MEASURING THE EFFECTS OF INSTRUCTIONAL ENVIRONMENT AND STUDENT ENGAGEMENT ON READING ACHIEVEMENT FOR STRUGGLING READERS IN MIDDLE SCHOOL

by

PATRICK CHARLES KENNEDY

A DISSERTATION

Presented to the Department of Educational Methodology,
Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2014

DISSERTATION APPROVAL PAGE

Student: Patrick Charles Kennedy

Title: Measuring the Effects of Instructional Environment and Student Engagement on Reading Achievement for Struggling Readers in Middle School

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Educational Methodology, Policy, and Leadership by:

Dr. Joe Stevens	Chairperson
Dr. Gina Biancarosa	Core Member
Dr. Akihito Kamata	Core Member
Dr. Elizabeth Harn	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research and Innovation; Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2014

DISSERTATION ABSTRACT

Patrick Charles Kennedy

Doctor of Philosophy

Department of Educational Methodology, Policy, and Leadership

March 2014

Title: Measuring the Effects of Instructional Environment and Student Engagement on Reading Achievement for Struggling Readers in Middle School

Teaching is a complex and fundamentally collaborative process, through which knowledge and skills are acquired as a result of repeated interactions between teachers and students. The effectiveness of these interactions depends on both the instructional environment created by the teacher and the extent to which students engage with that environment. The current study combines these dimensions of teaching to (a) evaluate the construct validity of two instruments: the Middle School Intervention Project Classroom Observation Tool (MSIP-COT) and the Student Engagement Instrument (SEI), and (b) test the extent to which these measures predict differences in reading proficiency for middle school students who struggle with reading.

Observation, engagement, and reading achievement data were collected for a sample of 1,446 reading intervention students from 25 middle schools in the Pacific Northwest participating in an evaluation of state and local education programs. Instruments were evaluated by fitting a series of measurement models to the observation and engagement data. The results of the best fitting models were then used in a cross-classified, multilevel structural equation model to predict differences in reading proficiency and evaluate the direct and mediational effects of engagement and instructional environment.

Results provided reasonably strong evidence for both measures as indicators of their respective constructs but limited support for the direct and mediational effects of observed instructional environment and self-reported student engagement on reading proficiency for these students. Limitations and directions for future research are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Patrick Charles Kennedy

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Willamette University, Salem, OR

DEGREES AWARDED:

Doctor of Philosophy, Educational Methodology, Policy, and Leadership, 2014, University of Oregon
Bachelor of Arts, English Literature, 2002, Willamette University

AREAS OF SPECIAL INTEREST:

Measurement and Assessment
Item Response Theory
Psychometrics

PROFESSIONAL EXPERIENCE:

Data Management and Analysis Group Director, Center on Teaching and Learning, 2013-Present

Research Assistant, Center on Teaching and Learning, 2002-2013

GRANTS, AWARDS, AND HONORS:

Magna Cum Laude, Willamette University, 2002
Phi Beta Kappa, Delta of Oregon Chapter, 2002

PUBLICATIONS:

Nese, J. F. T., Biancarosa, G., Cummings, K. D., Kennedy, P. C., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across grades 1–8. *Journal of School Psychology, 51*, 625-642. doi: 10.1016/j.jsp.2013.05.006.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Drs. Joe Stevens, Gina Biancarosa, Akihito Kamata, and Elizabeth Harn for their valuable conceptual guidance and feedback in the preparation of this manuscript. Additional thanks are due to my colleagues at the Center on Teaching and Learning, particularly Dr. Edward Kame'enui, who helped foster my initial interest in research, and Drs. Scott Baker, Deanne Crone, and Hank Fien, who co-directed the Middle School Intervention Project (MSIP), the research project on which this study was based. MSIP was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E100041 awarded to the University of Oregon.

To Amanda, for her patience and support, my parents, who always believed in me, and my children, Connor and Rowan, for their love of life and infectious smiles.

TABLE OF CONTENTS

Chapter Page

Chapter | Page

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER I

STATEMENT OF THE PROBLEM

Teaching is a complex and fundamentally collaborative process, mediated through repeated interactions between teachers and students, a complexity that becomes particularly pronounced as students transition to middle school (Eccles et al., 1993). For decades, researchers have consistently found that some teachers are better at affecting student outcomes than others (e.g., Hanushek, 1971; Kane, Taylor, Tyler, & Wooten, 2010; Rivkin, Hanushek, & Kain, 2005), but identifying and accurately measuring the factors that explain these differences has proven substantially more challenging. Though appealing in their simplicity and availability, proxies of teacher effectiveness (e.g., level of education, years of experience) seldom effectively predict student academic outcomes (Hanushek & Rivkin, 2006), regardless of analytic sophistication used to detect relationships (Fielding & Goldstein, 2006). Rather, student outcomes are better predicted by observational measures that capture important nuances of the context in which those interactions take place (Kane & Staiger, 2012; Kane et al., 2010; MET Project, 2013; Stronge et al., 2011).

Efforts to measure the effects of instruction are further complicated by individual differences in students. Teaching clearly does not benefit all learners equally, even when the context of the instruction is taken into account. Consequently, researchers are increasingly studying the influence of malleable student characteristics such as engagement, the extent to which students connect with school (Appleton, Christenson, & Furlong, 2008; Fredricks et al., 2011). Engagement has been linked to several important academic outcomes, including completion of assignments and test performance (Pintrich

& De Groot, 1990), grades earned (Furrer & Skinner, 2003; Klem & Connell, 2004), and dropout rate (Finn, 1989), illustrating the importance of measures that accurately capture variation in these characteristics, most of which cannot be directly observed.

Surprisingly, despite independent research traditions that span multiple decades, there has been relatively little work investigating the combined effects of instructional context and student engagement on academic achievement, what Maehr & Meyer (1997) termed the co-creation of meaning between teachers and students. Studies on classroom climate and instructional quality seldom include measures of student engagement (e.g., Pianta, La Paro, Payne, Cox, & Bradley, 2002). Those that do generally focus on only one specific aspect of either the instructional context (e.g., the quality of the relationship between teacher and students; Decker, Dona, & Christenson, 2007; Hughes, Luo, Kwok, & Loyd, 2008; Patrick, Ryan, & Kaplan, 2007), or student engagement (e.g., (Downer, Rimm-Kaufman, & Pianta, 2007). In addition, nearly all such studies have been conducted in elementary school settings (i.e., kindergarten through grade 5). Much less is known about the relationship between classroom context and engagement for older students. Nevertheless, when considered together, these studies provide initial evidence of engagement as an important mediator of the effects of instructional context.

Further, results from the few studies that have examined the mediational effects of engagement on classroom context more directly provide promising results. One recent study of urban middle school students found that students' perceptions of their instructional environment in seventh grade affected their observed school engagement in eighth grade, which in turn, was positively associated with higher eighth grade GPA (Wang & Holcombe, 2010). Another found that psychological and behavioral engagement mediated the relationship between classroom context and academic achievement in fifth grade, but only for those students without prior academic difficulties (Dotterer & Lowe, 2011). Notably, these studies focused only on school engagement, emphasizing an individuals' connection to school without accounting for the impact of other potentially important factors, such as family and peer support (Appleton et al., 2008; Furrer & Skinner, 2003), and found no effects for struggling students, those who could reasonably be expected to benefit most from the effects of increased engagement.

Defining the Instructional Environment

Identifying and operationalizing those aspects of the instructional environment that have the greatest impact on student outcomes is a complex task (Lewis et al., 1998), but a number of researchers have proposed conceptual frameworks that attempt to do just that. Two widely used multi-subject approaches are the Framework for Teaching (Danielson, 2007) and the Classroom Assessment Scoring System (CLASS; Hamre et al., 2013; Pianta & Hamre, 2009; Pianta, La Paro, & Hamre, 2008). Each has been rigorously studied and shows substantial evidence of being positively associated with student achievement gains in both English language arts and mathematics (Kane & Staiger, 2012; MET Project, 2013). The Framework for Teaching organizes the responsibilities of teaching into four domains: (a) planning and preparation, (b) classroom environment, (c) instruction, and (d) professional responsibilities. Each domain is further divided into five or six components. Of the four, only classroom environment and instruction reflect behaviors that occur during instruction, the focus of the current study. The other domains represent knowledge and skills required of teachers in their responsibilities outside the classroom and so are not considered here.

The CLASS is more narrowly focused on behaviors that can be observed during interactions between teachers and students, and divides teaching into three domains: (a) emotional supports, (b) classroom organization, and (c) instructional supports. As in the Framework for Teaching, each domain in the CLASS is further divided into several dimensions, and each dimension is measured by a series of indicators. A third framework, the Protocol for Language Arts Teaching Observations (PLATO; Grossman et al., 2010), focuses almost entirely on factors that are included in the other two frameworks as part of the instructional domain, but is relevant to the current study due to its emphasis on middle school ELA instruction. The PLATO identifies 12 essential elements of ELA instruction, ranging from intellectual challenge and explicit strategy instruction to behavior and time management.

Despite obvious differences in their areas of focus, these frameworks are notable for their significant similarities, as illustrated by efforts to synthesize these and other related lines of research (e.g., Stronge et al., 2011; Stronge, Ward, Tucker, & Hindman, 2008). Stronge and colleagues categorized the components that comprise the instructional environment into domains related to the classroom environment (i.e., classroom organization and personal qualities of the teacher), and domains related to instructional practice (i.e., instructional delivery and the use of student assessments). Importantly, these constructs are drawn from and consistent with findings from a long tradition of research on effective instructional practices (e.g., Kane et al., 2010; National Institute of Child Health and Human Development Early Child Care Research Network, 2002, 2005). A closer look at these constructs and their relationship to the frameworks described above provides a useful comparison for the conceptual framework evaluated in the present study.

Components of the classroom environment. Classroom environment refers to both the structure and climate created and fostered by teachers. Structural components include the use of routines, establishment and enforcement of clear expectations, and protection of instructional time. Classroom climate refers to the level of rapport between teachers and students, and the respect they show each other. Each of the frameworks described emphasizes the importance of the classroom environment in facilitating effective teaching. For instance, the Framework for Teaching defines a classroom environment by the extent to which the teacher (a) creates an atmosphere of respect and rapport, (b) establishes a culture for learning, (c) manages classroom procedures, (d) manages student behavior, and (e) organizes physical space (Danielson, 2007). The CLASS further divides classroom environment into two domains: emotional supports and classroom organization (Pianta et al., 2008). The former represents the emotional climate of the classroom (e.g., sensitivity and respect for student perspectives), while the latter represents the structural components (e.g., behavior management, productivity, and instructional learning formats). The PLATO does not organize its essential elements into domains, but includes two elements related to classroom structure: behavior and time management (Grossman et al., 2010).

Components of instructional practice. Instructional practice refers to the ways in which teachers communicate information to facilitate learning. According to Stronge et al. (2011), high quality instructional delivery consists of six effective teacher behaviors: (a) use of varied instructional practices, depending on student need; (b) modeling critical thinking skills, maximizing instructional time, and minimizing the amount of time spent on classroom management; (c) communicating clearly; (d) recognizing complexity and focus on concepts and connections rather than facts; (e) demonstrating high expectations for all students; and (f) using technology appropriately. Further, they argue that differentiated instruction is most effective when learning is monitored and students receive feedback on their performance (Stronge et al., 2011).

Not surprisingly, all three frameworks propose similar conceptualizations of effective instruction, although they differ slightly in their emphasis. For instance, the Framework for Teaching is organized primarily around what teachers do, whereas the CLASS framework emphasizes how teachers interact with their students. Indicators of effective instruction in the Framework for Teaching include the extent to which teachers (a) communicate effectively with students, (b) use questioning and discussion techniques, (c) engage students in learning, (d) use assessment in instruction, and (e) demonstrate flexibility and responsiveness (Danielson, 2007). In the CLASS framework, instructional support is represented by three constructs: (a) concept development, (b) quality of feedback, and (c) language modeling (Hamre et al., 2013; Pianta & Hamre, 2009). The PLATO framework is primarily focused on this domain, and includes elements that are conceptually similar to each of the other frameworks, operationalized using terms like representation of content, explicit strategy instruction, and classroom discourse.

The current framework. The framework for the instructional environment evaluated in the present study has features in common with each of the other frameworks just described. Like the PLATO, it is designed for use in middle school literacy classrooms; like the CLASS, it emphasizes the observable characteristics of classroom interactions; and like the Framework for Teaching, it posits a single classroom environment domain and a separate instructional practice domain. However, the current framework also hypothesizes that the effective use of monitoring to make instructional adjustments represents a third factor, distinguishable from instructional practice. In each of the other three frameworks, elements of monitoring and adjustments are included as part of the instruction domain.

In the current framework, classroom environment is operationalized much like it is in both the Framework for Teaching and the CLASS. It includes indicators for classroom climate, organization and routines, classroom management, and student participation and engagement. Similarly, instructional practice is measured along several dimensions. A primary component is intellectual challenge, the extent to which teachers demonstrate high expectations for all students, extend student responses by asking for clarification and elaboration, and devote ample time to high-level analysis and inferential thinking. Also included are representation of content, the frequent use of effective and varied activities and examples; and relevance of learning, the extent to which teachers make explicit connections between new material and students' prior knowledge and personal experiences. It also includes an item labeled teaching for reading proficiency, designed to measure the extent to which teachers display confidence in the classroom, are at ease teaching, and emit a high sense of efficacy, which helps facilitate clear, concise, and consistent communication. Finally, the current framework includes items for student monitoring and instructional adjustments, based on the premise that effective teachers regularly monitor their students for understanding, re-teach as necessary, and adjust instruction to meet individual student needs. The current study tests the hypothesis that variation in all three factors of instructional environment can be measured, and that together, they are predictive of reading proficiency of struggling readers.

Measuring the Instructional Environment

Attempts to measure the impact of teachers on the instructional environment typically take one of three approaches. Many evaluations of instructional effectiveness rely on relatively indirect measures of teacher competence (e.g., education level, years of experience, and other similar proxies). Although they are easier and less costly to collect than more nuanced measures, attempts to explain differences in the effectiveness of teachers using these variables suggest that they are relatively poor predictors of student academic success (Hanushek & Rivkin, 2006), even when used with sophisticated analytic methods (Fielding & Goldstein, 2006). Similarly, the results of studies on the effects of licensure or certification status are equivocal at best (Goldhaber & Brewer, 2000; Kane, Rockoff, & Staiger, 2008; Strong, 2011). Thus, despite being reasonably easy to measure, these proxies do not reliably differentiate between effective and ineffective instructional environments.

A second approach to quantifying differences between teachers is known as value-added modeling. This approach uses statistical modeling to associate measurable gains in student achievement with individual teachers (Sanders, 2000). There is substantial debate about the validity and subsequent policy implications of these models (e.g., Glazerman et al., 2010; Hanushek & Rivkin, 2010; Rothstein, 2010), but even if value-added analyses do provide an estimate of teachers' contribution to gains on standardized assessments, they do not provide any information about why those gains occurred (Pianta & Hamre, 2009). A third approach to quantifying instruction is the use of expert ratings. Such summative evaluations consistently perform poorly in identifying effective teaching as measured by value-added modeling, especially when based on short observations of instruction. In one study, experts (administrators and other educators) correctly categorized teachers only about one third of the time (Strong, Gargani, & Hacifazlioglu, 2011). In a second study reported by the same authors, well-trained judges using the CLASS (Pianta et al., 2008) to rate a full-length lesson were only able to correctly identify effective teachers 50% of the time. However, subsequent analyses by these same researchers demonstrated that classification accuracy could be improved using only those items from the instructional support domain (Strong et al., 2011), suggesting that when focused on important features of instruction and operationalized accurately, classroom observations may in fact be a relatively efficient way to identify effective instructional environments.

Somewhat paradoxically, despite the development of a myriad of observational systems and approaches, evidence linking observational outcomes to student academic success has only recently begun to accumulate (Strong et al., 2011). One recent study of the extent to which classroom observations predicted student outcomes found that highly systematic observations focused on the instructional environment and conducted by well-trained, neutral observers could both (a) accurately identify effective instruction, and (b) provide descriptive information about what made those teachers effective (Kane et al., 2010). Similar evidence is provided by the Measures of Effective Teaching (MET) Project, a large-scale longitudinal study of teaching effectiveness (MET Project, 2013). Results from this effort suggest that repeated classroom observations conducted by trained observers provide important information about instructional effectiveness, beyond the information provided by measures of student feedback and value added estimates (Kane & Staiger, 2012; MET Project, 2013). Across these and similar studies, what seems to best predict differences in teachers are robust measures of the learning environment and the quality of instruction (Kane & Staiger, 2012; Kane et al., 2010; MET Project, 2013; Stronge et al., 2011), such as those operationalized in the frameworks described above.

