Middleton, M., Urdan, T., & Midgley, C. (2002). Achievement goals and goal structures. In C. Midgley (Ed.), *Goals, goal structures, and patterns of adaptive learning* (pp. 21-53). Mahwah, NJ: Erlbaum.

Karabenick, S. A. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology, 96*(3), 569-581.

Koskey, K. L., Karabenick, S. A., Woolley, M. E., Bonney, C. R., & Dever, B. V. (2010). Cognitive validity of students' self-reports of classroom mastery goal structure: What students are thinking and why it matters. *Contemporary Educational Psychology*, *35*(4), 254–263.

Kunter, M., Tsai, Y. M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction*, *18*(5), 468-482.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249-277.

Maehr, M. L., & Midgley, C. (1991). Enhancing student motivation: A school-wide approach. *Educational Psychologist, 26,* 399–427.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106-124.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732-749.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology*, *57*(1), 487-503.

Middleton, M. J., & Midgley, C. (2002). Beyond motivation: Middle school students' perceptions of press for understanding in math. *Contemporary Educational Psychology*, *27*(3), 373-391.

Miller, A. D., & Murdock, T. B. (2007). Modeling latent true scores to determine the utility of aggregate student perceptions as classroom indicators in HLM: The case of classroom goal structures. *Contemporary Educational Psychology*, *32*(1), 83-104.

Moos, R. H. (1979). *Evaluating educational environments*. San Francisco: Jossey-Bass.

Murray, H. A. (1938). Explorations in personality. New York: Oxford University Press.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Patrick, H., Anderman, L. H., Ryan, A. M., Edelin, K. C., & Midgley, C. (2001). Teachers' communication of goal orientations in four fifth-grade classrooms. *The Elementary School Journal*, 102(1), 35-58.

Pianta, R.C., Hamre, B.,Hayes, N.,Mintz, S.& LaParo, K.M.(2008). Classroom Assessment Scoring System—Secondary (CLASS-S). Charlottesville, VA: University of Virginia.

Pianta, R.C., La Paro, K.M., & Hamre, B. (2007). Classroom Assessment Scoring System—CLASS. Baltimore, M.D.: Brookes.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209-233.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*(3), 700-712.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*(4), 958.

Rosenshine, B. (1970). Evaluation of classroom instruction. *Review of Educational Research*, 40(2), 279-300.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (*under review*). How teacher emotional support motivates students: The mediation roles of peer relatedness and autonomy support.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68-78.

Schenke, K., Lam, A.C., Conley, A.M., & Karabenick, S.K. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology, 41*, 133-146.

Schenke, K., Lam, A.C., Ruzek, E., Conley, A.M., Karabenick, S.A., & Eccles, J.S. (2015). Heterogeneity of student perceptions of the classroom climate: A latent profile approach. Paper presented at the annual meeting of the American Educational Research, Chicago, IL.

Shouse, R. C. (1996). Academic press and sense of community: Conflict, congruence, and implications for student achievement. *Social Psychology of Education*, *1*(1), 47-68.

Simon, A., & Boyer, E. G. (1974). Mirrors for Behavior III. An Anthology of Observation Instruments. Communication Materials Center, 110 Rices Mill Road, Wyncote, Pennsylvania 19095 (Library of Congress Catalog Card Number 67-31735, $19.95).

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), 571-581.

Stein, M.K., Engle, R.A., Smith, M.S., & Hughes, E.K. (2008). Orchestrating productive mathematical discussions: Helping teachers learn to better incorporate student thinking. *Mathematical Thinking and Learning, 10*(4), 313-340.

Trickett, E. J., & Moos, R. H. (1973). Social environment of junior high and high school classrooms. *Journal of Educational Psychology*, *65*(1), 93-102.

Turner, J. C., & Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist*, *35*(2), 69-85.

Turner, J. C., Midgley, C., Meyer, D. K., Gheen, M., Anderman, E. M., Kang, Y., & Patrick, H. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. *Journal of Educational Psychology*, *94*(1), 88-106.

Urdan, T. (2010). The challenges and promise of research on classroom goal structure. In J. L.Meece & J. S. Eccles (Eds.), *Handbook of Research on Schools, Schooling and Human Development* (pp. 92-108), 1st Edition. New York: Routledge.

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, *28*, 1–11.

Wallace, T.L., Kelcey, B., & Ruzek, E. (*under review*). Dimensionality and predictive validity of a student perception survey: What do adolescent perceptions of a classroom experience tell us?

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573.

Weinstein, R. S. (1982, February). Student perceptions of schooling. Paper presented at the National Invitational Conference, "Research on Teaching: Implications for Practice". Warrenton, VA.

Wittrock, M. C.(1986). Teacher behavior and student achievement. In Wittrock, M.C. (Ed.) *Handbook of research on teaching* (pp. 297-315). London: Collier Macmillan Publishers.

Table 4.1

Standardized factor loadings of student survey items for the middle school sample (n=3,190)

Student survey items	λ	R sq
Interest (α=0.82)		
This class does not keep my attention—I get bored.	0.61	0.37
My teacher makes learning enjoyable.	0.87	0.75
My teacher makes lessons interesting.	0.88	0.76
Emotional support (α=0.81)		
My teacher really tries to understand how students feel about things.	0.78	0.61
My teacher respects my ideas and suggestions.	0.75	0.57
My teacher in this class makes me feel that s/he really cares about me.	0.78	0.61
Academic press (α=0.68)		
In this class, my teacher accepts nothing less than our full effort.	0.58	0.33
My teacher doesn't let people give up when the work gets hard.	0.67	0.43
My teacher wants us to use our thinking skills, not just memorize things.	0.61	0.37
My teacher wants me to explain my answers—why I think what I think.	0.54	0.29
Instructional support (α=0.91)		
If you don't understand something, my teacher explains it another way.	0.71	0.51
My teacher has several good ways to explain each topic that we cover in this class.	0.76	0.57
My teacher explains difficult things clearly.	0.73	0.53
My teacher asks questions to be sure we are following along when s/he is teaching.	0.56	0.32
In this class, we learn to correct our mistakes.	0.70	0.48
My teacher gives us time to explain our ideas.	0.69	0.48
My teacher takes the time to summarize what we learn each day.	0.59	0.34
My teacher checks to make sure we understand what s/he is teaching us.	0.74	0.55
The comments that I get on my work in this class help me understand how to improve.	0.70	0.49
We get helpful comments to help us know what we did wrong on assignments	0.66	0.44
This class makes me a better thinker	0.67	0.44

Note. Negatively worded items were reverse coded before entered into the model.

Table 4.2

Standardized factor loadings of student survey items for the elementary school sample (n= 3,507)

Item	λ	R sq
Emotional support (α=0.83)		
My teacher is nice to me when I ask questions.	0.60	0.36
My teacher wants us to share our thoughts.	0.45	0.2
I like the way my teacher treats me when I need help.	0.74	0.54
If I am sad or angry, my teacher helps me feel better.	0.74	0.54
My teacher in this class makes me feel that s/he really cares about me.	0.78	0.61
My teacher knows when the class understands, and when we do not.	0.52	0.27
My teacher seems to know if something is bothering me.	0.66	0.43
Academic press (α=0.55)		
In this class, my teacher accepts nothing less than our full effort.	0.46	0.21
My teacher pushes everyone to work hard.	0.42	0.17
The teacher in this class encourages me to do my best.	0.57	0.33
My teacher wants me to explain my answers—why I think what I think.	0.52	0.27
Instructional support (α=0.85)		
If you don't understand something, my teacher explains it another way.	0.54	0.29
My teacher has several good ways to explain each topic that we cover in this class.	0.62	0.39
My teacher explains difficult things clearly.	0.54	0.29
My teacher explains things in very orderly ways.	0.60	0.36
My teacher gives us time to explain our ideas.	0.61	0.38
My teacher takes the time to summarize what we learn each day.	0.57	0.32
My teacher checks to make sure we understand what s/he is teaching us.	0.66	0.44
When my teacher marks my work, he/she writes on my papers to help me understand how to do better.	0.54	0.3
When he/she is teaching us, my teacher asks us whether we understand.	0.56	0.31
My teacher asks questions to be sure we are following along when he/she is teaching.	0.59	0.35
In this class, we learn a lot almost every day.	0.56	0.32
I understand what I am supposed to be learning in this class.	0.47	0.22

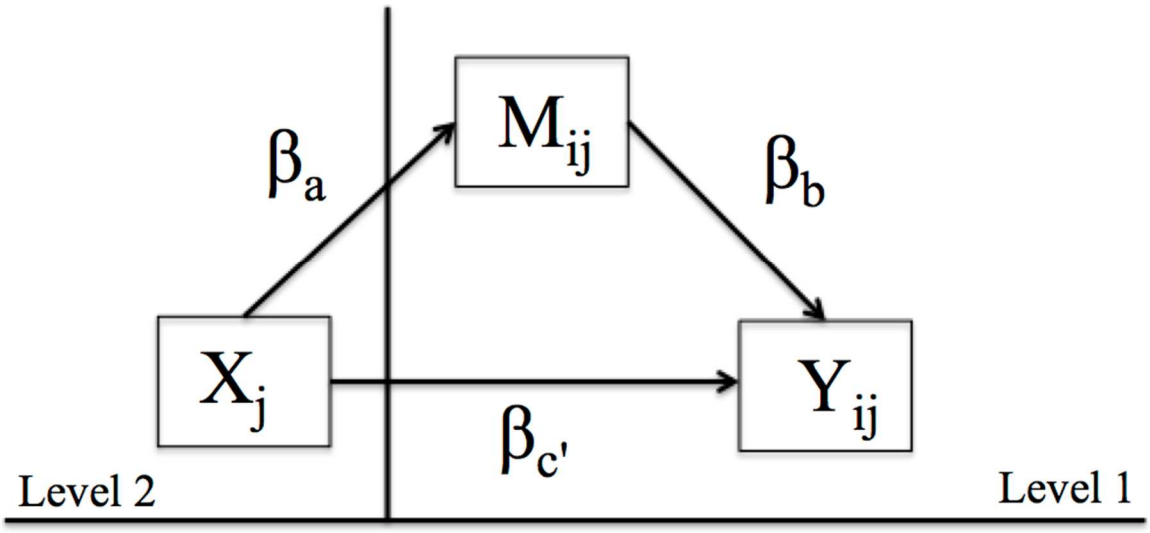


Figure 4.1. Multilevel mediation model. From Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249-277.

Table 4.3

Means, standard deviations, and correlations for middle school sample (n=3,190)

	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. L2 Emotional support	3.20 (0.56)	1.00															
2. L2 Classroom organization	5.16 (0.62)	**0.33**	1.00														
3. L2 Instructional support	3.29 (0.76)	**0.71**	**0.50**	1.00													
4. L1 Affect	3.72 (0.95)	**0.18**	**0.16**	**0.14**	1.00												
5. L1 Emotional support	3.78 (0.89)	**0.15**	**0.14**	**0.11**	**0.72**	1.00											
6. L1 Instructional support	3.98 (0.77)	**0.17**	**0.17**	**0.16**	**0.76**	**0.80**	1.00										
7. L1 Teacher expectation	4.17 (0.72)	**0.16**	**0.15**	**0.15**	**0.55**	**0.63**	**0.72**	1.00									
8. T2 Math achievement	0.28 (0.89)	**0.13**	**0.09**	**0.17**	**0.07**	**0.07**	**0.06**	**0.09**	1.00								
9. T1 Math achievement	0.20 (0.88)	**0.12**	**0.05**	**0.16**	0.00	0.02	-0.01	**0.05**	**0.80**	1.00							
10. Male	50%	0.01	-0.00	0.00	0.01	**-0.04**	-0.02	**-0.06**	0.01	0.02	1.00						
11. English language learner	13%	**-0.11**	**-0.06**	**-0.09**	0.02	0.03	0.02	**-0.05**	**-0.23**	**-0.24**	0.03	1.00					
12. Free or reduced lunch	61%	**-0.13**	**-0.16**	**-0.22**	-0.02	-0.01	-0.03	**-0.04**	**-0.28**	**-0.27**	-0.01	**0.23**	1.00				
13. White	27%	**0.08**	**0.08**	**0.11**	-0.02	-0.01	-0.02	**0.04**	**0.19**	**0.20**	0.03	**-0.19**	**-0.33**	1.00			
14. Hispanic	36%	**-0.07**	**-0.09**	**-0.12**	-0.02	0.00	-0.02	**-0.06**	**-0.13**	**-0.10**	-0.01	**0.36**	**0.35**	**-0.46**	1.00		
15. Black	23%	**-0.07**	**-0.04**	**-0.07**	0.01	-0.03	0.01	0.00	**-0.25**	**-0.25**	-0.03	**-0.19**	-0.01	**-0.34**	**-0.41**	1.00	
16. Other ethnicity	14%	**0.08**	**0.07**	**0.10**	**0.04**	**0.04**	**0.04**	0.02	**0.25**	**0.20**	0.00	-0.03	**-0.08**	**-0.24**	**-0.30**	**-0.22**	1.00
17. 6th grade	51%	**0.32**	**0.21**	**0.39**	**0.13**	**0.11**	**0.13**	**0.11**	0.02	0.02	0.01	**-0.07**	**-0.11**	0.03	-0.00	**-0.04**	0.02
18. 7th grade	43%	**-0.22**	**-0.11**	**-0.24**	**-0.13**	**-0.09**	**-0.11**	**-0.08**	0.03	**0.04**	-0.01	**0.04**	0.03	0.04	**-0.07**	**0.04**	-0.01
19. 8th grade	6%	**-0.21**	**-0.20**	**-0.30**	-0.00	**-0.04**	**-0.04**	**-0.06**	**-0.10**	**-0.11**	0.00	**0.05**	**0.14**	**-0.13**	**0.14**	0.01	-0.03

Note. Significant correlations are bolded.