The current study. Using a similar operationalization of instruction, the current study evaluates a new observational measure, the Middle School Intervention Project Classroom Observation Tool (MSIP-COT) (Nelson-Walker, Kennedy, Cohen, & Crone, 2011), developed specifically for the project as a way to efficiently assess middle school instructional practices along three dimensions: the environment of interactions, the quality of instructional practices, and the use of instructional adjustments. Several key features differentiate the current study from other recent analyses. First, the project on which this study was based was not intended as a measurement study. Rather, the current study was conceived as a way to augment the original project by evaluating the construct validity of a measure that was developed to facilitate the collection of large-scale observation data. As such, this study attempted to measure instructional quality and the classroom environment using substantially fewer items than most observation measures (Kane & Staiger, 2012; Kane et al., 2010; MET Project, 2013). An instrument with fewer items has important implications for the scalability of classroom observation efforts. Second, this study specifically tested whether the inclusion of a separate measure of instructional adjustments improved the evaluation process. Third, the analyses in this study were conducted using confirmatory factor analysis (CFA) and structural equation modeling (SEM), rather than principal components analysis (Kane et al., 2010), approaches that allow for the separation of common item and residual variance.

Defining Student Engagement

Researchers have long sought to understand what factors mediate the impact of quality instruction. Recognizing the limitations of traditional measures of student background characteristics (e.g., intelligence, prior achievement, demographics), recent attempts to explain differences in students' academic performance have focused increasingly on engagement as a potential mediator of success (Fredricks et al., 2011). However, simply defining the term is a significant challenge: one literature review identified nearly 20 different operational definitions of the construct (Appleton et al., 2008). For instance, although literature on engagement in the context of schools often uses the terms "school engagement" and "student engagement" interchangeably (Appleton et al., 2008; Fredricks, Blumenfeld, & Paris, 2004), an important distinction can be made between the two: school engagement implies that engagement is a characteristic of the environment, rather than the individual. In addition, school engagement emphasizes the effects of the school environment, obscuring the impact of other potentially important influences, such as family or community (Appleton et al., 2008). Although affected by the social contexts experienced, engagement is an attribute of the person, not the context. Consequently, the term student engagement is used here.

As a construct, student engagement subsumes multiple types of engagement. It is almost universally described as multidimensional, and is typically operationalized as the combination of two or more of the following: behavioral, affective, cognitive, and academic engagement (Appleton et al., 2008). Early engagement research distinguished between two types: behavioral and affective engagement (Finn, 1989). Behavioral engagement refers to observable actions that serve as indicators of the extent to which a student is an active participant in school (Appleton et al., 2008; Fredricks et al., 2004). Individual actions can be either positive (e.g., following rules, class participation) or negative (e.g., disrespecting a teacher, skipping class). Distinctions can also be made among degrees of engagement, ranging from a direct response to teacher direction representing relatively low engagement on one end, to a student-initiated interaction representing relatively high engagement on the other (Finn, 1989; Fredricks et al., 2004).

Affective engagement refers to the extent to which an individual identifies with a particular group or social context and feels as though they belong (Finn, 1989; Fredricks et al., 2004). Many of the concepts associated with affective engagement are closely related to those described in the literature on motivation. For instance, one model of motivation research, self-determination theory, argues that one of three basic human needs is relatedness (Pintrich, 2003). Similarly, research on achievement motivation also posits that individuals are motivated by three basic needs, one of which is a need for affiliation or attachment (Pintrich, 2003). However, there is a key distinction between motivation and engagement: motivation research focuses primarily on the psychological processes affecting belonging, whereas research on affective engagement emphasizes the act of being involved in a situation (Appleton et al., 2008). Engaged individuals are thus, by definition, motivated, but the reverse is not necessarily true.

More recently, research on engagement in schools has emphasized cognitive engagement, the degree of personal investment in learning (Fredricks et al., 2004). Cognitive engagement is often operationalized as a combination of psychological investment in learning and the use of self-regulation, the extent to which students perceive and demonstrate control over their learning experience (Fredricks et al., 2011).

As with affective engagement, cognitive engagement is closely related to components of motivational theory. Specifically, another of the basic human needs identified by self-determination theory is autonomy (Pintrich, 2003). A related concept is academic engagement (Appleton, Christenson, Kim, & Reschly, 2006), the observable, or behavioral, component of cognitive engagement. Academic engagement is often operationalized using measures like time on task, homework completion, and credit hours toward graduation. Importantly, some indicators of engagement reflect engagement at a specific moment in time, such as many of the examples just offered, while others provide a measure of engagement long term, such as rate of attendance over the course of an academic year. The stability of various facets of student engagement remains an area in need of further study.

The current framework. The framework for student engagement evaluated in the present study includes measures that attempt to differentiate between the sources of affective engagement and the context of cognitive engagement. As proposed by Appleton et al. (2006), affective engagement with school consists of three factors: relationships with teachers, peer support for learning, and family support for learning. Similarly, cognitive engagement can differ depending on the context; perceptions of control and relevance in one context (e.g., middle school) may differ markedly from the perceived relevance of future educational opportunities (e.g., after high school). The current study tests the hypothesis that all five factors of affective and cognitive engagement can be identified and measured, and that affective and cognitive engagement each provide unique contributions to predicting reading proficiency, partially mediating the effect of the instructional environment.

Measuring Student Engagement

Approaches to measuring engagement typically vary depending on the type of engagement being measured. As with many attempts to measure the instructional environment, student engagement has traditionally been assessed using those measures that are most easily obtained: proxies for behavioral engagement and academic engagement (i.e., data on observable behaviors), such as attendance and referrals, or number of credits earned towards graduation (Appleton et al., 2008). Although behavioral engagement and academic engagement are important components of student engagement, using only those variables that can be readily observed substantially simplifies a complex construct.

Affective engagement and cognitive engagement are typically measured via surveys, administered to teachers, students, or both. A strong argument in favor of measuring affective and cognitive engagement by asking students directly is that doing so reduces the degree of inference required (Appleton et al., 2006). This may lead to a more valid understanding of these constructs, although one drawback to asking students directly is the possibility of bias resulting from students who respond based on perceived social norms (Duncan & McKeachie, 2005). Not surprisingly, evidence suggests that higher levels of engagement are positively correlated with academic outcomes (Dotterer & Lowe, 2011; Fredricks et al., 2004). The current study evaluates the extent to which the affective and cognitive engagement of students identified as struggling readers can be measured using an established, validated student self-report measure of engagement, the Student Engagement Instrument (SEI).

Study Purpose

This study draws on insights from previous research measuring the instructional environment and student engagement, both separately and together, in an attempt to better understand the complex interaction between teachers and students that serves as the foundation for learning in the critical middle school years. Emphasizing the dynamic, interactive nature of the learning process, the current study builds on prior research in two ways. First, it evaluates the construct validity of two measures in the context of struggling students: an observation tool designed to measure the instructional environment of middle school English language arts (ELA) and reading intervention classrooms, and a survey designed to measure students' perceptions of their engagement across multiple contexts: school, family, and peers. Second, it uses those measures to assess the extent to which self-reported student engagement mediates the relationship between the instructional environment and reading achievement for middle school students with reading difficulties.

Specifically, this study uses a direct observation measure of the instructional environment, the Middle School Intervention Project Classroom Observation Tool (MSIP-COT), and a student self-report measure of affective and cognitive engagement, the Student Engagement Instrument (SEI) to answer six research questions (RQ):

1. Does the MSIP-COT measure the three hypothesized dimensions of instruction (environment of interaction, quality of instructional practices, and use of instructional adjustments) in 7th grade English Language Arts (ELA) and reading intervention classes?

2. How much variability on the dimensions of the MSIP-COT is captured between ELA and reading intervention classrooms?
3. Does the SEI measure the five hypothesized factors of student (engagement teacher-student relationships, peer support for learning, family support for learning, control and relevance of schoolwork, and future aspirations) for struggling readers?
4. How much variability between students does the SEI capture on these dimensions?
5. Do the MSIP-COT and SEI significantly predict differences in reading proficiency for students identified as struggling readers?
6. Does student engagement, as measured by the SEI, mediate the effect of the instructional environment, as measured by the MSIP-COT, on student reading proficiency for struggling readers?

CHAPTER II

METHOD

The current study examined the direct and mediational effects of instructional practices and student engagement on reading achievement for struggling readers in a sample of 1,446 seventh grade reading intervention students.

Participants

Participants in this study included literacy teachers and seventh grade reading intervention students from 25 Pacific Northwest middle schools that participated in the Middle School Intervention Project (MSIP; Baker, Fien, & Crone, 2009), a federally funded Evaluation of State and Local Education Programs (hereafter, the evaluation), during the 2011-12 school year. The full evaluation used a non-experimental regression discontinuity design to evaluate the effectiveness of an intervention consisting of three parts: (a) intensive reading instruction, (b) individualized student engagement programs, and (c) ongoing data based monitoring and adjustment. Teachers in participating schools administered the complete intervention to as many struggling readers as possible, given budgetary and other resource constraints.

Districts. Districts with an existing working relationship with the evaluating organization were invited to participate in the evaluation if (a) the district demonstrated a commitment to evaluating reading instruction practices in middle school, (b) a majority of schools were already implementing all three components of the intervention prior to the start of evaluation, and (c) evaluation team leadership believed that implementation across districts was sufficiently consistent to allow for meaningful comparisons (Baker et al., 2009). In 2011-12, six districts in the Pacific Northwest participated in the evaluation.

Schools. Within districts, individual schools were allowed to opt out of the evaluation in cases where the district, school, and evaluation team leadership mutually agreed that one or more characteristics of the school prevented an adequate evaluation. In all cases, this decision was made because a school did not offer a traditional middle school curriculum (*n* = 9) and/or less than 10% of seventh grade students were eligible for intervention (*n* = 6). Across the project, participating schools represented 25 of the 40 district schools that served seventh grade students (63%). On average, participating schools were larger and served a more academically diverse range of students than schools that did not participate. All teachers in participating schools who taught at least one reading intervention or English Language Arts class were included in the study (*n* = 188). Because participating schools agreed to implement the full evaluation as part of standard district practice, active consent was not obtained from teachers.

Students. All seventh grade students who began the 2011-12 school year at a participating middle school and who received a reading intervention for at least one term (*n* = 1,635) were eligible for the current study, with one exception: students with an IEP requiring them to attend a life skills or similar class for cognitively low functioning students (*n* = 78) were excluded because (a) these students were not required to take the standard version of the statewide assessment test (a primary outcome measure in this study), and (b) the instruction in these classes typically did not include reading as a goal.

Eligibility procedures. Eligibility for intervention was determined as part of the larger evaluation project on a school-by-school basis. At the conclusion of the 2010-11 school year, project staff created a composite *z*-score for each incoming seventh grade student based on the linear combination of an easyCBM passage reading fluency (PRF;

Alonzo, Tindal, Ulmer, & Glasgow, 2006) measure administered in the spring of 2011 and the sixth grade state reading assessment (Oregon Department of Education Office of Assessment and Information Services [ODE OAIS], 2011a) scores that were strongly positively correlated, $r(6314) = .61$, $p < .001$. Administrators at each school were then given a rank ordered list of *z*-scores and asked to choose the cut score corresponding to the number of students they had the resources to serve. Students below the cut score were assigned to intervention, students above the cut score were assigned to the comparison group. To increase project participation, schools were also allowed to exclude up to five percent of students from the assignment procedures (i.e., these students were not assigned to the condition indicated by their position on the composite *z*-score list). Across the project, 3.95% of students (242 of 6,129) were excluded from the project based on school decision. At the conclusion of each term, actual intervention participation was determined by cross-referencing student schedules with a list of classes identified by the schools as reading interventions.

Reading interventions were implemented as standard practice by participating schools. Consequently, student consent for the current study was based on whether they completed the student engagement measure. The instructions read aloud prior to administration of the SEI clearly stated that completion of the survey was voluntary. Therefore, students who did not complete the measure ($n = 111$, 6.8%) were considered not to have given their consent to participate in the current study and were excluded from analyses. The final sample consisted of the 1,446 students who received a reading intervention in seventh grade and took the engagement measure that fall. Demographics for the final student and school samples are reported in Table 1. Fifty-five percent of participants were male, 84% were Hispanic or white, 17% were identified as limited English proficiency, and more than a third were identified for special education services.

Table 1

Student-level Demographic Data and School-level Poverty and Class Size Indicators

Variable	*N*	%
Total students	1446	100.0
Demographics not reported	18	1.2
Gender		
Female	630	43.6
Male	798	55.2
Race/Ethnicity		
American Indian/Alaska Native	38	2.6
Asian	41	2.8
Black/African American	50	3.5
Hawaiian/Pacific Islander	17	1.2
Hispanic	531	36.7
Multiracial	56	3.9
Not reported	17	1.2
White	678	46.9
Limited English proficiency (LEP)		
Not identified as LEP	1168	80.8
Identified as LEP	244	16.9
Not reported	16	1.2
Special Education (SPED)		
Not identified for SPED	895	61.9
Identified for SPED	533	36.9

Note. Percentages within each demographic item do not sum to 100% due to students for whom demographic data were not reported. Free or reduced price lunch and pupil-teacher ratio statistics were available at the school level from the National Center for Education Statistics (NCES; 2011). Across participating schools, an average of 48.77% of students qualified for free or reduced price lunch, and the average pupil-teacher ratio was 21.3.

Measures

To investigate the research questions of interest, four measures were used: (a) the Middle School Intervention Project Classroom Observation Tool (Nelson-Walker et al., 2011), a researcher-developed classroom observation tool; (b) the Student Engagement Instrument (Appleton et al., 2006), a student-completed attitudinal measure of engagement with school, family, and peers; (c) the Oregon Assessment of Knowledge and Skills Reading and Literature test (ODE OAIS, 2011b), a standards-based measure of reading comprehension and vocabulary; and (d) easyCBM Passage Reading Fluency (Alonzo et al., 2006), a measure of fluency with connected text. The structure and technical adequacy of each measure is described below.

MSIP Classroom Observation Tool (MSIP-COT). The MSIP-COT (Nelson-Walker et al., 2011) provides measures of the three hypothesized dimensions of effective instructional practices: (a) the environment of interaction between teachers and students, (b) the quality of instructional practices, and (c) the use of instructional adjustments. The complete observation protocol includes a cover sheet, completed before the observed class begins; a version of the Classroom Observation of Student-Teacher Interactions (COSTI; Smolkowski & Gunn, 2012), completed during the class period; and an observational rating form, completed immediately following the observation. The cover sheet identifies the observation and summarizes general information about the classroom (e.g., group size). The COSTI documents, for each classroom activity, the group structure, reading content domain, and the frequency of observed behaviors by the teacher and students. At the conclusion of an observation, the observer uses the rating form to rate the class on seven questions regarding overall fidelity (1 = not observed, 2 = occasionally observed, 3 = frequently observed, 4 = consistently observed throughout the class period), and eleven questions regarding the instructional environment (1 = not present, 2 = somewhat present, 3 = present, 4 = highly present). These last 11 items are the focus of the current study. Appendix A provides a description of each item and operational definitions of each score; the complete observation protocol is provided in Appendix B.

Inter-rater agreement for field checkouts of the MSIP-COT exceeded 91% for interaction (COSTI) observations at each round, and 94% for the rating form. Inter-rater agreement for maintenance reliability was above 88% for interaction observations at each round, and above 93% for fidelity observations (Nelson-Walker, Turtura, & Cohen, 2012). Additionally, results from a version of the COSTI used to measure student-teacher interactions in kindergarten showed similar estimates of reliability (Smolkowski & Gunn, 2012). Further, student behaviors in the kindergarten analyses predicted gains in several reading outcomes (Smolkowski & Gunn, 2012). Predictive validity coefficients of reading and mathematics outcomes in elementary school classrooms using the 11 instructional environment items range from .26 to 42. Additional information about the development of the measure, its components, and its function are provided in the MSIP-COT technical report (Nelson-Walker et al., 2012).

Student Engagement Instrument (SEI). The SEI (Appleton et al., 2006) is a measure of cognitive and affective engagement with school, family, and peers, as reported by students. In the current study, the 2006 version of the SEI was administered. The measure consists of 35 statements, each targeting a specific facet of engagement (e.g., “School is important for achieving my future goals” and “At my school, teachers care about students”). For each statement, students indicate their level of agreement by selecting a response from a four-point Likert rating scale (1 = strongly agree, 2 = agree, 3 = disagree, and 4 = strongly disagree). The complete protocol is available in Appendix C.

Items on this version of the SEI are organized into six factors: (a) teacher-student relationships, (b) peer support for learning, (c) family support for learning, (d) control and relevance of schoolwork, (e) future aspirations and goals, and (f) extrinsic motivation. The first three are components of affective engagement; the last three are components of cognitive engagement. All factors are measured by at least four items, with the exception of extrinsic motivation, which is represented using only two items. Coefficient alphas for each of the six factors based on an initial development sample of 1,931 ninth grade students ranged from .72 to .88 (Appleton et al., 2006). Later versions of the SEI dropped the two extrinsic motivation questions, resulting in a five-factor model (Betts, Appleton, Reschly, Christenson, & Huebner, 2010). Reliability estimates for the five-factor model of the 2010 version of the SEI were similar, ranging from .70 to .80 (Betts et al., 2010). Although both extrinsic motivation questions were administered to the current sample, the analyses reported here also exclude the extrinsic motivation factor because the primary interest was in evaluating students' psychological investment in learning, rather than externally motivating factors.