Table 4.4

Means, standard deviations, and correlations for elementary school sample (n=3,269)

	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. L2 Emotional support	3.41 (.50)	1.00															
2. L2 Classroom organization	5.41 (.59)	**0.35**	1.00														
3. L2 Instructional support	3.70 (.69)	**0.69**	**0.52**	1.00													
4. L1 Emotional support	4.11 (.74)	**0.04**	0.01	-0.01	1.00												
5. L1 Instructional support	4.21 (.55)	0.00	0.01	**-0.04**	**0.75**	1.00											
6. L1 Teacher expectation	4.41 (.59)	0.02	0.02	-0.02	**0.54**	**0.60**	1.00										
7. T2 Math achievement	.13 (.91)	**0.05**	**0.09**	**0.08**	**0.06**	0.01	**0.07**	1.00									
8. T1 Math achievement	.09 (.92)	**0.04**	0.03	**0.07**	**0.04**	-0.01	**0.04**	**0.81**	1.00								
9. Male	49%	0.00	-0.01	0.03	**-0.11**	-0.08	**-0.08**	0.01	0.03	1.00							
10. LL	14%	**0.05**	**0.05**	**0.07**	0.02	0.02	**-0.04**	**-0.12**	**-0.15**	0.00	1.00						
11. Free or reduced lunch	51%	0.02	-0.02	-0.01	0.01	0.04	-0.01	**-0.14**	**-0.16**	0.01	**0.22**	1.00					
12. White	28%	**0.09**	**0.11**	**0.11**	**-0.04**	**-0.13**	**-0.05**	**0.23**	**0.26**	0.01	**-0.18**	**-0.22**	1.00				
13. Hispanic	27%	**0.07**	0.00	**0.06**	**0.05**	**0.04**	-0.02	**-0.04**	**-0.06**	0.00	**0.42**	**0.31**	**-0.38**	1.00			
14. Black	35%	**-0.15**	**-0.09**	**-0.17**	-0.02	**0.08**	**0.08**	**-0.24**	**-0.24**	-0.01	**-0.26**	**-0.10**	**-0.46**	**-0.45**	1.00		
15. Other ethnicity	10%	0.01	-0.01	0.03	0.02	0.01	-0.01	**0.11**	**0.09**	0.01	0.02	0.00	**-0.20**	**-0.20**	**-0.24**	1.00	
16. 4th grade	45%	0.01	-0.03	-0.03	**0.07**	**0.04**	**-0.05**	**-0.07**	**-0.07**	0.00	0.05	0.03	0.00	0.02	-0.01	-0.02	1.00
17. 5th grade	55%	-0.01	0.03	0.03	**-0.07**	**-0.04**	**0.05**	**0.07**	**0.07**	-0.00	-0.05	-0.03	-0.00	-0.02	0.01	0.02	-1.00

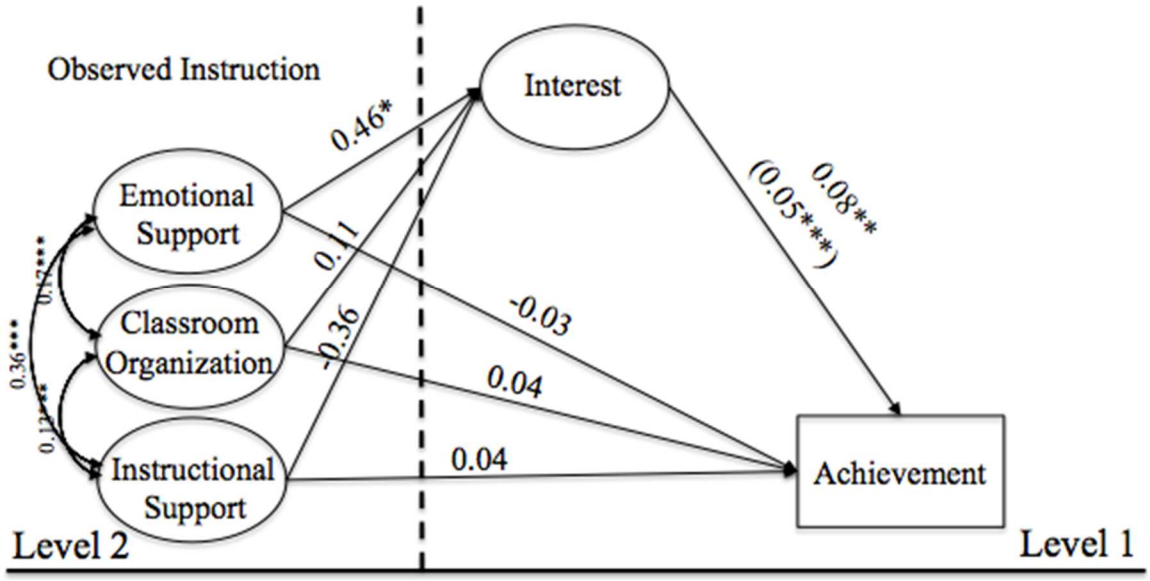


Figure 4.2. Mediation model with middle school students' perceptions of interest (3,190 students in 190 classrooms). * $p < .05$. ** $p < .01$. *** $p < .001$. Within effects are depicted in parentheses.

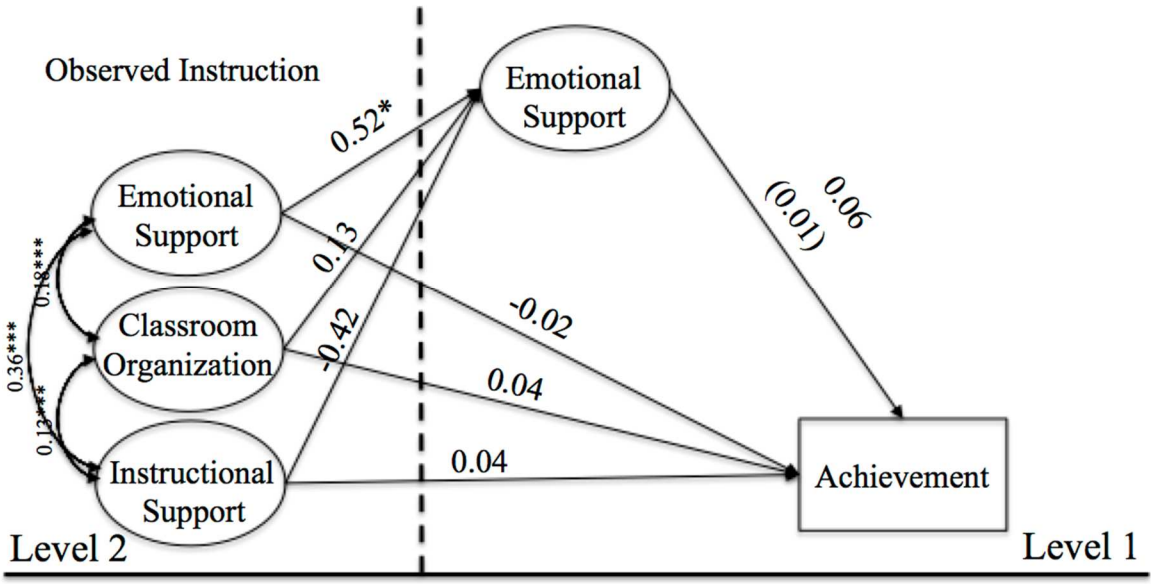


Figure 4.3. Mediation model with middle school students' perceptions of emotional support (3,190 students in 190 classrooms). * p < .05. ** p < .01. *** p < .001. Within effects are depicted in parentheses.