Validity evidence for the SEI is limited, although correlations between the SEI and a variety of academic and behavioral outcomes provide some evidence of convergent and discriminant validity (Appleton et al., 2006). Correlations between item totals for each factor and GPA ranged from .001 for control and relevance of schoolwork to .253 for future aspirations and goals, while correlations with reading achievement ranged from -.287 for control and relevance to .171 for teacher-student relationships. Similarly, correlations with math achievement ranged from -.249 for control and relevance to .162 for teacher-student relationships. Conversely, correlations with suspension status ranged from -.201 for teacher-student relationships to .032 for control and relevance (Appleton et al., 2006).

Oregon Assessment of Knowledge and Skills Reading/Literature (OAKS-R). The OAKS (ODE OAIS, 2012) are a series of standards-based tests aligned to grade-level content in a variety of subject areas. The Reading and Literature knowledge and skills assessment is a computer-adaptive, multiple choice test linking student performance to reading standards of word meaning, comprehension (literal, inferential, and evaluative), and identification and use of literacy elements. The assessment tests students' ability on six skills: (a) vocabulary, (b) reading to perform a task, (c) demonstrating general understanding, (d) developing an interpretation, (e) examining the content and structure of informational text, and (f) examining the content and structure of literary text (ODE OAIS, 2011b). The OAKS-R is a computer adaptive test and is not timed, but most students complete the assessment within two 50 minute sessions (ODE OAIS, 2012). In 2011-12, all Oregon students in grades 3-8 and 10 were required to take the OAKS-R. Students were allowed three attempts over the course of the school year, receiving credit for their highest score.

In seventh grade, the OAKS-R is comprised of 50 multiple-choice items. Between six and 12 of the items correspond to each literacy skill. Questions are one of three difficulty levels, and are distributed across several passages, selected to represent a range of writing types: literary, informative, and instructional. The specific passages and questions presented to each student are selected using a computer algorithm that aligns the difficulty of the passages with the estimated ability level of the student. Each question has four response choices, only one of which is correct ("none of the above" and "all of the above" are not used as options). Scoring is dichotomous; each question is worth one raw score point. The total raw score is converted into a RIT scale score based on application of a one-parameter (Rasch) item response theory (IRT) model. The IRT scale scores across tested grades typically range from 150 to 300. Through a standard setting procedure, three performance levels have been defined on the IRT scale. For seventh grade students in 2011-12, scale scores below 229 were labeled does not meet expectations; scores between 229 and 240 were labeled meets expectations; and scores of 241 and above were labeled exceeds expectations (ODE OAIS, 2011b).

Extensive reliability and validity information is available for the OAKS (ODE OAIS, 2007). For instance, the standard error of measurement (SEM) for the seventh grade OAKS-R is between three and four points across nearly the entire range of ability scores (except at the extreme tails of the distribution), suggesting that the OAKS-R provides similar and consistent information across the range of ability. Similarly, OAKS-R classification accuracy is high; although the seventh grade version was not evaluated specifically, classification accuracy across four grades and two content areas exceeded 84% in all cases. Content validity for the OAKS-R was demonstrated through expert judgment of the alignment between the assessment and content standards, test specifications, and the item development process (ODE OAIS, 2007). Construct validity for the seventh grade OAKS-R is strong; correlations with two different measures of reading achievement were .80 and .82 (ODE OAIS, 2007). Criterion validity was provided only for the tenth grade version of the OAKS, but evidence suggested that higher OAKS-R scores were associated with higher SAT scores, a higher first year college GPA, and a greater likelihood of reenrolling in a second year of college (ODE OAIS, 2007).

***easyCBM* Passage Reading Fluency (PRF).** *easyCBM* PRF (Alonzo et al., 2006) is a standardized, individually administered measure of fluency (the speed and accuracy with which students read connected text). Students are asked to read aloud from a short narrative passage for one minute. The assessor follows along on a separate copy of the passage, marking student deviations from print, and at the end of one minute counts the total number of words read and the number of errors made. Average alternate form reliability across 20 seventh grade passages was .89 (*SD* = .04; Alonzo, Park, & Tindal, 2008). Although it ostensibly measures a separate skill, PRF is moderately predictive of performance on the OAKS-R. Based on a recent analysis of a convenience sample from three districts, fall seventh grade PRF scores correlated .38 with seventh grade OAKS-R scores, accounting for 15% of OAKS-R variance (Anderson, Alonzo, & Tindal, 2010).

Data Collection

Data for the study were collected using a variety of methods, depending on the data source, and in some cases, the district. Data were collected using direct observation, student report, computer adaptive testing, and individualized administration. The data collection and scoring procedures used for each data source are described below.

MSIP-COT. A direct observation measure of classroom practices, the MSIP-COT was collected by trained data collectors. The data collection team was comprised of three groups of individuals: (a) evaluation specialists, individuals with expertise in literacy instruction who were employed by the evaluating institution and had part-time duties related to the evaluation project; (b) teachers on special assignment (TOSA), full-time employees of the participating public school districts who had expertise in literacy instruction; and (c) individuals hired specifically to conduct classroom observations and collect other data related to the evaluation. The complete data collection team received 12 hours of observation training in the fall, and an additional six hours of training in both winter and spring. Each training session consisted of a review of project goals and the purpose of observation, detailed coverage of the measure and procedural manual, and coding practice using video clips of middle school literacy instruction.

Data collectors were required to demonstrate reliability with the ratings of the observation coordinator on each section of the MSIP-COT before beginning observations in classrooms (Nelson-Walker et al., 2012). Data collectors who did not meet specified reliability criteria (between .85 and .90 inter-rater agreement on both item and total scores for each section; see Appendix D for more details) were required to participate in additional training activities prior to observing study classrooms. Data collectors hired during the school year (i.e., after the fall training) received individualized training with the observation coordinator prior to participating in observations.

Each instructional grouping (a unique combination of teacher, period, curriculum, and group size) was observed three times during the year, once each in the fall, winter, and spring, unless the class did not meet for the full academic year. On average, instructional groups were observed 2.35 times per year. For 20% of observations, observers were paired to measure inter-rater agreement. Across the project, average observer agreement exceeded 88%.

SEI. The SEI was group administered by either a classroom teacher or a trained data collector, depending on the preference of the district. Administration was standardized, using the directions presented in Appendix E. After introducing the survey, administrators read each item aloud, pausing between items to allow students to select their response. Completed surveys were returned to the evaluation team and processed using TeleForm (Cardiff Software, 2009), a computerized forms processing application used to collect and efficiently process large quantities of data. The software evaluated each scanned form individually to produce an initial response set for each student. A data verification operator then reviewed each form to ensure the accuracy of each response. Prior to analysis, items were reverse-coded using SPSS 20 (IBM Corporation, 2011), so that higher scores represented higher levels of engagement. Students were administered the SEI in the fall of 2011, prior to the start of classroom observations.

OAKS-R. The Oregon Department of Education maintains an online interface through which students access the computer adaptive test (ODE OAIS, 2012). Between October 2011 and May 2012, students logged on and took the OAKS-R between one and three times, according to standard state and district procedures. At the conclusion of the testing window, the TOSA at each district provided the evaluating institution with a data file containing project ID numbers and the highest OAKS-R score for each student.

***easyCBM* PRF.** The evaluation team gave districts three options for administering the seventh grade PRF passage, depending on their existing formative assessment administration procedures, if any. Each district chose the administration method that caused the least disruption to existing practices. Districts that administered easyCBM measures independent of the evaluation gave the project passage without modifying their existing procedures ($n = 2$). Districts that used a different formative assessment system administered the common project passage prior to administering any measures from the other assessment system ($n = 2$). Districts that did not have a formative assessment system in place prior to the start of the evaluation received training in and assistance administering the PRF measure from the evaluation team ($n = 2$). To maximize the extent to which administration was standardized across districts and prevent form effects, all options required districts to administer the project passage before administering any other measures.

In all cases, the passage was collected by a combination of classroom teachers and trained data collectors between late May and early June 2012, after all classroom observations were completed. In five of the six districts, district personnel completed scoring and data entry onsite and transferred a completed data file containing student project ID numbers and the number of words each student read correctly in one minute to the evaluation team. In the sixth district, protocols were scored by district personnel and transferred to the evaluation team, where data collectors entered and verified scores directly in the project database.

Analyses

The analyses for this study are organized into three parts: (a) a measurement model to validate the MSIP-COT and calculate trait estimates for English Language Arts (ELA) and reading intervention classes, (b) a measurement model to validate the SEI for struggling readers and calculate engagement trait estimates for students, and (c) a cross-classified structural regression model incorporating the results of both measurement models to estimate the direct and mediational impact of the instructional environment and student engagement on reading achievement for seventh grade reading intervention students. For the measurement model analyses, classroom observation and student engagement data were randomly divided into separate validation and re-specification samples (*n* = 399 observations and 723 students, respectively). The validation samples were used to evaluate the fit of the hypothesized measurement models. The re-specification samples were kept separate in case substantial changes to the measurement models were needed. After the measurement models were validated, trait estimates were calculated for the complete samples and used in the structural regression model. Correlations and descriptive statistics were calculated using SPSS 20 (IBM Corporation, 2011). All other analyses were conducted using Mplus 7 (Muthén & Muthén, 2012).

MSIP-COT measurement model. The instructional environment portion of the MSIP-COT was developed to assess three dimensions of instruction: the environment of interactions, the quality of instructional practices, and the use of instructional adjustments. A preliminary exploratory factor analysis (EFA) of the measure using observation data from sixth grade ELA and reading intervention classes indicated that a three-factor solution was appropriate, with four items loading primarily onto the environment of interactions factor, four items loading primarily onto the instructional practices factor, and three items loading primarily onto the instructional adjustments factor. To evaluate the structure of MSIP-COT data from seventh grade ELA and reading intervention classes, four competing models were specified and tested. To investigate the hypothesis that the MSIP-COT measured the three instructional context factors described above (RQ 1), a multiple factor CFA model was fit to the validation sample as shown in Figure 1.

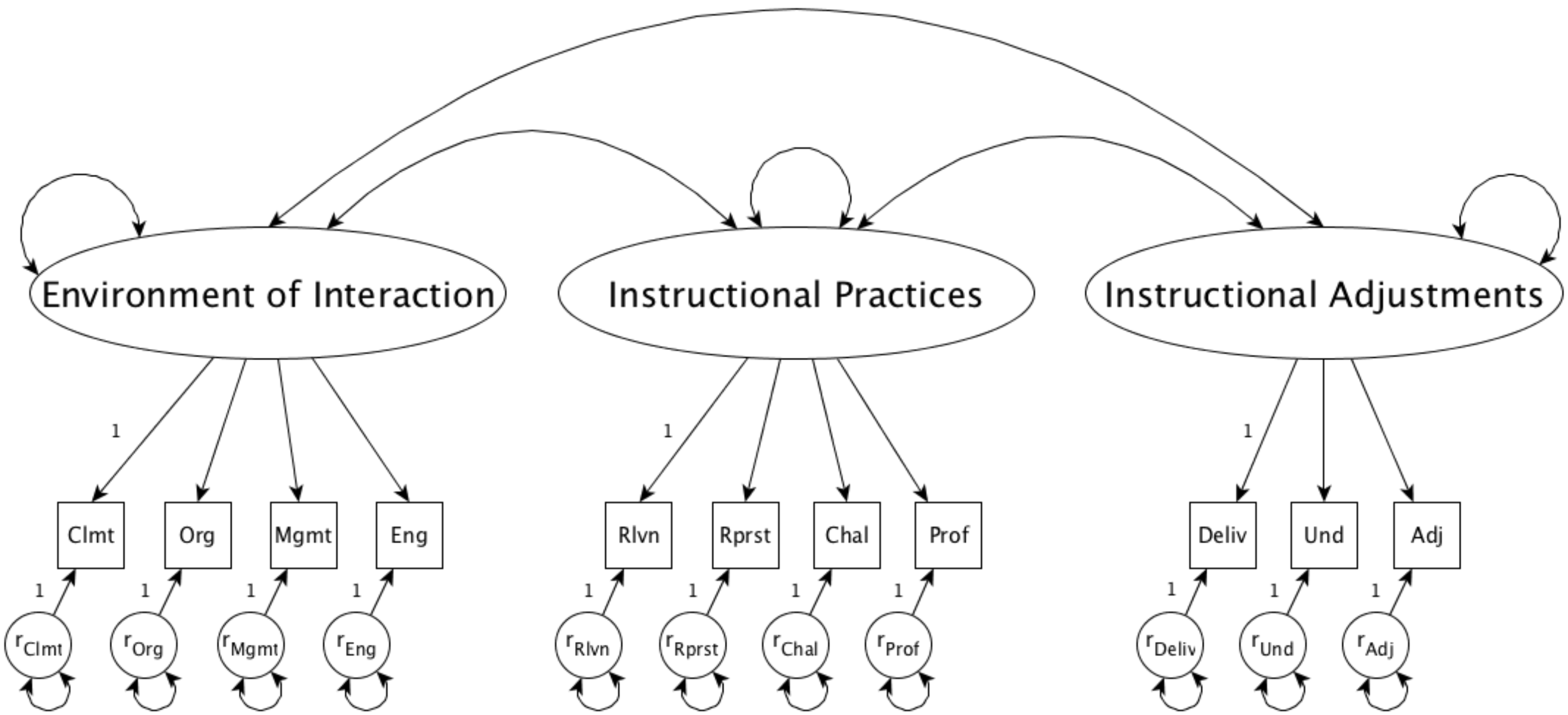

Figure 1. Hypothesized three-factor model of the MSIP-COT observation data. Clmt = Classroom climate; Org = Organization; Mgmt = Classroom management; Eng = Student participation and engagement; Rlvn = Relevance of learning; Rprst = Representation of content; Chal = Intellectual challenge; Prof = Teaching for reading proficiency; Deliv = Clear and consistent delivery of instruction; Und = Checks of student understanding; Adj = Instructional adjustments.

Also evaluated were (a) one- and two-factor models, to test whether instructional context was more accurately represented either as a single factor or as separate environment and instruction factors, with the latter including items representing both instructional practice and instructional adjustments; and (b) an alternative three-factor model based on the framework proposed by the authors of the CLASS (Pianta & Hamre, 2009), with domains for emotional supports, classroom organization, and instructional supports. For this analysis, items on the MSIP-COT were mapped to the dimensions defined on the CLASS based on alignment between key terms and item descriptions. Items representing classroom climate, relevance of learning, and teaching for reading proficiency were categorized as similar to the CLASS emotional supports factor; items representing organization, classroom management, representation of content, and student participation and engagement, were categorized as similar to the CLASS classroom organization factor; and items representing clear and consistent delivery of instruction, checks of student understanding, instructional adjustments, and intellectual challenge were classified as similar to the CLASS instructional supports factor.

Indicators on the MSIP-COT represent negatively skewed ordinal categories. Because results from simulation studies suggest that maximum likelihood estimation may not provide accurate estimates for categorical outcomes measured using fewer than five categories (Kline, 2011; Muthén & Muthén, 2012), all models were estimated using mean and variance adjusted weighted least squares (WLSMV). Model fit was evaluated using a combination of four fit statistics: (a) the χ^2, (b) the comparative fit index (CFI; Hu & Bentler, 1999), (c) the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), and (d) the standardized root mean square residual (SRMR).

Conventional model interpretation guidelines (i.e., non-significant χ^2 statistics, CFI $\geq$ .95, RMSEA < .06, SRMR < .08; Hu & Bentler, 1999) were used, with an emphasis on model fit statistics, given that χ^2 is known to be sensitive to sample size (i.e., with large samples, χ^2 is often significant, even when deviations from perfect fit are minor).

Missing data for individual items on the MSIP-COT was minimal. Two observations (.3%) did not complete any of the 18 items, 18 observations (2.3%) were missing one of the 18 items, and the remaining 780 observations (97.5%) had complete data. Prior to estimation of the measurement models, missing values were estimated using the data imputation command in Mplus to generate and save five imputed data sets (Muthén & Muthén, 2012). Reported results represent parameter estimates averaged across the five data sets, as computed by Mplus. Following evaluation of the measurement models, instructional context factor scores for the complete sample were estimated for use in the structural regression model. Results from estimation of the MSIP-COT factor scores were averaged across instructional grouping and used to answer RQ 2.

SEI measurement model. The authors of the SEI concluded, based on the results of a multiple group CFA, that measurement invariance of their proposed five-factor model of the SEI was appropriate for grades six through 12 and both male and female students (Betts et al., 2010). To confirm measure dimensionality in the current sample of seventh grade reading intervention students (RQ 3), a multiple factor CFA model was fit to the validation sample, as shown in Figure 2. Also tested were one- and two-factor CFA models, to test hypotheses that (a) student engagement is better represented as a single construct, and (b) affective and cognitive engagement can be distinguished as separate constructs without any sub-dimensions.

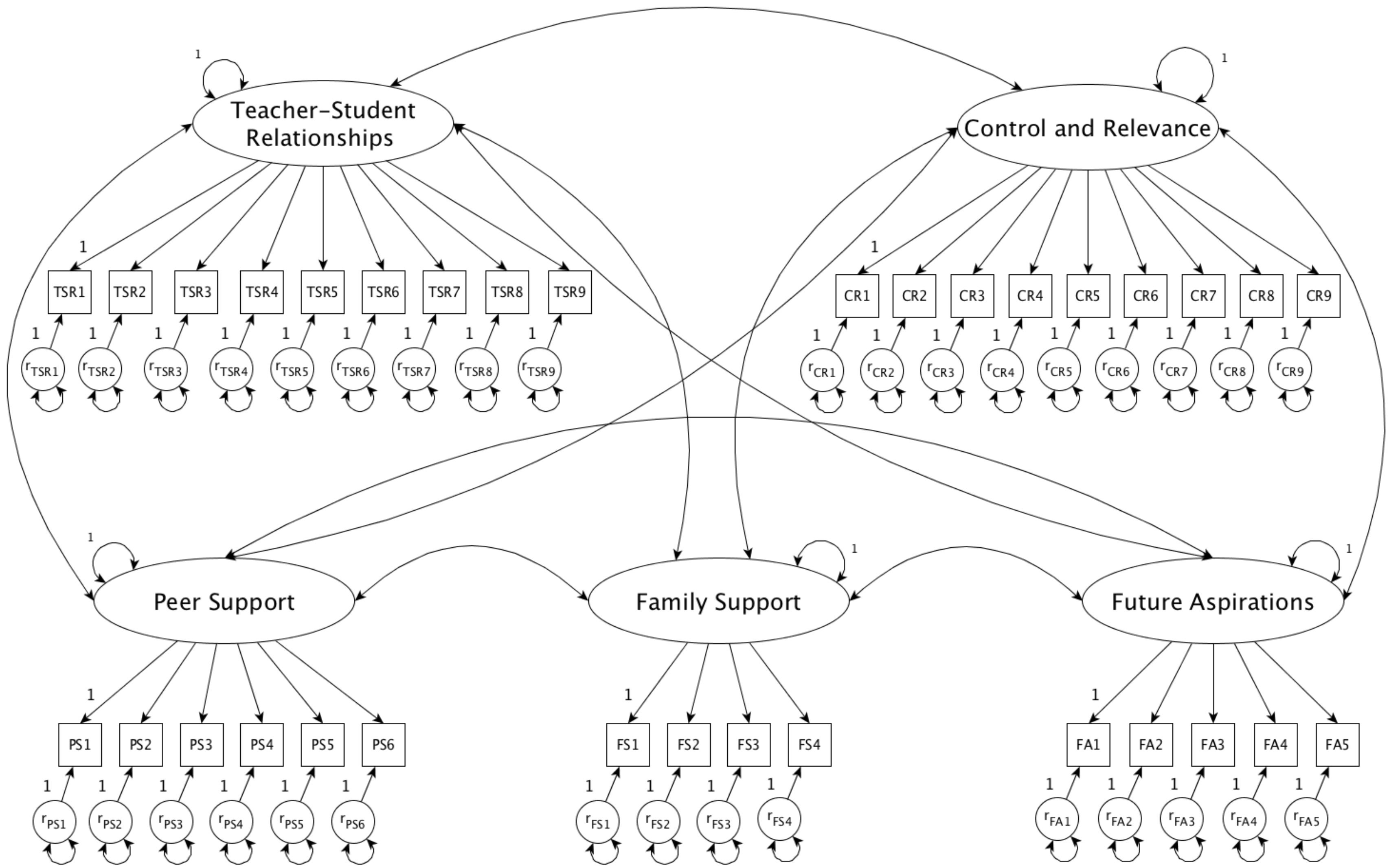

Figure 2. Hypothesized five-factor model of student engagement data. TSR = Teacher-Student Relationships; PS = Peer Support; FS = Family Support; FA = Future Aspirations; CR = Control and Relevance of Schoolwork.

As with the MSIP-COT, each of the five factors was represented with between four and nine negatively skewed ordinal items. Consequently, models were estimated using WLSMV (Kline, 2011; Muthén & Muthén, 2012). Model fit was evaluated using conventional model interpretation guidelines for the χ^2, the CFI, the RMSEA, and the SRMR, with an emphasis on the model fit statistics.

Missing data on the SEI was minimal. Four protocols (.3%) were missing three of the 33 items, 15 protocols (1.0%) were missing two items, 126 protocols (8.7%) were missing one item, and the remaining 1,301 protocols (90.0%) had complete data. Prior to estimation of the measurement models, missing values were estimated using the data imputation command in Mplus to generate and save five imputed data sets (Muthén & Muthén, 2012). Reported results represent parameter estimates averaged across the five data sets, as computed by Mplus. Following evaluation of the measurement model, trait estimates for student engagement were estimated for each student for use in the structural regression model. Results from estimation of the SEI factor scores were used to answer RQ 4.

Structural regression model. The trait estimates of instructional context and student engagement from the measurement models were then included in a multilevel structural regression (SR) model to estimate the direct and mediational impact of instructional practices and engagement on student reading achievement (RQ 5 and 6). However, the process of associating students to ELA and reading intervention classes was complex. By design, the composition of reading instruction for each student was tailored to the perceived needs of that particular student. That is, over the course of the school year and even within a single school day, many students experienced multiple intervention programs, reading classes, and teachers. Further complicating matters, schools did not always meet five days each week, not all instructional groupings met every day, and not all groupings ran for the same length of time each day they met. In an attempt to account for these complexities, a cross-classified multilevel model was specified (Fielding & Goldstein, 2006), in which the factor scores for each instructional grouping were weighted by the proportion of instruction each student received in each grouping, and students were cross-classified by unique combinations of ELA and reading intervention instruction.

The hypothesized multilevel relationship between student engagement and the instructional environment is shown in Figure 3. The student-level model, labeled the within model is on the top, and the instructional grouping-level model, labeled the between model is on the bottom, with a dashed line separating the two models. The SR model was estimated using two-step modeling (Anderson & Gerbing, 1988; Kline, 2011; McDonald, 2010), in which an SR model is respecified as separate CFA measurement and structural models. This allows the fit of the measurement model to be tested independently, avoiding interpretational confounding. Once the measurement model is adequately specified, the fit of the full SR model can be compared to alternate versions of the structural models. Further, the student level model was run as a separate step before the full model was tested. At the student level, the engagement trait estimates for teacher-student relationships, peer support, and family support were specified as indicators of affective engagement, while trait estimates for future aspirations and control and relevance of school were specified as indicators of cognitive engagement.

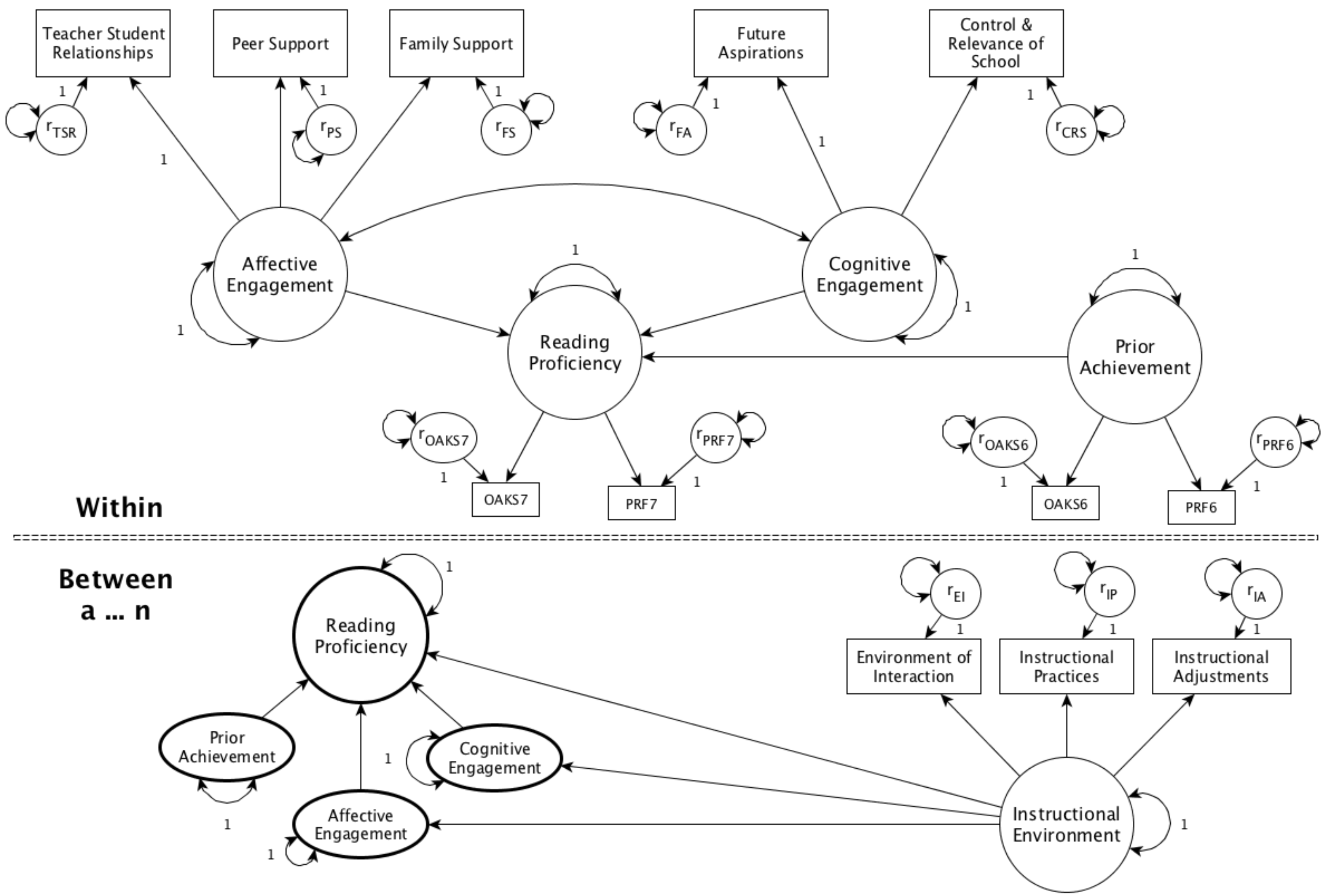

Figure 3. Hypothesized multilevel structural regression model for predicting reading proficiency for struggling readers.

Affective engagement and cognitive engagement were predicted to have direct effects on reading proficiency. Also evaluated were a one-factor CFA model, to test the hypothesis that engagement can be represented as a single factor, and a model with no measurement component, to test whether any of the individual engagement factor scores had direct effects on reading proficiency. Because the indicators for the student engagement factor scores were continuous, maximum likelihood estimation was used. Model fit was evaluated using conventional model interpretation guidelines for the χ^2, the CFI, the RMSEA, and the SRMR, with an emphasis on model fit statistics.

At the instructional grouping level, trait estimates for environment of interactions, instructional practices, and instructional adjustments were specified as indicators of a single latent factor termed “instructional environment.” In addition, student-level parameters modeled as random effects (i.e., those that are hypothesized to vary across instructional groupings) have been outlined in bold. Prior achievement, affective engagement, and cognitive engagement were modeled as random, between-groups effects. Group-level affective and cognitive engagement, the instructional environment, and prior achievement were predicted to directly affect reading proficiency. Engagement was also predicted to partially mediate the effect of the instructional environment. Mediation was tested by estimating the indirect effects of the instructional environment through engagement, using nonparametric bootstrapping to estimate the standard errors of the indirect effects. Because all indicators were continuous, maximum likelihood estimation was used. Model fit was evaluated using conventional model interpretation guidelines for the χ^2, the CFI, the RMSEA, and the SRMR, with an emphasis on model fit statistics.

CHAPTER III

RESULTS

MSIP-COT Measurement Model

Correlations and descriptive statistics for items included in the MSIP-COT measurement model are reported in Table 2. Means and SDs were similar across items, with means ranging from 2.984 to 3.361 (out of four points possible; *mdn* = 3.306), and SDs ranged from .594 to .699 (*mdn* = .671). Correlations between items within a construct ranged from .455 to .618 (*mdn* = .552). Correlations between items across constructs ranged from .327 to .577 (*mdn* = .462). The somewhat lower correlations between most items hypothesized to measure different constructs and somewhat higher correlations between most items within a hypothesized construct provided partial, though highly imperfect (Campbell & Fiske, 1959) evidence that the MSIP-COT measured multiple dimensions of the instructional environment. Notably, however, several correlations between items intended to measure different factors of instruction were higher than anticipated, in some cases exceeding correlations between items intended to measure the same factor.

As shown in Table 3, all four models converged successfully. The two- and three-factor models represented a significant improvement over the one-factor model. The alternative three-factor specification based on the CLASS was a slight improvement over the one-factor model (χ^2 difference = 10.99, $df = 3$), but not the others. Although the differences in model fit between the primary two- and three-factor models were substantially smaller than the differences between the one- and two-factor models, both were statistically and theoretically significant. For instance, the χ^2 difference between the

Table 2

Items in the MSIP-COT Measurement Model by Hypothesized Construct: Correlations and Descriptive Statistics (N = 798)

	Environment of Interactions (EI)				Instructional Practices (IP)				Instructional Adjustments (IA)		
Item	EI1	EI2	EI3	EI4	IP1	IP2	IP3	IP4	IA1	IA2	IA3
EI1 Classroom climate	-										
EI2 Organization	**0.465**	-									
EI3 Classroom mgmt.	**0.618**	**0.618**	-								
EI4 Student engagement	**0.585**	**0.538**	**0.590**	-							
IP1 Relevance of learning	0.343	0.351	0.356	0.327	-						
IP2 Representation of content	0.409	0.485	0.387	0.442	**0.455**	-					
IP3 Intellectual challenge	0.457	0.508	0.483	0.457	**0.546**	**0.569**	-				
IP4 Teaching for proficiency	0.469	0.496	0.520	0.465	**0.546**	**0.546**	**0.552**	-			
IA1 Clear delivery	0.436	0.561	0.472	0.452	0.458	0.566	0.544	0.577	-		
IA2 Checks understanding	0.403	0.427	0.442	0.425	0.374	0.491	0.474	0.521	**0.513**	-	
IA3 Instructional adjustments	0.413	0.478	0.468	0.444	0.393	0.536	0.524	0.540	**0.558**	**0.592**	-
M	3.330	3.321	3.307	3.167	3.001	3.106	2.984	3.353	3.306	3.361	3.175
SD	0.646	0.672	0.699	0.679	0.673	0.661	0.669	0.671	0.683	0.632	0.594
Skew	-0.496	-0.610	-0.593	-0.315	-0.447	-0.511	-0.334	-0.829	-0.830	-0.618	-0.362
Kurt	-0.416	-0.162	-0.457	-0.464	0.547	0.721	0.264	0.680	0.869	0.154	0.963

Note. Within-construct correlations are in bold. All correlations are significantly different than zero, $p < .001$. Cronbach's alpha coefficient for the 11 items = .911

one- and two- factor models of 130.05 (*df* = 1) and the χ^2 difference between the two- and three-factor models of 25.64 (*df* = 2) each represented statistically significant ($p < .001$) improvements in model fit.

Table 3

Goodness-of-Fit Indicators for Models of the Instructional Environment (n = 399)

Fit Statistic	One-Factor	Two-Factor	Three-Factor	Alternative Three-Factor
χ^2	300.909	170.862	145.227	289.919
df	44	43	41	41
χ^2 *diff*	-	53.248	22.131	-
CFI	0.957	0.979	0.983	0.958
RMSEA	0.121	0.086	0.080	0.123
SRMR	0.076	0.054	0.049	0.075

Note. df = degrees of freedom; χ^2 *diff* = χ^2 *diff*erence; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All fit statistics (except the χ^2 *diff*) are weighted least squares mean and variance adjusted averages based on five multiply imputed data sets. The χ^2 *diff* was computed on the first imputed data set using the second order correction for the χ^2 statistic as recommended in Mplus (Asparouhov & Muthén, 2010). $p < .001$.

Similarly, the CFI indicated that while the one-factor model explained 95.7% of the unexplained variance in the baseline model, the two- and three-factor models explained additional variance (97.9% and 98.3%, respectively). The RMSEA and SRMR were also substantially smaller for the two- and three-factor models. Although the RMSEA for the three-factor model was .080 (PCLOSE < .001), which did not meet the *a priori* decision rule of .05, the SRMR of .049 exceeded the *a priori* decision rule of .08. The χ^2 for all four models was statistically significant. However, Hoelter's Critical N (Hoelter, 1983) for the three-factor model was 157, indicating that the χ^2 for this model would have been significant even with half as many observations. Model-implied structure coefficients for the three-factor model provided moderate support for the discriminant validity of the MSIP-COT, which for off-dimension items were between .516 and .780 while on-pattern coefficients ranged from .668 to .884. All parameter estimates were statistically significant ($p < .001$), indicating that these values were different from zero. Taken together, these results show that the three-factor model was the best representation of the MSIP-COT of the models tested, although the RMSEA showed that the fit was slightly less than ideal. Standardized parameter estimates for the three-factor model are reported in Figure 4.