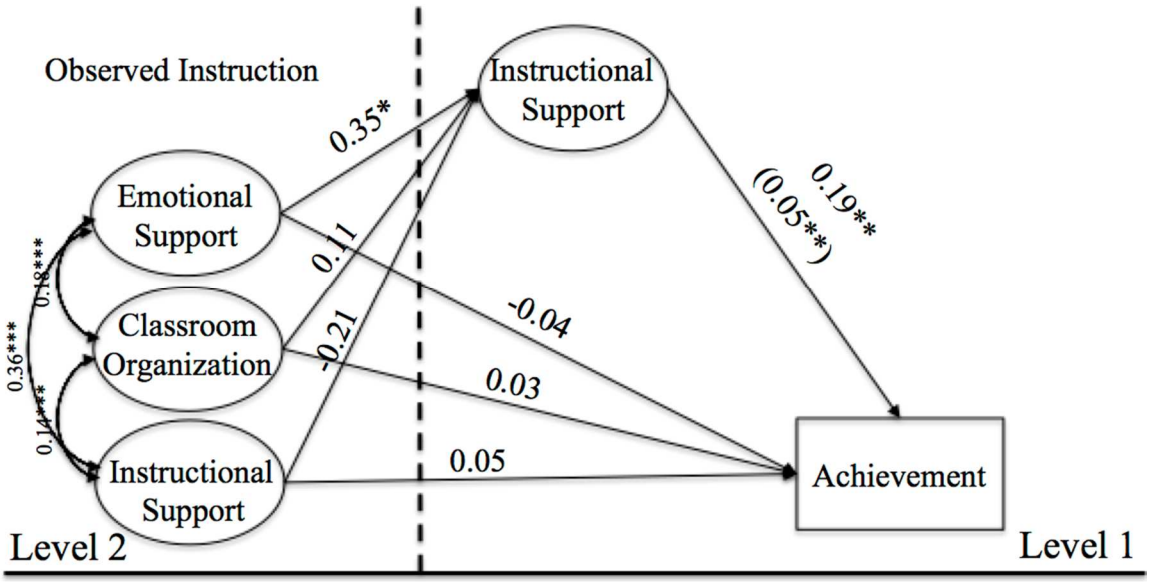


Figure 4.4. Mediation model with middle school students' perceptions of instructional support (3,190 students in 190 classrooms). * $p < .05$. ** $p < .01$. *** $p < .001$. Within effects are depicted in parentheses.

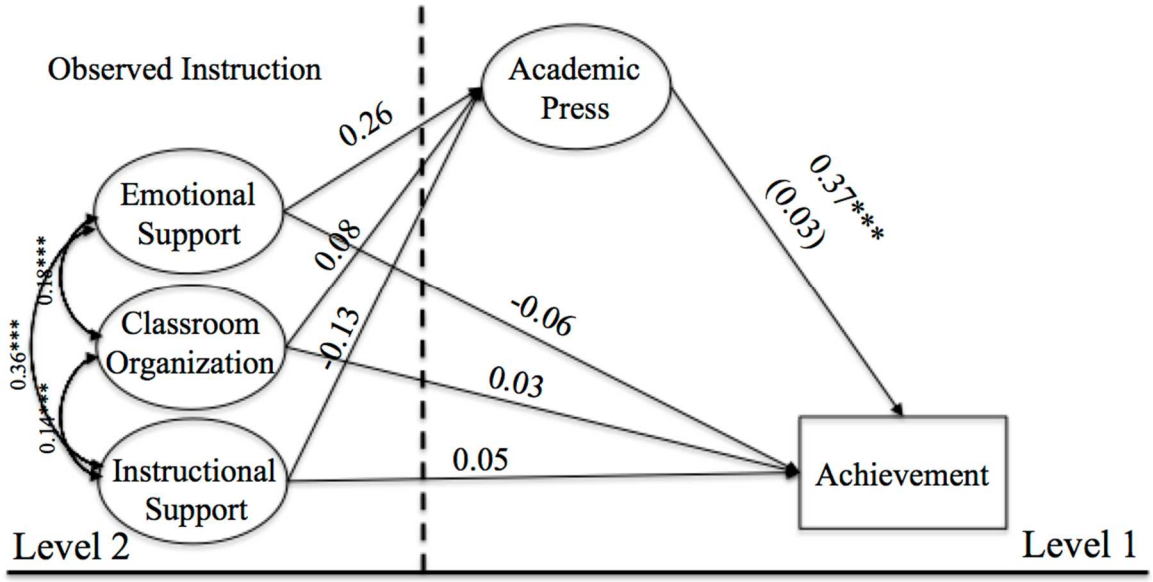


Figure 4.5. Mediation model with middle school students' perceptions of academic press (3,190 students in 190 classrooms). * $p < .05$. ** $p < .01$. *** $p < .001$. Within effects are depicted in parentheses.

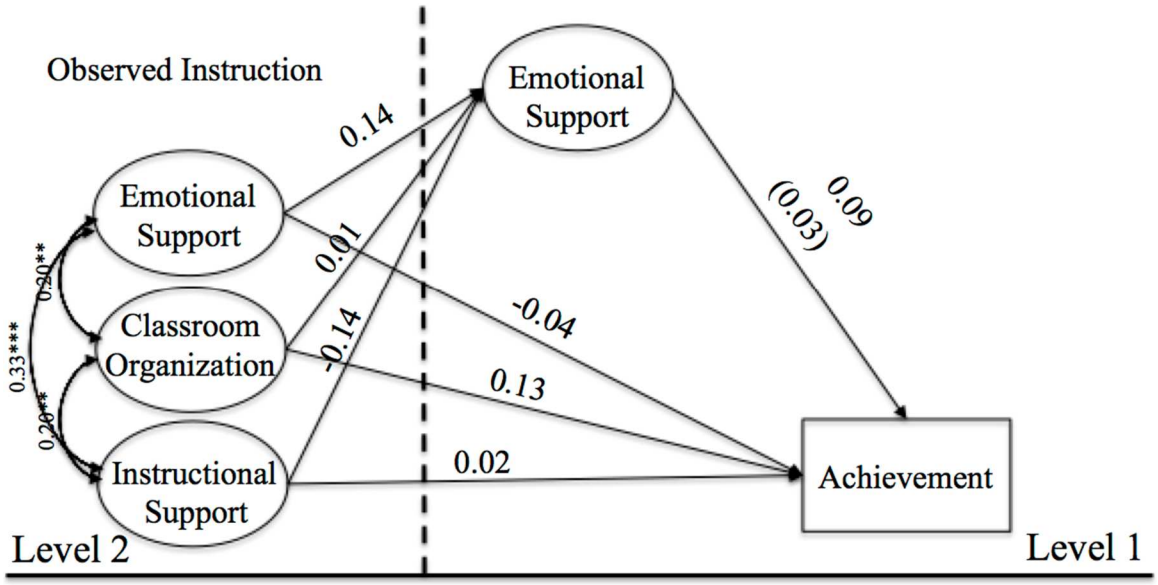


Figure 4.6. Mediation model with elementary school students' perceptions of emotional support (3,269 students in 322 classrooms). * p < .05. ** p < .01. *** p < .001. Within effects are depicted in parentheses.