An examination of the modification indices (MI) provided by Mplus revealed five item cross-loadings and eight correlated error terms that would have reduced the χ^2 value by at least 10. Two cross-loadings were between factors for the environment of interactions and quality of instructional practices; two were between instructional practices and the use of instructional adjustments; and one was between environment of interactions and the use of instructional adjustments. Items for organization and relevance of learning were each listed twice, indicating that they loaded relatively highly on all three instructional factors. Further, the two highest MI for cross-loadings were for the cross-loadings of those items on the use of instructional adjustments. The MI for the loading of organization on use of instructional adjustments was 26.812, while the MI for the loading of relevance of learning on use of instructional adjustments was 29.713. The highest MI for correlated error terms was between the items for instructional adjustments and checks for student understanding (MI = 34.134). Future research should consider the extent to which these modifications are supported by theory.

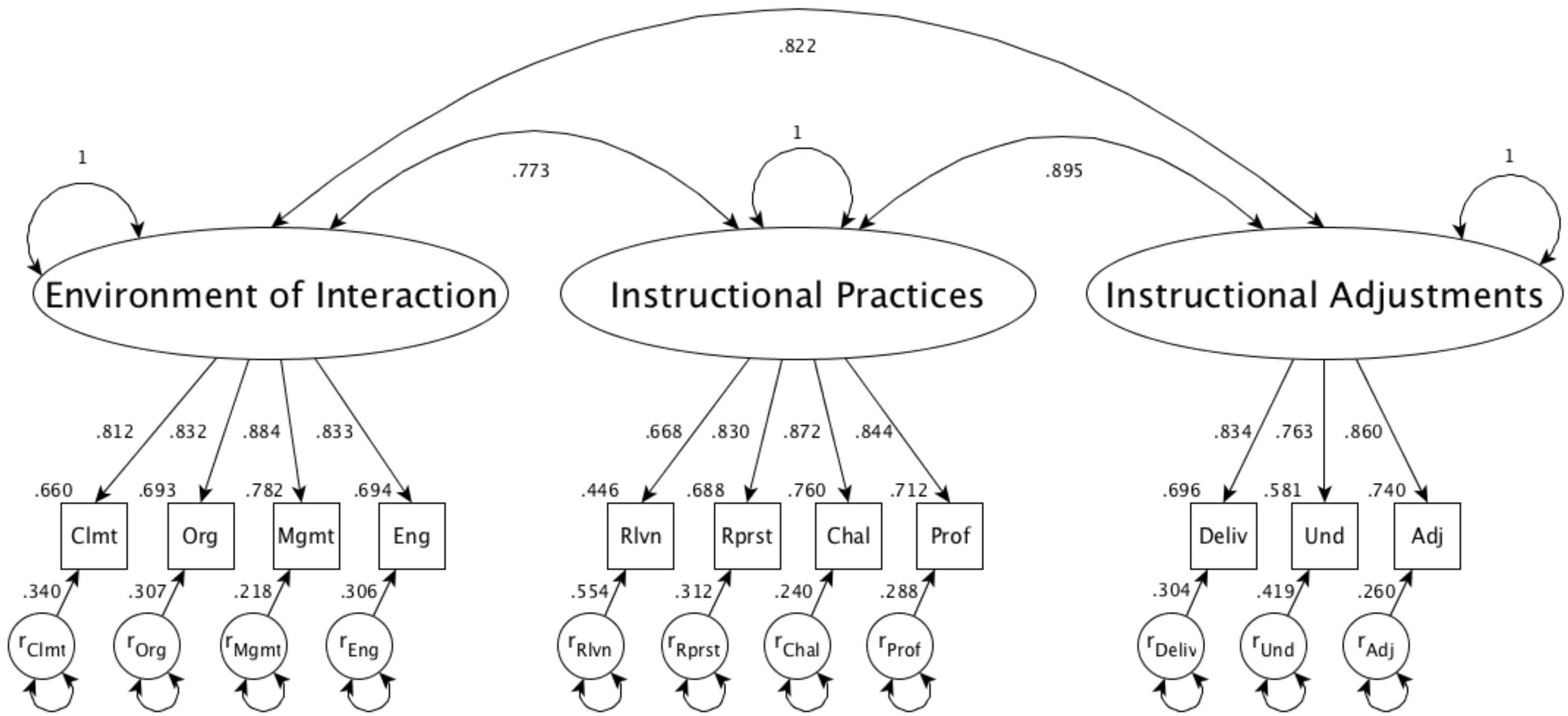

Figure 4. Final three-factor model of the MSIP-COT observation data. Values underneath double headed arrows represent correlations between factors. Values for arrows pointing from latent variables to measured variables represent correlations between items and factors. Values in the upper left corner of each observed variable represent r^2 for measured variables. Values between residual variables and observed variables represent proportions of unexplained variance.

Variability of MSIP-COT factor scores. Averaged across instructional groupings, factor score standard deviations for the environment of interactions, instructional practices, and instructional adjustments were .59, .49, and .61, respectively, which suggests that (a) the measure does capture variability between instructional groups, as intended; and (b) the variability across the three factors is similar. For the environment of interactions factor, 80% of instructional groupings had an average score between -.91 and .67. Similarly, 80% of instructional groupings had an average score for the instructional practices factor between -.67 and .57, and between -.80 and .74 for the instructional adjustments factor.

SEI Measurement Model

Correlations and descriptive statistics for items included in the SEI measurement model are reported in Table 4. Means and SDs were fairly consistent across items. Means ranged from 2.77 to 3.69 (*mdn* = 3.23), and SDs ranged from .531 to .845 (*mdn* = .698). Correlations between items within a construct displayed substantial variation, ranging from .171 to .637 (*mdn* = .397), while correlations between items across constructs ranged from .087 to .531 (*mdn* = .218). Low to moderate correlations between most items hypothesized to measure different constructs and moderate to large correlations between items within a construct provided partial, though highly imperfect (Campbell & Fiske, 1959) evidence that the SEI measured multiple dimensions of student engagement. As with the MSIP-COT, several correlations between items intended to measure different factors of engagement were higher than anticipated, in some cases exceeding correlations between items intended to measure the same factor.

Table 4

Items in the SEI Measurement Model by Hypothesized Construct: Correlations and Descriptive Statistics (N = 1446)

	Teacher-student relationships (TSR)								
Item	3	5	10	13	16	21	22	27	31
3. My teachers are there for me ...	–								
5. Adults ... listen to the students.	**.503**	–							
10. The school rules are fair.	**.368**	**.372**	–						
13. Most teachers at my school ...	**.445**	**.392**	**.330**	–					
16. Overall, my teachers are ...	**.516**	**.479**	**.395**	**.506**	–				
21. Adults ... treat students fairly.	**.480**	**.535**	**.468**	**.421**	**.534**	–			
22. I enjoy talking to ... teachers ...	**.444**	**.347**	**.393**	**.410**	**.465**	**.459**	–		
27. I feel safe at school.	**.339**	**.360**	**.340**	**.312**	**.392**	**.437**	**.333**	–	
31. ... teachers care about students.	**.511**	**.534**	**.386**	**.484**	**.569**	**.602**	**.532**	**.420**	–
4. Other students here like me ...	.212	.166	.136	.172	.193	.160	.152	.350	.162
6. … students ... care about me.	.212	.157	.136	.225	.197	.148	.148	.335	.202
7. Students ... are there for me ...	.221	.172	.156	.217	.225	.188	.161	.340	.206
14. Students here respect what I ...	.240	.275	.217	.267	.269	.305	.195	.402	.258
23. I enjoy talking to ... students ...	.117	.128	.107	.079**	.188	.168	.166	.320	.156
24. I have some friends at school.	.134	.121	.087	.104	.119	.081**	.111	.231	.121
1. My family ... is there for me ...	.183	.212	.213	.168	.192	.245	.129	.222	.141
12. When something good ...	.167	.169	.220	.215	.185	.213	.148	.178	.161
20. When I have problems ...	.201	.205	.245	.223	.259	.290	.220	.264	.236
29. My family ... wants me to ...	.207	.161	.174	.168	.203	.245	.249	.225	.223

Table 4

Continued

Item	Teacher-student relationships (TSR)								
	3	5	10	13	16	21	22	27	31
2. After finishing schoolwork ...	.234	.183	.254	.216	.199	.226	.199	.208	.192
9. Most of what is important ...	.312	.341	.302	.231	.340	.350	.300	.234	.307
15. When I do schoolwork ...	.232	.185	.255	.205	.245	.271	.248	.234	.235
25. When I do well in school ...	.229	.205	.246	.171	.234	.255	.259	.250	.214
26. The tests in my classes ...	.312	.324	.266	.297	.291	.362	.303	.312	.376
28. I feel like I have a say about ...	.207	.216	.206	.237	.223	.296	.222	.269	.269
33. Learning is fun because I get ...	.271	.239	.312	.293	.297	.352	.388	.292	.344
34. What I'm learning in my ...	.215	.204	.231	.246	.286	.277	.300	.227	.283
35. The grades in my classes ...	.305	.303	.283	.292	.333	.354	.290	.310	.372
8. My education will create ...	.190	.197	.176	.166	.224	.225	.216	.181	.192
11. Going to school after high ...	.199	.161	.211	.189	.229	.257	.257	.183	.197
17. I plan to continue ...	.200	.155	.189	.153	.227	.234	.287	.215	.184
19. School is important for ...	.185	.184	.195	.213	.251	.237	.275	.207	.251
30. I am hopeful about my future.	.219	.169	.166	.167	.205	.240	.239	.192	.226
M	3.19	3.14	2.88	2.91	3.21	3.06	3.10	3.12	3.23
SD	.694	.731	.845	.762	.698	.761	.740	.786	.694

Table 4

Continued

Item	Peer support at school (PS)						Family support for learning (FS)			
	4	6	7	14	23	24	1	12	20	29
4. Other students here like me ...	–									
6. … students ... care about me.	**.594**	–								
7. Students ... are there for me ...	**.532**	**.637**	–							
14. Students here respect what I ...	**.498**	**.537**	**.539**	–						
23. I enjoy talking to ... students ...	**.478**	**.476**	**.491**	**.426**	–					
24. I have some friends at school.	**.390**	**.385**	**.383**	**.297**	**.463**	–				
1. My family ... is there for me ...	.208	.231	.230	.229	.153	.188	–			
12. When something good ...	.167	.192	.211	.237	.145	.135	**.424**	–		
20. When I have problems ...	.173	.190	.212	.195	.200	.222	**.508**	**.477**	–	
29. My family ... wants me to ...	.206	.180	.228	.161	.243	.245	**.361**	**.364**	**.405**	–
2. After finishing schoolwork ...	.111	.121	.126	.173	.100	.053*	.195	.217	.195	.155
9. Most of what is important ...	.131	.105	.094	.162	.135	.106	.222	.179	.216	.260
15. When I do schoolwork ...	.198	.191	.221	.243	.220	.156	.284	.284	.290	.316
25. When I do well in school ...	.212	.166	.216	.193	.262	.217	.260	.272	.323	.398
26. The tests in my classes ...	.187	.179	.170	.265	.198	.144	.217	.187	.228	.258
28. I feel like I have a say about ...	.126	.152	.167	.209	.140	.151	.165	.232	.252	.168
33. Learning is fun because I get ...	.163	.150	.183	.223	.235	.139	.198	.197	.282	.312
34. What I'm learning in my ...	.194	.169	.180	.220	.280	.214	.203	.210	.243	.383
35. The grades in my classes ...	.167	.122	.154	.243	.200	.110	.231	.179	.269	.263

Table 4

Continued

Item	Peer support at school (PS)						Family support for learning (FS)			
	4	6	7	14	23	24	1	12	20	29
8. My education will create ...	.153	.148	.168	.174	.189	.192	.306	.263	.301	.331
11. Going to school after high ...	.171	.180	.211	.192	.249	.240	.283	.221	.259	.355
17. I plan to continue ...	.173	.182	.258	.214	.276	.211	.276	.272	.279	.367
19. School is important for ...	.172	.154	.202	.165	.271	.239	.223	.220	.294	.420
30. I am hopeful about my future.	.172	.186	.160	.189	.259	.264	.249	.214	.286	.351
M	3.18	3.07	3.07	2.86	3.48	3.69	3.62	3.42	3.50	3.64
SD	.768	.754	.803	.767	.684	.557	.565	.706	.664	.531

Item	Control and relevance of schoolwork (CR)								
	2	9	15	25	26	28	33	34	35
2. After finishing schoolwork ...	–								
9. Most of what is important ...	**.196**	–							
15. When I do schoolwork ...	**.395**	**.263**	–						
25. When I do well in school ...	**.248**	**.296**	**.357**	–					
26. The tests in my classes ...	**.194**	**.325**	**.299**	**.265**	–				
28. I feel like I have a say about ...	**.174**	**.175**	**.171**	**.177**	**.261**	–			
33. Learning is fun because I get ...	**.301**	**.374**	**.354**	**.380**	**.345**	**.236**	–		
34. What I’m learning in my ...	**.212**	**.406**	**.317**	**.448**	**.392**	**.224**	**.457**	–	
35. The grades in my classes ...	**.243**	**.335**	**.275**	**.325**	**.509**	**.248**	**.397**	**.407**	–

Table 4

Continued

Item	Control and relevance of schoolwork (CR)								
	2	9	15	25	26	28	33	34	35
8. My education will create ...	.203	.305	.336	.411	.277	.136	.282	.400	.273
11. Going to school after high ...	.193	.333	.268	.328	.239	.159	.362	.462	.267
17. I plan to continue ...	.174	.262	.305	.397	.260	.167	.345	.441	.248
19. School is important for ...	.186	.349	.327	.402	.314	.188	.392	.531	.307
30. I am hopeful about my future.	.162	.274	.295	.418	.261	.174	.318	.420	.295
M	2.77	3.34	3.23	3.55	3.19	2.95	3.17	3.54	3.23
SD	.709	.636	.695	.603	.714	.842	.766	.635	.794

Item	Future Aspirations (FA)				
	8	11	17	19	30
8. My education will create ...	–				
11. Going to school after high ...	**.408**	–			
17. I plan to continue ...	**.453**	**.603**	–		
19. School is important for ...	**.466**	**.472**	**.525**	–	
30. I am hopeful about my future.	**.438**	**.390**	**.446**	**.448**	–
M	3.57	3.60	3.65	3.62	3.62
SD	.598	.613	.566	.577	.583

Note. Within-construct correlations are in bold. All correlations are significantly different than zero, $p < .001$, unless otherwise marked. Cronbach's alpha coefficient for the 33 items = .920.

$^*p < .05$. $^{**}p < .01$.

As shown in Table 5, all three models converged successfully. All WLSMV fit statistics indicated that (a) the two-factor model represented a significant improvement in model fit over the one-factor model, and (b) the five-factor model represented a significant improvement in model fit over both of the other models. For instance, the χ^2 difference of 890.20 (*df* = 1) for the two-factor model and the χ^2 difference of 2621.92 (*df* = 9) for the five-factor model each represented statistically significant ($p < .001$) improvements in model fit. Similarly, the CFI for the one-factor model explained 73.8% of the unexplained variance in the baseline model, whereas the five-factor model explained 94.7% of the unexplained variance, an increase of 21%. In addition, the RMSEA (.05) and SRMR (.058) for the five-factor model met the *a priori* decision rules of .06 and .08 respectively, whereas the one- and two-factor models did not. The *p* of close fit for the RMSEA was .488, further indicating relatively close model fit.

Table 5

Goodness-of-Fit Indicators for Models of Student Engagement (n = 723)

Fit Statistic	One-Factor	Two-Factor	Five-Factor
χ^2	4874.440	3984.236	1362.312
df	495	494	485
χ^2 *diff*	–	204.009	778.062
CFI	0.738	0.791	0.947
RMSEA	0.111	0.099	0.050
SRMR	0.116	0.107	0.058

Note. df = degrees of freedom; χ^2 *diff* = χ^2 *diff*erence; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. All fit statistics (except the χ^2 *diff*) are weighted least squares mean and variance adjusted averages based on five multiply imputed data sets. The χ^2 *diff* was computed on the first imputed data set using the second order correction for the χ^2 statistic as recommended in Mplus (Asparouhov & Muthén, 2010). $p < .001$.

Although the χ^2 for all three models was statistically significant, Hoelter's Critical N for the three-factor model was 285, which indicates that the χ^2 would have been significant even with half as many students. It also provides another indication that overall model fit was acceptable (Hoelter, 1983). The discriminant validity of the five-factor model was partially supported by the model-implied structure coefficients, which were below .500 for 99 of 132 off-dimension items (75%), while on-pattern coefficients exceeded .500 for 31 of 33 items (94%). All parameter estimates were statistically significant ($p < .001$), indicating that these values were different from zero. Together, these results show that the five-factor model was the best representation of the SEI of the models tested. Standardized parameter estimates for the five-factor model are reported in Figure 5.