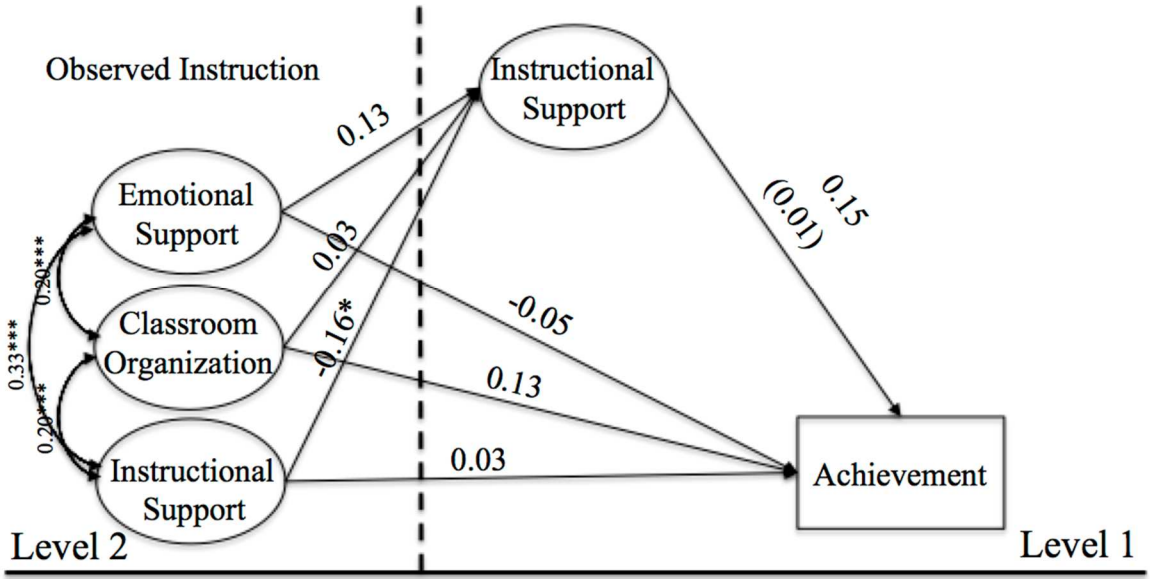


Figure 4.7. Mediation model with elementary school students' perceptions of instructional support (3,269 students in 322 classrooms). * p < .05. ** p < .01. *** p < .001. Within effects are depicted in parentheses.

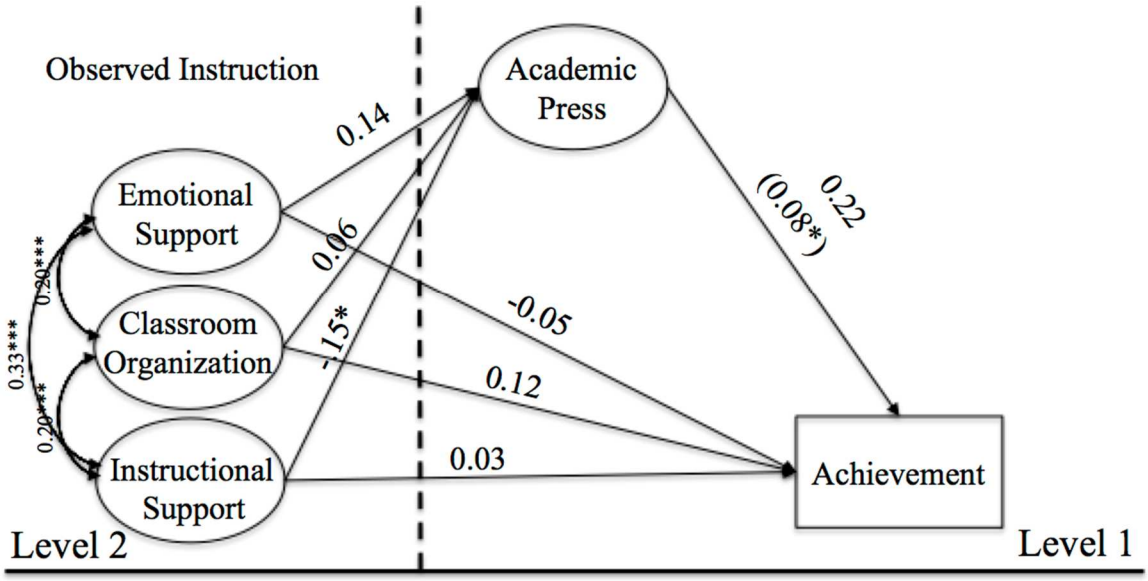


Figure 4.8. Mediation model with elementary school students' perceptions of academic press (3,269 students in 322 classrooms). * $p < .05$. ** $p < .01$. *** $p < .001$. Within effects are depicted in parentheses.

Table 4.5

Unstandardized coefficients of variables predicting to standardized mathematics achievement for middle school (n=3,190)

	Interest	Emotional Support	Instructional Support	Academic Press
		Level-2 variables		
Emotional support	-0.03	-0.02	-0.04	-0.06
	(0.08)	(0.01)	(0.08)	(0.08)
Classroom organization	0.04	0.04	0.03	0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Instructional support	0.04	0.04	0.05	0.05
	(0.11)	(0.11)	(0.10)	(0.10)
Grade 7	0.02	0.02	0.03	0.03
	(0.04)	(0.04)	(0.04)	(0.04)
Grade 8	0.09	0.09	0.10	0.11
	(0.09)	(0.09)	(0.09)	(0.08)
		Level-1 variables		
Student perception	.05***	0.01	.05***	0.03
	(0.01)	(0.01)	(0.01)	(.02)
Male	0.00	0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
English language learner	-0.05	-0.05	-0.05	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)
White	0.09**	0.09**	0.09**	0.09**
	(0.03)	(0.03)	(0.03)	(0.09)
Hispanic	0.05*	0.05*	0.05*	0.05*
	(0.03)	(0.03)	(0.03)	(0.03)
Other	0.20***	0.20***	0.20***	0.20***
	(0.04)	(0.02)	(0.02)	(0.02)
Previous math achievement	0.65***	0.65***	0.65***	0.65***
	(0.02)	(0.02)	(0.02)	(0.02)

Note. * p <.05. ** p < .01. *** p < .001. Unstandardized regression coefficients. Standard errors are shown in parentheses. District dummies are not shown and none are statistically significant. Reference group is female African American sixth grade students. Only effects from variables to student achievement are shown. The coefficients of level-2 variables can be interpreted as the observed classroom dimension's effect on classroom's average achievement. The between effect from student perception to mathematics achievement is not shown in this table. The number of classrooms was 190.

Table 4.6

Unstandardized coefficients of variables predicting to standardized mathematics achievement for elementary school (n=3,269)

	Emotional Support	Instructional Support	Academic Press
	Level-2 variables		
Emotional support	-0.04	-0.04	-0.05
	(0.09)	(0.09)	(0.09)
Classroom organization	0.13	0.13	0.12
	(0.10)	(0.10)	(0.08)
Instructional support	0.02	0.02	0.03
	(0.09)	(0.09)	(0.03)
Grade 4	0.00	0.00	0.01
	(0.03)	(0.03)	(0.03)
	Level 1		
Student perception	0.04	0.01	0.08*
	(0.02)	(0.03)	(.04)
Male	-0.02	0.00	-0.02
	(0.02)	(0.02)	(0.02)
English language learner	0.00	-0.05	0.00
	(0.03)	(0.03)	(0.03)
White	0.16***	0.16***	.16***
	(0.03)	(0.03)	(0.03)
Hispanic	0.09**	0.09**	0.09**
	(0.03)	(0.03)	(0.03)
Other	0.19***	0.19***	0.19***
	(0.04)	(0.04)	(0.04)
Previous math achievement	0.75***	0.75***	0.75***
	(0.02)	(0.02)	(0.02)

Note. * p <.05. ** p < .01. *** p < .001. Unstandardized regression coefficients. Standard errors are shown in parentheses. District dummies are not shown and none are statistically significant. Reference group is female African American fifth grade students. Only effects from variables to student achievement are shown. The coefficients of level-2 variables can be interpreted as the observed classroom dimension's effect on classroom's average achievement. The between effect from student perception to mathematics achievement is not shown in this table. The number of classrooms was 322.

Chapter 5: Key Findings and Future Research

What is good instruction? How do we measure it? How can we design professional development to educate teachers to better their practice? Researchers, policymakers, and educational administrators are continually trying to answers these questions. The current predominant model for understanding teacher quality has been evaluating a teacher's value added to student achievement (Raudenbush, 2004; Rivkin, Hanushek, & Kain, 2005; Rowan, Correnti & Miller, 2002). Though unobserved teacher characteristics are nicely captured by the teacher value-added models, the value that teachers add cannot be directly observed in their instructional practices (Hill, Kapitula, & Umland, 2011). Though these models purport to identify who the "effective" teachers are, they are limited in their usefulness in determining what behaviors these teachers do to make them effective.

The purpose of this dissertation is to identify effective instructional practices for student achievement and to understand how these instructional practices influence student achievement. To do this, I focused on examining the link between observed instruction and student achievement using two statistical methods, and tested a model that hypothesized that student perceptions of the classroom mediate the association between observed instructional practices and student achievement.

Summary of Findings

In chapter 2, I used patterns to describe instructional practices using domain-general (CLASS) and domain-specific (MQI) observation protocols across four time points of observed instruction. I also examined the correlations among teachers' ratings of instructional practices across the four time points of observed instruction. I found

patterns that emerged using the domain-general protocol described low, medium, and high ratings of instruction and did not look any different than if they had been modeled using a continuous variable. On the domain-specific protocol, three patterns emerged: a pattern that described a teacher's practice as high in errors and imprecision and low opportunity for student interactions; a pattern that described teachers as low in error and imprecisions and also low opportunity for student interactions; and a final pattern that described teachers as low in error and imprecisions but high in opportunity for student interaction. In examining the correlations among instructional practices across the four observation points, I found low to medium correlations among the same practices measured over the four time points. These correlations were larger in magnitude for aspects of the domain-general protocol than they were for the domain-specific protocol. This is a striking finding, especially since many policy decisions and tests of theory are based on one observation of a teacher's practice.

In chapter 3, I used the patterns obtained in chapter 2 and related them to students' achievement at the classroom level (classroom mean and standard deviation of achievement). I compared this approach with traditional variable-centered approaches and found that at most time points, instructional practices were not significantly related to classroom achievement (using both variable-centered and pattern centered approaches) for both the domain-general and domain-specific protocols (one exception is the domain-general protocol on the first time point of observed instruction). When using an average of the four time points, observed instructional practices on the domain-general protocol were related to changes in classroom achievement. This effect was larger in magnitude when using the pattern-centered approach compared with the variable-centered approach,

suggesting that the pattern-centered approach is explaining more variance in the outcome than using continuous variables; however, it is still unclear what the pattern-centered approach buys researchers. In only one instance (video 4), I found that aspects of the domain-general protocol were significantly associated with the standard deviation of classroom-level achievement—specifically, emotional support was related to reduced variance of student achievement within the classroom and instructional support was related to increased variance in students' achievement in the same classroom.

In chapter 4, I examined one possible mechanism to explain the findings obtained in chapter 3 (the association between domains of the CLASS and student achievement using a variable centered approach for video 1). I tested a multilevel model with student perceptions mediating the association between instructional practices (as assessed by the domain-general protocol) and students' achievement. Though I did not find significant mediation, I found that the association between instructional practices and achievement could be explained by students' perceptions. I found that middle school students' perceptions of value, emotional support, and instructional support were significantly influenced by observed measures of emotional support. For elementary school students, I found that their perceptions of instructional support and academic press were significantly negatively associated with an observer's rating of a teacher's instructional support. This suggests that different aspects of observed classroom practices may influence different aspects of students' perceptions for different age groups.

Key Lessons Learned

Teacher's instructional practices are not stable within the academic year. In chapter 2, I presented correlations between observed instructional practices at four time

points of instruction. Correlations ranged from -0.40 at their lowest to .39 at their highest, with the domain-general protocol yielding higher overall correlations than the domain-specific protocol. This is troubling, especially when many decisions about theory and practice are based on just one observation (e.g., Stigler, Gonzalez, Kwanaka, Knoll, & Serrano, 1999). It could be that some null results obtained in previous studies and in chapter 3 of this dissertation are due to the timing of when observations were conducted. For example, are some time points during the year critical for setting the tone of student learning? Previous research has pointed to the first two months (Deci, Schwartz, Sheinman, & Ryan, 1981) and maybe even the first three weeks of instruction (Evertson & Emmer, 1982) are critical to establishing classroom management norms. Perhaps students aiming to link instructional practices with student achievement should focus their observations at the beginning of the school year. However, this assumption should be replicated with more recent studies and with different domains of the classroom.