An examination of the MI provided by Mplus revealed 64 item cross-loadings and 53 correlated error terms that would have reduced the χ^2 value by at least 10. Of the 64 item cross-loadings, 23 were for items hypothesized to represent control and relevance, 18 were for items hypothesized to represent peer support, 15 were for items hypothesized to represent teacher student relationships, six were for items hypothesized to represent peer support, and two were for items hypothesized to represent future aspirations. Six items were listed four times, seven items were listed three times, and six items were listed twice, indicating that these items loaded relatively highly on five, four, and three different factors, respectively.

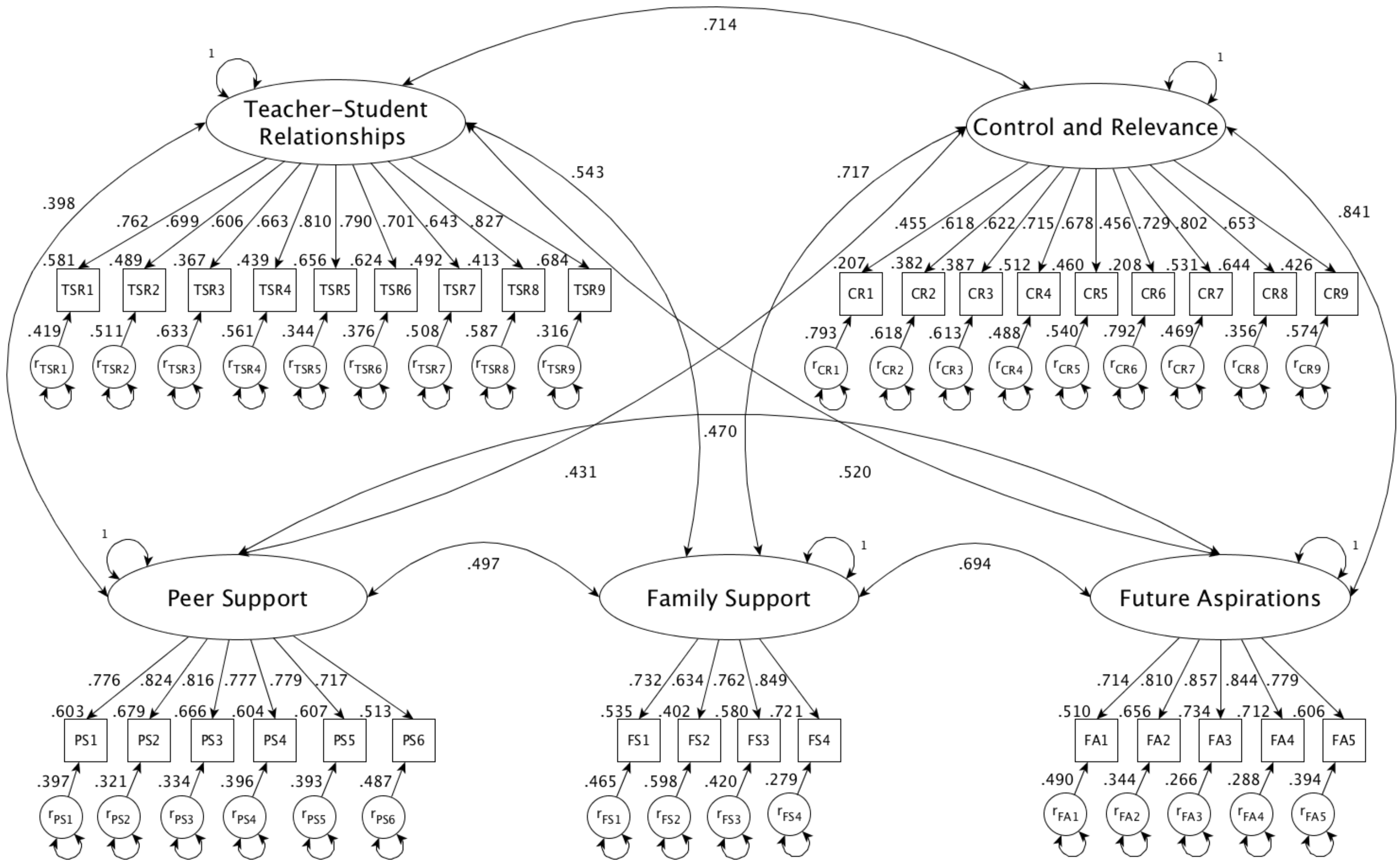

Figure 5. Final five-factor model of student engagement data with standardized factor loadings. TSR = Teacher-Student Relationships; PS = Peer Support; FS = Family Support; FA = Future Aspirations; CR = Control and Relevance of Schoolwork

The highest MI was for the cross loading of “I feel safe at school” (MI = 353.041), which was hypothesized to represent teacher-student relationships but also loaded highly on peer support at school. The second highest MI for cross-loadings was for “Students here respect what I say” (MI = 127.412), which was hypothesized to represent peer support at school, but also loaded highly on teacher-student relationships. The next two highest MI for cross-loadings were for items hypothesized to represent control and relevance that also loaded highly on future aspirations (MIs = 109.141 and 103.686). Of the 53 correlated error terms, 20 were between items hypothesized to represent the same factor and 37 were between items hypothesized to represent the same higher order factor (i.e., affective or cognitive engagement). Two of the three highest MI for correlated error terms were between items hypothesized to represent control and relevance (MIs = 80.004 and 79.488). Future research should consider the extent to which these modifications are supported by theory.

Variability of SEI factor scores. Across students, standard deviations for factor scores for teacher-student relationships, peer support, family support, control and relevance, and future aspirations were .69, .69, .61, .42, and .63, respectively, showing that students differed slightly more on their ratings of teacher-student relationships and peer support than on the other factors. For the teacher-student relationships and peer support factors, 80% of students had an average score between -.91 and .91, and -.90 and .90, respectively. Similarly, 80% of students had average scores between -.84 and .73 for the family support factor, between -.52 and .52 for the control and relevance factor, and between -.87 and .74 for the future aspirations factor.

Structural Regression Model

Students received reading instruction in an average of 2.43 (*SD* = .92, *range* = 1 to 6) total instructional groupings over the course of the academic year. Nearly three quarters received all of their reading instruction in either two (*n* = 675, 46.7%) or three (*n* = 398, 27.5%) classes. By definition, all students in the current study received at least one class that was identified as a reading intervention, and the majority participated in either one (*n* = 862, 59.6%) or two (*n* = 432, 29.9%) intervention classes. Students received an average of approximately 92 instructional days in a given intervention (*SD* = 48.10) and 134 total instructional days in at least one intervention class (*SD* = 45.67). That is, on average, students participated in a reading intervention for about three quarters of the school year. More than 70% of students (*n* = 1047, 72.4%) attended a single ELA class. However, nearly a fifth of all participants (*n* = 276, 19.1%) did not receive any ELA instruction. The average length of attendance in ELA was about 143 days (*SD* = 43.80).

Correlations and descriptive statistics for items included in the structural regression model are reported in Tables 6-8. At the student level, correlations between engagement factor scores were moderate to high, as expected. The magnitudes of correlations between achievement scores were similar, but most correlations between engagement and achievement were not statistically significant, $p > .05$. Similar patterns were observed at both the ELA and reading intervention group levels (i.e., only variables measuring the same basic construct were significantly correlated.

Table 6

Items in the Structural Regression Model: Student Level Correlations and Descriptive Statistics (N = 1446)

Variable	1	2	3	4	5	6	7	8	9	10
Student level										
1. TSR factor	–									
2. PS factor	.486	–								
3. FS factor	.620	.568	–							
4. CR factor	.806	.526	.814	–						
5. FA factor	.596	.521	.812	.910	–					
6. PRF grade 6	-.008	.052	.032	.027	.056	–				
7. PRF grade 7	.010	.040	.047	.019	.045	.506	–			
8. OAKS grade 6	.011	.064	.026	.022	.045	.314	.137	–		
9. OAKS grade 7	.012	.022	.043	.010	.016	.106	.615	.087	–	
10. RPCPA	.009	-.031	-.059	-.020	-.030	-.438	-.468	-.415	-.625	–
M	-0.004	-0.015	-0.030	-0.011	-0.033	117.642	120.473	215.043	214.131	1.054
SD	0.691	0.691	0.612	0.415	0.634	41.660	41.543	31.179	48.830	2.939

Note. TSR = Teacher-Student Relationships; PS = Peer Support; FS = Family Support; CR = Control and Relevance; FA = Future Aspirations; PRF = Passage Reading Fluency; OAKS = Oregon Assessment of Knowledge and Skills; RPCPA = Reading Proficiency Controlling for Prior Achievement.

$p < .05$ for all $r \geq |.051|$, $p < .01$ for all $r \geq |.068|$, $p < .001$ for all $r \geq |.087|$.

Table 7

Items in the Structural Regression Model: ELA Classroom Level Correlations and Descriptive Statistics (N = 128)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
English language arts group level													
1. TSR factor	–												
2. PS factor	.531	–											
3. FS factor	.683	.679	–										
4. CR factor	.874	.566	.836	–									
5. FA factor	.672	.549	.854	.909	–								
6. PRF grade 6	-.032	-.003	.050	.036	.055	–							
7. PRF grade 7	-.077	-.006	-.026	-.068	-.019	.782	–						
8. OAKS grade 6	-.243	-.180	-.145	-.139	-.024	.405	.358	–					
9. OAKS grade 7	-.152	-.090	-.085	-.137	-.015	.390	.473	.554	–				
10. RPCPA	.045	.075	-.005	-.071	-.046	-.193	.229	-.421	.416	–			
11. EI factor	.086	-.107	.006	.056	-.005	.169	.143	.177	.038	-.139	–		
12. IP factor	.028	-.118	-.013	.003	-.037	.168	.155	.125	.048	-.079	.912	–	
13. IA factor	.043	-.150	-.025	.015	-.026	.172	.159	.158	.045	-.109	.933	.980	–
M	-0.042	-0.063	-0.031	-0.032	-0.051	126.44	130.90	220.29	226.12	0.001	-0.003	-0.004	-0.037
SD	0.500	0.533	0.409	0.293	0.441	22.796	20.054	5.030	5.378	0.414	0.505	0.421	0.513

Note. TSR = Teacher-Student Relationships; PS = Peer Support; FS = Family Support; CR = Control and Relevance; FA = Future Aspirations; PRF = Passage Reading Fluency; OAKS = Oregon Assessment of Knowledge and Skills; RPCPA = Reading Proficiency Controlling for Prior Achievement; EI = Environment of Interactions; IP = Instructional Practices; IA = Instructional Adjustments. $p < .05$ for all $r \geq |.174|$, $p < .01$ for all $r \geq |.228|$, $p < .001$ for all $r \geq |.289|$.

Table 8

Items in the Structural Regression Model: RI Classroom Level Correlations and Descriptive Statistics (N = 420)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
Reading intervention group level													
1. TSR factor	–												
2. PS factor	.468	–											
3. FS factor	.602	.564	–										
4. CR factor	.801	.510	.813	–									
5. FA factor	.531	.514	.807	.882	–								
6. PRF grade 6	.084	.116	.151	.150	.152	–							
7. PRF grade 7	.095	.114	.145	.166	.185	.817	–						
8. OAKS grade 6	-.055	-.092	.045	-.032	.007	.416	.366	–					
9. OAKS grade 7	.039	-.037	.076	.043	.083	.413	.446	.640	–				
10. RPCPA	.096	.052	.023	.084	.098	-.162	.225	-.394	.338	–			
11. EI factor	-.008	.046	.053	-.006	-.034	-.002	.013	.045	.051	.017	–		
12. IP factor	.027	.076	.075	.039	.016	.003	.021	-.008	.026	.047	.895	–	
13. IA factor	.020	.057	.062	.021	-.009	.011	.004	.007	.014	.000	.915	.978	–
M	-0.084	-0.072	-0.047	-0.048	-0.082	120.30	124.29	219.03	224.71	-0.018	-0.058	-0.018	-0.033
SD	0.600	0.595	0.507	0.342	0.533	29.838	27.437	6.328	6.609	0.468	0.473	0.369	0.461

Note. TSR = Teacher-Student Relationships; PS = Peer Support; FS = Family Support; CR = Control and Relevance; FA = Future Aspirations; PRF = Passage Reading Fluency; OAKS = Oregon Assessment of Knowledge and Skills; RPCPA = Reading Proficiency Controlling for Prior Achievement; EI = Environment of Interactions; IP = Instructional Practices; IA = Instructional Adjustments. $p < .05$ for all $r \geq |.096|$, $p < .01$ for all $r \geq |.126|$, $p < .001$ for all $r \geq |.161|$.

At the student level, an initial one-factor CFA with all five of the engagement factor scores on a single dimension was a poor fit to the data (CFI = .873, RMSEA = .319, SRMR = .050). A two-factor CFA, with three factor score indicators of affective engagement and two factor score indicators of cognitive engagement fit the data slightly worse than the one-factor model (CFI = .874, RMSEA = .349, SRMR = .047). An alternative specification using the item-level SEI scores to model a second-order factor structure where the factor scores specified in the five-factor model loaded onto second-order factors for affective and cognitive engagement provided a better fit (CFI = .941, RMSEA = .053, SRMR = .058).

Adding latent factors for both prior achievement and reading proficiency resulted in a psi (ψ) matrix that was not positive definite. Although this may have occurred for a number of reasons, one possibility was that the latent factors for achievement were too highly correlated; observed correlations between variables comprising prior achievement and reading proficiency ranged from $r(1331) = .371$, $p < .001$ for the relationship between sixth grade OAKS and seventh grade PRF to $r(1298) = .815$, $p < .001$ for the relationship between PRF in sixth and seventh grade. To simplify model estimation, a composite score for reading proficiency controlling for prior achievement was created by computing the difference between composite scores for reading proficiency and prior achievement. A second-order factor model using this composite difference score converged successfully and displayed a model fit similar to the previous second-order CFA (CFI = .941, RMSEA = .052), but neither affective engagement nor cognitive engagement were statistically significant predictors of 7^{th} grade reading proficiency controlling for prior achievement ($p > .05$).

At the group level, the cross-classified model including predictors for English Language Arts instruction and reading intervention instruction failed to converge, both with and without the engagement factors in the model. To test whether reading proficiency controlling for prior achievement could be predicted using only the instructional environment factor scores for reading intervention classes, a simplified multilevel model was specified in which reading proficiency controlling for prior achievement was regressed on environment of interactions, instructional practices, and instructional adjustments. The regression path for the environment of interactions factor was not significant ($p > .05$), but regression paths for instructional practices and instructional adjustments were.

The unstandardized estimate for instructional practices was 1.05 ($p < .001$), indicating that a one point increase in the instructional practices score for a reading intervention class resulted in a predicted increase in reading proficiency controlling for prior achievement of 1.05 points, holding all other instructional environment scores constant. In contrast, the unstandardized estimate for instructional adjustments was -0.91 ($p < .001$), indicating that a one point increase in the instructional adjustments score for a reading intervention class resulted in a predicted decrease in reading proficiency controlling for prior achievement of 0.91 points, again holding all other instructional environment scores constant. Standardized parameter estimates for the five-factor model are reported in Figure 6. According to this model, reading intervention classrooms accounted for 22% of the variance in reading proficiency controlling for prior achievement.