Domain-general protocols have a stronger association with achievement than domain-specific protocols. In chapter 3, I found that classrooms characterized as low on the domain-general protocol were associated with significantly lower achievement than classrooms characterized as medium or high on the protocol. However, we must be cautious in concluding that domain-general protocols are better at discriminating within the low end of the distribution (e.g., Brophy, 1986), because I was unable to replicate this curvilinear effect using a quadratic term in the variable-centered approach using a general rating of classrooms. Future research on the development of observation protocols should conduct detailed analyses of ability to discriminate along the distribution of teaching quality.

Student perceptions of instruction explain the association between observed instructional practices and student achievement. Another key finding of this research is that student perceptions explained the association between observed instructional practices and student achievement. Student perceptions serve as a critical point of explaining how students behave and achieve. If we know which elements of the observed environment students respond to, then we can design teacher preparation programs to focus on those elements of instruction. Additionally, I hope that this study is read by people outside of my immediate field, as it provides useful information on the mechanism of how instructional practices are transmitted to students. I feel that the perspective of students taking ownership of their learning is lost in large-scale evaluation studies of teacher quality.

Different aspects of observed instruction matter for different aspects of student perceptions. The final key finding is that different aspects of observed instructional practices (i.e. different domains of the CLASS) were found to be associated with different perceptual domains and some developmental differences emerged. However, various aspects of students' perceptions were significantly and highly correlated with one another suggesting a halo effect of these perceptions. Indeed, this effect has been found in other studies using other samples (e.g., Patrick, Kaplan, & Ryan, 2011; Schenke, Lam, Conley, & Karabenick, 2015). Yet, it is still unclear if it is a measurement issue—that is, students cannot distinguish between the domains that researchers have theorized—or if classrooms that are high in one dimension are also high in another.

Future Research

This dissertation provides further evidence that instructional practices and student perceptions matter for student achievement. In the end, I replicated what process-product researchers had found regarding the importance of classroom organization on student achievement (for a review, see Brophy & Good, 1986) and found some evidence for a theoretical model that is implicit in many social cognitive models of learning (e.g., Bandura, 1986; Eccles et al., 1983; Weiner 1985). What I hope will be taken away from this work is the need to incorporate older conceptualizations of effective instruction into the design of large-scale studies with the purpose of generating new knowledge on students in their classrooms. Part of this involves compiling studies from different disciplines to understand where conclusions of good instruction converge and diverge. Stipek and colleagues (1998), for example, have done some of this work where they looked at research from the motivation literature as well as literature on mathematics reform to create a list of teacher practices that lead to student achievement. Some of these practices include: the degree to which teachers emphasize student effort, the degree to which teachers encourage student understanding, and the level of teacher enthusiasm. Other researchers such as Candace Walkington are also doing this work (Walkington, Petrosino & Sherman, 2013). Research on *what* a teacher teaches does not preclude focusing on *how* a teacher teaches.

The need for a cohesive theoretical model for how instruction influences student achievement. Moving forward, we need a cohesive model for how classroom contexts and teaching influence students' behavior and achievement. Such a cohesive model could address Berliner's criticism of the process-product literature (and I would also argue the current literature): "how could the number or percentage of teacher verbal

communications coded as praise statements in November influence results on achievement test item given in May?" (p.122, Berliner, 1979). Though these relationships have been found among teachers' actions and behaviors during one point of the year and student achievement at the end of the year (e.g., Allen, Mikami, Lun, Hamre, & Pianram 2013) it is still unclear, what the explanations behind these relationships are and how they come about.

Establishing such a cohesive model would entail compiling a comprehensive list of domain-general and domain-specific practices that matter for student achievement. Though many reviews of instructional practices exist (Brophy & Good, 1986; Fang, 1996), they do not incorporate recent studies on instructional practices and may not include studies from different theoretical orientations such as math education. A cohesive model could also include how certain teacher characteristics such as mathematical knowledge for teaching (Hill, Rowan, & Ball, 2005), credentials, and teacher preparation programs influence student achievement. Thus far, there has not been a consensus on whether or not teacher credentials or mathematical for knowledge for teaching matter for student achievement (Clotfelter, Ladd, & Vigdor, 2007; Croninger, Rice, Rathbun, & Nishio, 2007). It makes logical sense that a teacher who has higher or more developed knowledge of mathematical knowledge for teaching should in fact be able to teach in ways that are more beneficial for student learning than a teacher with less well developed mathematical knowledge for teaching. Or that a teacher who has spent one or two more years taking courses about teaching is better equipped than a teacher who does not have this extra credential. By developing a cohesive model for how these aspects make their way to student behavior, researchers may be able to better understand how and under

what circumstances teacher credentials of mathematical knowledge for teaching matter. I would hypothesize that these elements do matter indirectly for instruction, and that rather than looking at distal outcomes such as student achievement, more proximal outcomes such as teacher's beliefs or self-efficacy should be investigated. Although I did not explore aspects of the teacher aside from observed instructional practices it is important to think about how these characteristics influence behaviors and instruction. Kersting and colleagues (2012), for example, found that observed instructional practices mediated the effect of teacher mathematical knowledge for teaching on student learning. Clements and colleagues (2013) looked at the effect of their professional development intervention on student achievement as mediated by teacher's instructional practices (Clements, Sarama, Wolfe, & Spitler, 2013). Such meditational models could be expanded to include other elements such as student perceptions as was tested in this dissertation. Another aspect to include in the model could be to consider student behaviors such as engagement or attention as mediators from students' perceptions to their achievement as is the case in Skinner and Belmont (1993) and others (e.g., Ing, Webb, Franke, Turrou, Wong, Shin, & Fernandez, 2014; Ruzek, Hafen, Allen, Gregory, Mikami & Pianta, *under review*). Additionally, students' cognitive processes have also been purported to mediate the association between instruction and student outcomes (Winne, 1987; Winne & Marx, 1980) and in my proposed model, could also mediate the association between student perceptions and their behavior.

Continuity of students' classroom experiences. Another promising avenue of research is in understanding the continuity of students' classroom experience both during the academic year and across academic years. Earlier work documenting the middle

school and high school transition (Hirsch & Rapkin, 1987; Midgley, Feldlaufer & Eccles, 1989; Wigfield, Eccles, Mac iver, Reuman, & Midgley, 1991) has considered this continuity and has found that it mattered for student outcomes. For example Midgley and colleagues (1989) looked at students' transition from elementary school to junior high and found that students who had rated their teacher as low on support in elementary school but went into classrooms where they rated their teacher as high in support experienced higher intrinsic motivation than students how had initially highly supportive teachers and transitioned into classrooms where they were low in support. In understanding students' continuity of classroom experiences, information on the cumulative effects of classroom environments could be better explored. Considering cumulative effects will temper expectations that researchers, policymakers, and even parents have about the magnitude of the true effect of a student having one stellar teacher. We cannot expect students' experiences with any one particular teacher to have a large impact on student achievement especially since previous research has pointed out the strong differences that exist before kids even start school (Claessens & Engel, 2013; Duncan et al., 2007). Though researchers have used econometric student-fixed effects methods for understanding the impact of teaching in which unobserved student characteristics are pulled out to get a causal estimate of a particular teacher characteristic (Clotfelter, Ladd, & Vigdor, 2007; Rockoff, 2004), these analyses do not allow us to understand between-student differences in experiences with multiple teachers and over time. Exploring this continuity would include longitudinal modeling of students' experiences within a particular domain such as mathematics, and also across domains. Does what happens in a student's ELA class impact how they perceive their mathematics

teacher? Does having three good teachers provide different results than a student who has one stellar teacher and two bad ones? We might not get large effects from observations and student surveys of one classroom but these associations could be larger in magnitude if researchers considered all of the classrooms students encounter during the school year. To really understand students' experiences in school we need to think about how instruction across all courses and domains work together and at pivotal points of development.