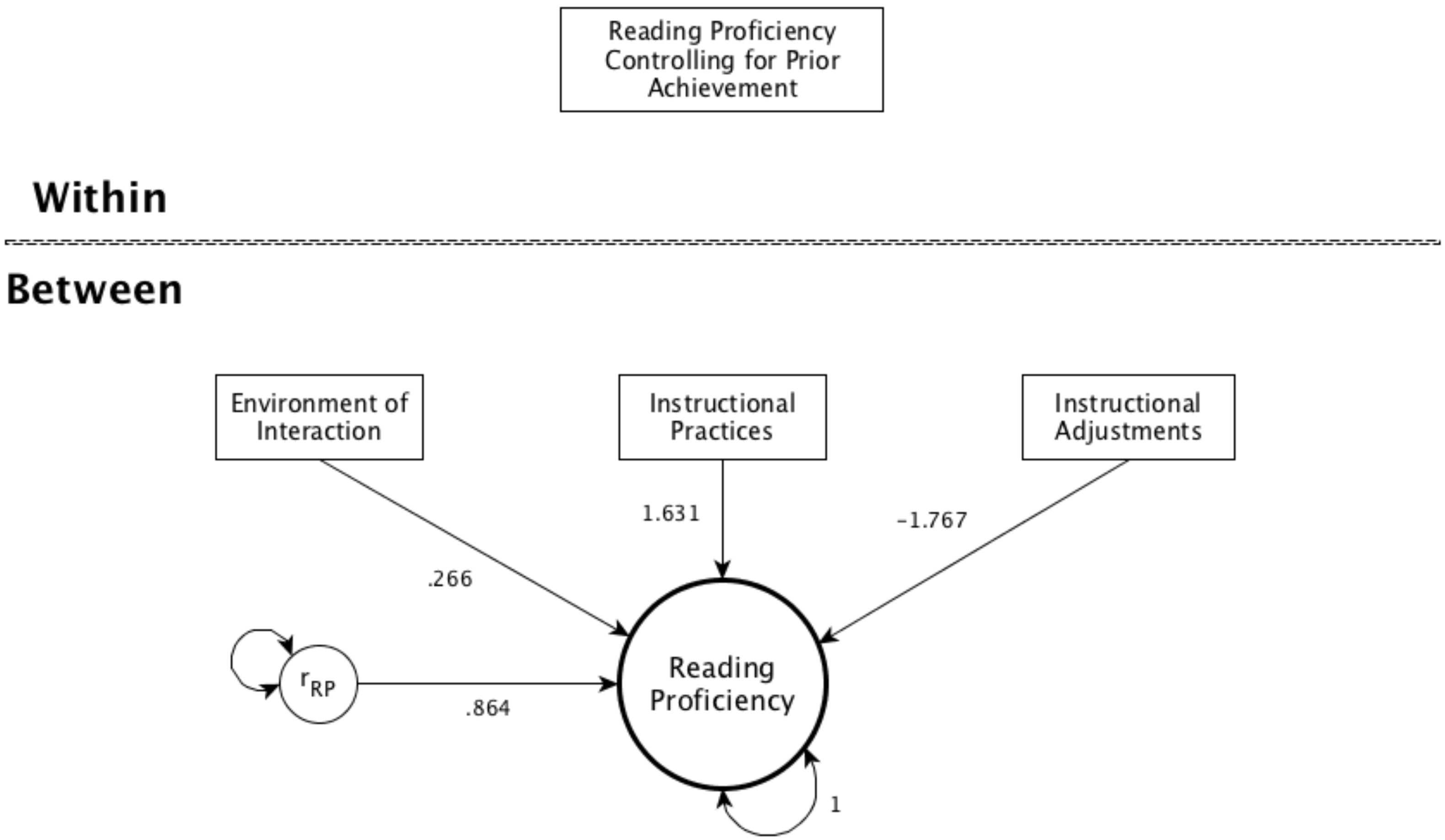

Figure 6. Final multilevel structural regression model predicting reading proficiency for struggling readers with standardized regression coefficients. Values for arrows pointing from latent variables to measured variables represent correlations between items and factors. Values between residual variables and observed variables represent proportions of unexplained variance.

CHAPTER IV

DISCUSSION

This study evaluated the construct validity of two measures: (a) an observation tool designed to measure the instructional environment of English language arts (ELA) and reading intervention classrooms, and (b) a survey designed to measure students' perceptions of their engagement in school, including measures of support from family and peers. Previous research has found that direct observation of instruction both provides important information about the effectiveness of instructional practice (Pianta & Hamre, 2009) and shows promise as a predictor of student achievement gains (Kane & Staiger, 2012; Pianta & Hamre, 2009). Results of the current study demonstrated that the Middle School Intervention Project Classroom Observation Tool (MSIP-COT) can be used to measure variability between teachers in instructional practices, but also that additional development and refinement of the tool is necessary. The current study also found that, consistent with previous research (Appleton et al., 2006; Betts et al., 2010), differences in self-reported affective and cognitive engagement can be measured using a student-completed survey of engagement, the Student Engagement Instrument (SEI), although it too could be refined with additional research.

This study also used the results of the MSIP-COT and SEI measurement models to evaluate the extent to which self-reported student engagement mediates the relationship between the instructional environment and reading achievement for middle school students with reading difficulties. Prior research has found relationships between various components of instructional context, student engagement, and academic performance (Downer et al., 2007; Fredricks et al., 2004; Hamre & Pianta, 2005).

Although evidence regarding these relationships is most robust with respect to behavioral engagement for elementary school children, other relationships have also been observed. For instance, research has correlated (a) decreased affective engagement for students in middle school with similar decreases in the quality of teacher-student relationships relative to elementary school students (Eccles et al., 1993), and (b) increased cognitive engagement with classrooms that have higher ratings of intellectual challenge (Fredricks et al., 2004). Limited research with middle school students on engagement as a mediator of the instructional context provides evidence of a relationship between instructional context and behavioral engagement, but also suggests that (a) affective and cognitive engagement may not measurably mediate the effects of differences in instructional context, and (b) changes in behavioral engagement may not translate to improvements in academic outcomes, especially for students identified as at-risk (Dotterer & Lowe, 2011).

Consistent with previous research, the current study found that (a) higher ratings on instructional practices predicted gains in reading proficiency controlling for prior achievement, although the magnitude of those gains were moderate; and (b) variability in self-reported affective and cognitive engagement neither predicted differences in student outcomes directly nor mediated the effects of instructional context on reading outcomes for struggling readers. Although more research is necessary, these findings have two important instructional implications. First, they provide support for the importance of not only what is taught (i.e., range and relevance of content), but also how it is taught (i.e., approach and disposition) and the depth to which it is taught (i.e., intellectual challenge). Second, these results highlight that these important skills can be measured, and thus used as formative feedback to improve teachers' instructional competence.

Contributions to Research

Despite the lack of significant findings in the full structural regression model, the current study made several contributions to the existing research regarding instructional context and student engagement. First, this study provided empirical support for the validity of MSIP-COT and SEI as multi-factor measures of the instructional environment and student affective and cognitive engagement, respectively. Although all three factors on the MSIP-COT were highly correlated, an examination of model fit and modification indices (MI) showed that the measure provided statistically significant and theoretically meaningful information regarding three distinct factors of the instructional environment. The hypothesized three-factor MSIP-COT model had relatively few significant model modification indices. Given evidence that such modifications rarely generalize well (MacCallum, Roznowski, & Necowitz, 1992), no MI-based modifications were made, although future research should consider the extent to which modifications from the three-factor model are supported by theory.

Further, model estimated factor scores show that the MSIP-COT did in fact capture variation between instructional groups. In addition, the variability in the current sample was likely underestimated relative to a random sample of schools, given that the districts in the current study had previously demonstrated a commitment to improving middle school literacy instruction. It is possible that in a more representative sample of districts, the MSIP-COT would show even more variability between instructional groups.

Similarly, results of the SEI measurement model indicated that the SEI does in fact measure five distinct factors of engagement, as hypothesized. Although several of the model fit indices indicated some degree of model misspecification based on conventional decision rules (i.e., Hu & Bentler, 1999), at least one group of researchers have argued that, for multifactor rating instruments when analyses are done at the item level (as the SEI analyses were), conventional criteria may be too restrictive (Marsh, Hau, & Wen, 2004). Correspondingly, a close examination of modification indices suggests that the relative lack of fit was due, in part, to significant cross-loadings of several items, a condition consistent with the reservations expressed by Marsh and colleagues. As with the MSIP-COT, results from estimation of SEI factor scores demonstrated that the measure does capture variability between students. It is also likely that the variability between students in the current sample, all of whom were identified as struggling readers, was less than it would be in a sample that included a representative number of typically performing readers.

A second distinguishing feature of the current study was its focus on struggling readers in middle school, students for whom improved instruction and engagement could reasonably be assumed to make a meaningful difference in outcomes, but who in this context are understudied, in part due to the additional complexities of class scheduling and compliance issues associated with middle school students (Fuchs, Fuchs, & Compton, 2010). This complexity was apparent in preparation of the data for the cross-classified model; some students received reading instruction as part of as many as six different instructional groupings over the course of 7^{th} grade, necessitating the use of weighted average factor scores in the cross-classified model. Because most of the research in this area has been conducted in elementary classes, it was important to begin to examine the extent to which the relationships observed in elementary settings generalize to middle school, particularly given research suggesting that as students age, engagement with school tends to decrease (Dotterer, McHale, & Crouter, 2009; Furrer & Skinner, 2003).

Third, the current study built on prior research by including previously understudied components of student engagement (i.e., measures of family and peer support for learning), using a measure of engagement that differentiated between the context (or sources) of affective engagement and the targets of cognitive engagement in an attempt to disentangle the various facets of psychological engagement. As reported, the relationships between these factors and student achievement were not statistically significant. However, the generalizability of these results may be limited, given that engagement was measured using a single instrument, and reported by a single individual on just one occasion.

Reading Intervention Ratings as Predictors of Achievement

Although the full cross-classified model (i.e., concurrent estimates for English language arts and reading intervention classes) did not support the notion that the MSIP-COT and SEI were significant indicators of reading proficiency for struggling readers, an analysis modeling the relationship between reading achievement for struggling readers and the instructional context ratings measured only in those classes identified as reading interventions did show some evidence that ratings of the instructional environment were predictive of reading achievement. Most notably, the instructional practices factor was a significant positive predictor of reading proficiency, controlling for prior achievement, results that are consistent with findings from other recent research (Kane, McCaffrey, Miller, & Staiger, 2013; Kane & Staiger, 2012).

Counterintuitively, higher ratings on a measure of instructional adjustments predicted lower reading proficiency, again controlling for prior achievement, although it is unclear exactly why that was the case. One possible explanation for the negative coefficient for the instructional adjustments factor relates to the measure itself. Many of the items on the MSIP-COT appear to be measuring more than one aspect of instruction. That is, most items are operationalized using multiple related, but arguably distinct behaviors or skills. For instance, the item labeled instructional adjustments was operationalized using three phrases: (a) “Teachers allow an adequate amount of think time before asking for student responses”; (b) “Student responses are used to adjust instruction for individual learners”; and (c) “Students receive sufficient guided practice before being allowed to independently practice skills and are given ample time and material to demonstrate independent mastery”.

Although few would argue against any of the three as important components of effective instruction, many would probably take exception with the categorization of all three as descriptions of a single characteristic that could be described as instructional adjustments. Clearly, it is possible for some teachers to merit a high ranking on one or two of the descriptors, but not all, making an accurate rating on this item exceptionally difficult. Thus, it is possible that the structure of the MSIP-COT is obscuring variability that otherwise would explain differences in reading proficiency after controlling for prior achievement. This hypothesis does not, however, explain why the instructional practices factor was a significant positive predictor, despite a similar design (i.e., items operationalized using multiple related, but distinct phrases).

From a practical perspective, another possibility is that the coefficient for instructional adjustments is negative because lower performing students tend to be those most in need of adjustments. Thus, it may be that teachers simply make more instructional adjustments when teaching struggling students, not that teachers who make adjustments have a negative impact on students. From a measurement perspective, the negative coefficient may also be evidence of a suppression effect. As shown in Figure 4, instructional practices and use of instructional adjustments are highly correlated, r(398) = .895, $p < .001$. It could be that, after accounting for the common variance, the remaining variance in the factor labeled use of instructional adjustments is actually measuring something else.

In the full model, the combination of all three factors (environment of interactions, instructional practices, and instructional adjustments) explained approximately 14% of the variability in reading proficiency controlling for prior achievement ($R^2 = .136$), Together, these findings highlight the challenges associated with measuring complex constructs such as instruction and engagement. Still, these results are consistent with findings from other recent research (Kane et al., 2013; Kane & Staiger, 2012). For example, the Measures of Effective Teaching (MET) Project, a large-scale longitudinal study of teaching effectiveness, found that classroom observation scores alone explained approximately 7% of the variance on state ELA assessments, half that explained in the current study. This suggests that, (a) although modest in an absolute sense, the variance explained by the current model is notable, at least for struggling readers, and (b) simply providing struggling readers in middle school with high quality instruction is likely not sufficient to significantly help them, relative to their peers.

Limitations

The validity of the current findings depends on many factors, including the study design, model specification, and the extent to which the measures were collected reliably and interpreted validly. Recognizing these limitations, this section identifies several important threats to validity and discusses the resulting implications. One significant limitation regarding the study design was the lack of random assignment. Given the correlational nature of the study, it is impossible to infer causality between the factors measured and student outcomes. Similarly, this study did little to account for a number of threats to internal validity, such as maturation or development. For instance, observations were only conducted in those classes that were identified by the school as language arts or reading intervention classes. Thus, this study did not account for the effects of any instruction occurring outside those contexts.

A related limitation was the timing and precedence of the outcome measures used in the study relative to the assessments of the instructional environment and student engagement, an issue that was particularly pronounced with respect to the testing schedule for the OAKS-R. Although designed as a summative assessment, students were allowed to take the OAKS-R up to three times during the year, and in some schools, students were able to begin taking the test as early as October. Further, test dates were not available as part of the data collected for the current study, so information regarding when each student took the OAKS-R could not be included in the models. Although most of the students in the current at-risk sample likely received their highest score relatively late in the year, it is likely that, at least for some students, the reading proficiency score used as an outcome was based on an assessment they had taken prior to receiving a substantial portion of their seventh grade reading instruction. In addition, the OAKS-R is a distal measure, relative to the observation of a classroom on a given day. Had the study used a more proximal measure of reading proficiency, the effects of instructional practices might have been more pronounced.

A further limitation related to construct validity is the single method, single instrument approach used to measure the instructional environment (i.e., direct observation using the MSIP-COT) and student engagement (i.e., student self-report using the SEI), rather than a more robust approach that made use of multiple measures of each. Based on prior research (Dotterer & Lowe, 2011), it is likely that the use of multiple assessment methods would have improved the accuracy and reliability with which the constructs were measured. That is, perhaps one reason why Dotterer and Lowe (2011) found a relationship between classroom context and engagement for typically performing students was because they utilized multiple measures of instruction and engagement, including teacher ratings, student surveys, and direct observational measures of instructional practice and behavioral engagement. For instance, a measure of the instructional environment that included ratings from other experts, such as the principal or other teachers, or even the students themselves (Kane et al., 2013; Kane & Staiger, 2012), might allow for more valid estimates of instruction. Similarly, a measure of student engagement that also included ratings from teachers likely would have provided better estimates of affective and cognitive engagement than self-report alone, and low inference measures of behavioral engagement, including direct observation, likely would have improved overall estimates of engagement.

It is also likely that the relationships between instruction and engagement are more complicated than could be represented in the current study, given available data, further suggesting the possibility of issues related to model specification. For instance, these constructs may represent feedback or cross-lagged relationships that interact with each other over the course of a school year that could not be represented accurately with a single measure of engagement and instructional ratings averaged across two to three class sessions. The results of any given study are ultimately dependent on the particular model specifications tested, which are a reflection not only of the variables included in the model, but those excluded as well. In the current study, there were a number of variables omitted entirely (e.g., attitudes, motivation, climate) that could have altered the reported parameter estimates, and potentially affected the conclusions. For instance, evidence regarding the malleability of student engagement is limited (Fredricks et al., 2004), but some studies have shown that teachers can affect levels of student engagement, at least with respect to observed behavioral engagement in young children (Downer et al., 2007; Hamre & Pianta, 2005). Having students complete one or more additional measures of engagement, attitudes, or motivation might have made possible (a) more accurate estimates of affective and cognitive engagement, (b) estimates of the stability of engagement, (c) testing of more complex relationships, or even all three.

Similar questions could be asked regarding the measurement of instructional context using the MSIP-COT. Although initial estimates suggest that the measure is relatively robust to differences between observers (reliability ICCs on the various components of the MSIP-COT range from .754 to .970; Kennedy & Nelson-Walker, 2013), stability estimates are substantially lower (stability ICCs range from .173 to .517;

Kennedy & Nelson-Walker, 2013), indicating that two or even three observations may not be sufficient to get a stable estimate of the instructional context. Further, the items themselves may not be sufficiently precise to generate accurate estimates of instructional context. As noted, a number of the items are operationalized using phrases that describe related, but arguably distinct behaviors or skills, potentially obscuring meaningful differences in instructional practices between teachers. In addition, there is evidence indicating that including more items per factor (perhaps by separating the descriptions of distinct skills) would make the MSIP-COT a more robust measure of instruction (Marsh, Hau, Balla, & Grayson, 1998).

A further potential limitation was the fact that the data were collected as part of an evaluation of existing district practices, rather than in the context of a focused evaluation of particular instructional features. As a result, students received a wide range of types and combinations of instruction. While every effort was made to accurately model the instruction each student received, complications related to representing the complex, simultaneous nesting of instruction inherent in middle school settings may have confounded the results. For instance, some levels of nesting in the data were ignored to allow tractable analyses, as described in the section on the MSIP-COT measurement model. Future research could clarify these relationships by studying the measures in a more controlled setting. Other potential misspecifications include the exclusion of one or more important variables and inaccurate assumptions regarding the form of the relationship between variables (e.g., inaccurately modeling a curvilinear relationship as a linear one).

Finally, the ecological and external validity of the current study are limited, given that it used a relatively homogenous sample, particularly with respect to the types of schools and districts studied. As described previously, districts were selected to participate as a result of a shared commitment to improving existing reading intervention practices and implementing systems designed to enhance student engagement in instructional settings. Further, a number of individual schools self-selected out of the evaluation. The schools that opted out tended to be either higher performing (i.e., they served relatively few lower performing students) or options schools that catered to students who struggled in the traditional middle school context. Thus, although the schools that did participate appeared to represent a fairly typical middle school environment, it is unknown how these findings would generalize to other contexts or settings, including those that differ with respect to locale, teacher or student population, or reading programs used. This is especially true of those contexts that lack a shared culture or are inexperienced in implementing intensive, system-wide interventions.