This continuity is important because students' perceptions and behaviors are influenced by their previous experiences. Collecting data on students over multiple years would help researchers understand how student perceptions develop and what aspects of these perceptions influence achievement and other outcomes. Unfortunately most datasets only measure students' perceptions twice a year. In preliminary work, I found that students' perceptions of their classroom are in part explained by their perceptions of their previous teacher (Schenke, Lam, Ruzek, Conley, Karabenick, & Eccles, 2015). This is striking, especially because large scale studies like the MET only measure students' perceptions once during the year and then use aggregated student perceptions to the classroom level as a measure of classroom quality without considering that students in the same classroom could have very different previous experiences. One limitation to the work on student perceptions is that students seem to perceive different aspects of their environments in very similar ways (called the halo effect). However, what remains puzzling is that we are still able to see movement on some constructs and not on other constructs of these perceptions (chapter 4 of this dissertation). If they do in fact mediate the association between observed instructional practices and student outcomes, they can

be used as a supporting mechanism for outcomes researchers and practitioners are interested in such as supporting achievement, career aspirations and help seeking.

Student-level moderators of instructional practices. Another avenue for future research is to understand student-level characteristics that moderate the association between instructional practices student achievement. Previous research in other domains has found that females and under represented minorities respond differently to instructional practices in undergraduate lecture halls (e.g., Reimer, Schenke, Nguyen, O'Dowd, Domina, & Warschauer, 2015), and that first-generation college students respond differently to value affirmation interventions than non first-generation college students (Harackiewicz, Canning, Tibbetts, Giffen, Blair, Rouse, & Hyde, 2014). Looking beyond demographic characteristics, other student-level characteristics such as students' value or self-efficacy beliefs could moderate the effect of instructional practices on student outcomes such that students who have high self-efficacy for a subject may be able to take advantage of the high quality instruction than students with lower self-efficacy for doing well in the class. Such moderators can be described as an Aptitude X Treatment Interaction (Cronbach & Snow, 1977), which suggests that individual differences should be taken into account when understanding the effects of a treatment. Considering the potential for differential effects of instruction on student outcomes could lead to the conclusion that asking what is a teacher's average gain on student test scores is an overly simplistic question, and maybe we should not assume that student achievement is uniformly influenced by the teacher.

Conclusion

Classrooms are complex environments. Even if researchers are able to isolate specific variables that are associated with student achievement, translating those specific variables into practice may not always have the effects they intend to produce. For example, Brophy (1986) suggested that adequate research supports the idea that teachers should wait a given time for students to respond after asking a question; however, just because this is an established research finding does not mean that waiting three seconds after asking a question should be directly translated into policy, and instead waiting three seconds to respond needs to be thought about in the context in which other instructional practices occur. Classrooms are messy environments and reducing their complexity into one equation will not be successful in implementing the policy implications that come from that.

References

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

Bandura, A. (1986) *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Berliner, D. C. (1979). Tempus educare. In P. L. Peterson & H. J. Walberg (Eds.) *Research on teaching* (pp.120-135). Berkeley, CA: McCutchan.

Bill & Melinda Gates Foundation, (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study. *MET Project Policy and Practice Brief.* Bill & Melinda Gates Foundation.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, *41*(10), 1069-1077.

Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328-375). New York: Macmillan.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*,*115*, 1-29.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies persistence of effects in the third year. *American Educational Research Journal*, *50*(4), 812-850.

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.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Oxford, England: Irvington.

Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, *26*(3), 312-324.

Deci, E. L., Schwartz, A. J., Sheinman, L., & Ryan, R. M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of educational Psychology*, *73*(5), 642.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meese, J. L. & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), *Achievement and achievement motives* (pp.75-146). San Francisco, CA: Freeman.

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552-565.

Everston, C. M., & Emmer, E. T. (1982). Effective management at the beginning of the school year in junior high classes. *Journal of Educational Psychology*, *74*(4), 485-498.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational research*, *38*(1), 47-65.

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, *106*(2), 375-389.

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.

Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, *118*(4), 489-519.

Hirsch, B. J., & Rapkin, B. D. (1987). The transition to junior high school: A longitudinal study of self-esteem, psychological symptomatology, school life, and social support. *Child Development*, 58(5), 1235-1243.

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2014). How student participation mediates the relationship between teacher practices and student achievement. Paper presented at the annual meeting of the American Educational Research, Philadelphia, PA.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & 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.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child Development*, *60*(4), 981-992.

Patrick, H., Anderman, L. H., Ryan, A. M., Edelin, K. C., & Midgley, C. (2001). Teachers' communication of goal orientations in four fifth-grade classrooms. *The Elementary School Journal*, 102(1), 35-58.

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

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

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

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

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (*under review*). How teacher emotional support motivates students: The mediation roles of peer relatedness and autonomy support.

Schenke, K., Lam, A. C., Conley, A. M, & Karabenick, S. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology, 41,*133-146.

Schenke, K., Lam, A.C., Ruzek, E., Conley, A.M., Karabenick, S.A., & Eccles, J.S. (2015). Heterogeneity of student perceptions of the classroom climate: A latent profile approach. Paper presented at the annual meeting of the American Educational Research, Chicago, IL.

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), 571-581.

Stigler, J. W., Gonzales, P., Kwanaka, T., Knoll, S., & Serrano, A. (1999). The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States. A Research and Development Report.

Stipek, D., Salmon, J. M., Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education*, *29*(4), 465-488.

Reimer, L., Schenke, K., Nguyen, T., O'Dowd, D., Domina, T. & Warschauer, M. (*under review*). Evaluating promising practices in STEM undergraduate lecture courses.

Walkington, C., Petrosino, A., & Sherman, M. (2013). Supporting algebraic reasoning through personalized story scenarios: How situational understanding mediates performance. *Mathematical Thinking and Learning*, *15*(2), 89-120.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573.

Winne, P. H. (1987). Why process-product research cannot explain process-product findings and a proposed remedy: The cognitive mediational paradigm. *Teaching and teacher education*, *3*(4), 333-356.

Winne, P. H., & Marx, R. W. (1980). Matching students' cognitive responses to teaching skills. *Journal of Educational Psychology*, *72*(2), 257-264.