Directions for Future Research

Future research in this area could expand on the current study in several ways. From an analytic perspective, there are a number of potential improvements in model specification that could be made. For instance, given the magnitude of the factor correlations in the current study (.773 to .895), it may be that a model that includes a second-order instructional factor more accurately represents the MSIP-COT. To the extent defensible by theory, the confirmatory factor analysis (CFA) models used to represent the MSIP-COT and SEI could also be modified to allow for cross-loadings of items on multiple factors, as indicated by the model modification indices reported previously. Alternatively, the patterns between the modification indices may also be an indication of the need to specify a more complex model. For instance, the consistent cross-loadings between family support, control and relevance, and future aspirations may indicate predictive relationships between those factors. These measures may also be useful in aggregate as predictors of classroom level engagement or school climate. That is, average SEI scores may be a more reliable estimate of classroom engagement than individual scores are of student level engagement. Similarly, average COT scores may provide a more reliable estimate of school climate than individual ratings of teachers. Regardless, it is important to recognize that this study also highlights the need for additional instrument development and refinement, as described in more detail below.

From a study design perspective, a number of changes would likely help provide a clearer picture of the relationship between instructional context and student engagement. For instance, future research would benefit from a multiple measure, multiple method approach to measuring both the instructional environment and student engagement. One possibility would be to aggregate individual engagement ratings at the classroom level, to allow for a comparison of average student reported engagement to observer ratings of student participation and engagement. Future research could also investigate the effect of more proximal measures of engagement in context of the framework evaluated in the current study, and explore ways to measure and affect additional malleable factors, such as struggling readers' sense of self-belief and competence. Further, this study evaluated the effectiveness of the MSIP-COT and SEI only within the context of reading outcomes, and only for a single group of students: those identified as needing reading intervention in seventh grade. Recent research has found that the effects of observed teacher behaviors are smaller in language arts classes than they are in math classes (Kane & Staiger, 2012). Thus, it would be informative to examine the extent to which these findings generalize to other subject areas and student populations.

Finally, from a measurement perspective, the analyses presented in the current study provided clear evidence for the need to further refine both the MSIP-COT and the SEI. For instance, the utility of both the diagnostic and evaluative information provided by the MSIP-COT would almost certainly benefit from efforts to increase the precision of the measure. Based on the analyses presented here, those efforts would likely include the development of a revised version that includes a greater number of indicators per factor, but more precisely operationalizes each item. A relatively straightforward and theoretically defensible way of accomplishing this would be to conduct a detailed review of the operational definition of each existing item, and separate items that are described using multiple related, but arguably distinct phrases into multiple items.

As an example, the individual item labeled instructional adjustments is operationalized in the current MSIP-COT using three different phrases: (a) “Teachers allow an adequate amount of think time before asking for student responses”, (b) “Student responses are used to adjust instruction for individual learners”, and (c) “Students receive sufficient guided practice before being allowed to independently practice skills and are given ample time and material to demonstrate independent mastery”. One could readily conceive of a situation in which a rater clearly observes the first operational anchor (adequate think time) but not either of the other two. Using the current version of the MSIP-COT, this hypothetical observer must decide how much to weight each of the three anchors when scoring the instructional adjustments item.

Variability in this decision creates a potential source of measurement error, as different observers are likely to weight the anchors differently. The same observer may also differ in how they make this decision over time. Thus, the development and implementation of a version of the measure that divides these anchors into separate items would increase the likelihood that these different facets of instruction are measured accurately. These revisions would then need to be tested, first for usability, and ultimately, for improvements in estimates of instructional effectiveness. Such revisions may also facilitate the extension of these analyses to an examination of longitudinal growth, with the goal of (a) using estimates of effective classroom practices to predict change in student engagement, (b) exploring the extent to which direct observational measures such as the MSIP-COT are sensitive to changes in classroom practices over time, or both.

The current study also demonstrated that the SEI would likely benefit from additional development and refinement. As with the MSIP-COT, one useful revision would be to increase the precision of the language used in a number of items. The authors of the SEI argue that a defining characteristic of their measure is that it is sensitive to the experiences of individual students. However, several items ask students to rate their perception of school generally, rather than their own specific experience. For example, one item in the teacher-student relationships factor asks students to rate the extent to which "Adults at my school listen to the students." Changing that item to "Adults at my school listen to me" would shift the focus back to the student completing the measure.

A complimentary refinement would involve examining those items that load highly on multiple factors, and rewriting them to more precisely target a single factor. This issue is most clearly illustrated by item 27 ("I feel safe at school"). Included in the five-factor model as an indicator of teacher-student relationships, it also loaded highly on the peer support factor. Given the wording of the item, this specific cross loading is not particularly surprising; students' feelings of safety at school are likely to be affected by their relationships with both teachers and peers. This lack of specificity could be addressed by replacing the current item with two parallel items, one asking specifically about teachers (e.g., "I feel safe around the teachers at my school") and the other asking specifically about peers (e.g., "I feel safe around other students at my school"). As with the MSIP-COT, these revisions would then need to be tested for both usability (i.e., the extent to which students complete the revised measure as intended) and improvements in model fit. Such revisions may also benefit efforts to model the extent to which the SEI is sensitive to changes in student engagement over time, which has implications regarding a preventative model of engagement, in which students are monitored for risk of disengagement, and intervened with as necessary.

APPENDIX A

MSIP-COT RUBRIC FOR QUALITY INDICATOR ITEMS

1. Classroom Climate (Rapport, Respect, Support, Sensitivity, Positive Attitudes)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Students demonstrate disrespectfulness toward the teacher and fellow classmates; Interactions are often inappropriate and/or disrespectful; Students and teacher demonstrate negative attitudes; Teacher ignores student needs/ problems; When addressed, supports are ineffective, lack warmth, and/or exceed time necessary	Students demonstrate limited rapport with teacher and fellow classmates; Interactions are sometimes disrespectful; Teacher or some students demonstrate positive attitudes, but not both; Teacher addresses student needs with limited respect and care; Supports and redirects are effective or efficient, but not both	Students demonstrate rapport with teacher and each other; Teacher-student and student-student interactions are generally respectful; Teacher and some students demonstrate positive attitudes; Teacher addresses student needs with general respect and care; Supports and redirects are somewhat effective and efficient	Students demonstrate strong rapport with teacher and each other; Teacher-student and student-student interactions are highly respectful; Teacher and majority of students demonstrate positive attitudes; Teacher addresses student needs with respect and care; Supports and redirects are highly effective and efficient

2. Organization (Preparation, Transitions, Routines, Accessibility, Task Completion, Scheduling, Use of Instructional Time)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Lesson appears disorganized and unprepared; Transitions between and within lesson activities lack efficiency and consume large amounts of instructional time; Materials are not immediately accessible to both teacher and students; Routines are not evident in the classroom, allowing regular demands to take away from instructional time; Instructional tasks are never completed in the time allowed; Reading instructional time is not regularly scheduled or protected from interference	Lesson appears partly prepared; Transitions between and within lesson activities are mostly inefficient, resulting in loss of instructional time; Materials are immediately accessible to teacher or students, but not both; Routines are present for some regular demands but not others; Instructional tasks are rarely completed in the time allowed; Reading instructional time may be regularly scheduled, but is not prioritized over other activities	Lesson appears organized and prepared; Transitions between and within lesson activities are efficient and only rarely detract from time for instruction; Learning materials are immediately accessible to the teacher and students; Routines are present for many regular demands, allowing the focus to remain on reading instruction; Instructional tasks are usually completed in the time allowed; Instructional time is regularly scheduled, but other activities may occasionally interfere with content	Lesson appears highly organized and prepared; Transitions between and within lesson activities are efficient and do not detract from instructional time; Learning materials are easily accessible to the teacher and students; Routines are present and contribute to the amount of dedicated instructional time; Lesson activities are almost always completed in the time allowed; Instructional time is scheduled at a regular time and protected from interference as much as possible

3. **Relevance of Learning** (Purpose, Prior Knowledge, Personal and Cultural Experiences)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Lesson is not linked to identifiable learning objective(s) relevant to the development of ELA skills; Lesson is not made relevant to students' own personal and cultural experiences; Teacher does not connect new material to students' prior knowledge; Teacher does not link rationale for learning to students' future goals	Lesson is linked to the broad development of ELA skills but not a specific learning goal; Connections to students' prior knowledge, personal and cultural experiences, and future goals are superficial or unclear and do not maximize student learning	Lesson is linked to clearly defined learning objective(s); however, objectives may not be linked to students' own personal and cultural experiences or students' future goals; Connections to prior knowledge are not made sufficiently clear to promote deep understanding of new material	Lesson is linked to clearly defined learning objective(s) that are made relevant to students' own personal and cultural experiences; Teacher makes explicit connections between students' prior knowledge and new material; Teacher extends purpose of lesson by linking rationale for learning to students' future goals

4. **Classroom Management** (Clear Expectations, Minimizes Non-Instructional Time, Addresses Appropriate Behavior)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Expectations are not established; Few behavior supports used; Addresses all inappropriate behavior and disregards appropriate behavior; Few, students follow class rules and/or expectations; Management duties greatly reduce time devoted to instruction	Expectations are established, but not applied in the classroom; Few pre-corrections are used; Addresses more inappropriate behavior than appropriate behavior; Some students follow class rules and/or expectations; Management duties compromise time devoted to instruction	Expectations are established and applied most of the time; General use of pre-corrections to avoid / anticipate future events; Addresses and reinforces both appropriate behavior and inappropriate behavior; Majority of students follow class rules and/or expectations	Clear expectations are established and applied in the classroom; Use of pre-corrections to avoid / anticipate future events; Addresses and reinforces appropriate behavior and limits attention for inappropriate behavior. All students follow class rules and/or expectations

5. **Representation of Content** (Range of Examples, Range of Activities, Content Appropriateness)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Teacher provides insufficient instructional examples across content and materials; Students only rarely engage in activities designed to develop a variety of skills; Majority of students seem confused with materials and tasks; Interactions are of low-quality for most students and/or are inappropriate for the content	Teacher provides limited instructional examples across content and materials; Students occasionally engage in activities designed to develop a variety of skills; Student-teacher interactions are sometimes lacking for some students (e.g., struggling learners) and are inconsistently appropriate for the content	Teacher provides frequent instructional examples that are usually effective and varied; Students often engage in activities designed to develop reading, writing, oral communication, and literature skills; Student-teacher interactions are appropriate for content and allow most students to demonstrate understanding	Teacher provides frequent, effective, and varied instructional examples; Students are consistently engaged in a variety of activities designed to develop reading, writing, oral communication, and literature skills; Student-teacher interactions are appropriate for all students and content
6. **Clear and Consistent Delivery of Instruction** (Teacher Demonstrations, Pacing, Consistent Language, Minimizes Student Confusion)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Demonstrations are infrequent, inconsistent, and often confuse students; Teacher dialogue is either limited or excessive; Does not anticipate student misconceptions; Re-teaching does not occur; Little opportunity for students to contribute; Pacing is slow and some students seem off-task or uninterested in the lesson	Demonstrations are clear or concise, but not both; Delivery is sometimes inconsistent and duration of teacher dialogue is often excessive or few explanations are provided: Re-teaching occurs infrequently; Time for student contribution is sometimes limited, excessive, or unrelated; Pacing of instruction is inconsistent and a majority of students experience difficulties; Instruction moves on despite student confusion	Demonstrations are consistent and clear, but sometimes long; Duration of teacher dialogue is mostly appropriate; Explanations are provided, but sometimes too short; Re-teaches when necessary; Allows time for student contribution, but can sometimes lack purpose; Pacing of instruction is appropriate with few students experiencing difficulties	Demonstrations are clear, concise, and consistent; Duration of teacher dialogue is appropriate; Explanations are in-depth, but not excessive; Anticipates misconceptions by pre-correcting and/or providing further examples/ explanations; Re-teaches when necessary Allows time for an appropriate amount of student contribution; Pacing of instruction is at a level that promotes understanding and engagement

7. **Student Participation and Engagement** (Active Involvement, Compliance, Completion of Work)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
A majority of students appear off-task throughout entire lesson; A majority of students do not participate during whole-class activities, possibly due to lack of interest or motivation; Individuals do not complete independent work; Few students follow teacher directions	Instruction fosters student interest for parts of the lesson; Some students appear engaged; Some students appear eager to participate and learn; When assigned, some students complete learning tasks and/or independent work; Some students follow teacher directions	Instruction fosters student interest throughout most of lesson; A majority of students appear engaged throughout most of lesson; A majority of students appear eager to participate and learn; When assigned, most students complete learning tasks and/or independent work; A majority of students follow teacher directions	Instruction fosters student interest throughout entire lesson (e.g., during unitary response opportunities, most students respond); Most students appear highly engaged throughout entire lesson, and appear eager to participate and learn; When assigned, most students complete learning tasks and/or independent work; Most students follow directions
8. **Checks of Student Understanding** (Timely Checks, Active Monitoring, Individual Response Opportunities)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Checks are not used before, during, and after activities to gauge student understanding of content and procedures; Student responses are not actively monitored and confirmed or corrected; Instruction focuses on teacher talk or modeling with few student opportunities to respond; Does not actively check in with students throughout independent practice activities	Inconsistently checks for student understanding; Teacher actively monitors during a few activities, providing some corrective or confirmatory feedback to students; Opportunities to respond are focused on a small number of students, so that the majority of the class doesn't benefit; Actively checks in with some students throughout practice activities	Teacher monitors students throughout a range of activities, providing corrective and confirmatory feedback to students when appropriate; Response opportunities may be focused on the whole class only or insufficient in number for all students to demonstrate understanding before moving on to the next activity; Actively checks in with many students throughout activities	Uses frequent response opportunities and active monitoring to check for student understanding in all activities. Provides in-depth feedback, firming concepts and skills; Structures response opportunities so that the majority of the class can demonstrate understanding before moving on; Actively checks in with majority of students throughout activities

9. **Instructional Adjustments** (Student Response Time, Accommodates Learning Needs, Allows Independent Learning)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Allows insufficient wait time before asking for student responses; Adjustments are not made for individual learners; Does not provide sufficient guided practice for students to demonstrate independent mastery or does not allow sufficient time or material for students to practice skills independently	Teachers sometimes provide wait time before asking for student responses; Adjustments are occasionally made for individual learners; Occasional opportunities for independent learning are provided, but do not allow students to demonstrate independent mastery of content	Teachers usually allow an adequate amount of think time before asking for student responses; Instruction may be inconsistently adjusted to address all individual student needs; Opportunities for independent learning are provided, and generally allow students to demonstrate independent mastery	Teachers allow an adequate amount of think time before asking for student responses; Student responses are used to adjust instruction for individual learners; Students receive sufficient guided practice before being allowed to independently practice skills and are given ample time and material to demonstrate independent mastery

10. **Intellectual Challenge** (Depth of Understanding, Analysis of Concepts, Inferential Thinking, Reading Discourse)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Teacher demonstrates low expectations for many students; Teacher ineffectively manages reading discourse and does not allow students to engage in dialogue about concepts and ideas or does not manage student discourse; Teacher does not extend student thinking or probe depth of understanding through responses	Teacher demonstrates high expectations for some students; Little management of reading discourse occurs such that discussion often ventures off topic; Students occasionally have the opportunity to demonstrate depth of understanding by expanding on answers, making inferences, and analyzing concepts	Teacher demonstrates high expectations for most students; Student understanding is often promoted through teacher request for clarification; Teacher generally manages reading discourse such that students are often encouraged to share inferences, analyze concepts, and extend thinking around topics	Teacher demonstrates high expectations for all students; Teacher extends student understanding by asking for clarification and elaboration in questioning; Ample time is devoted to students' high level analysis of concepts and inferential thinking; Teacher manages reading discourse effectively

11. **Teaching for Reading Proficiency** (Positive Outlook on Reading, Views Reading as Important, Confidence, Productive Disposition)			
Not Present	***Somewhat Present***	***Present***	***Highly Present***
Teacher appears uninterested in teaching reading; Appears very uncomfortable in the classroom and teaching reading; Structures lessons in a way that limits student-teacher and student-student interactions; Does not allow for idea development and intellectual challenge	Teacher appears somewhat interested in teaching reading; Appears somewhat uncomfortable in the classroom and teaching reading; Does not plan lesson activities to promote original idea development and intellectual challenge, but may allow some discourse or exchange of thought if initiated by students	Teacher appears interested in reading and teaching students; Emits a general sense of efficacy; Appears somewhat confident in the classroom and teaching reading; Supports student confidence in reading by scheduling some time for intellectual challenge and original idea development	Teacher appears enthusiastic about reading and teaching students; Emits a high sense of efficacy; Fosters a sense that knowing reading is important in our world; Appears confident in the classroom and at ease teaching reading; Teaches students to be confident readers by encouraging original idea development and intellectual challenge

Adapted by Nelson-Walker, N. J. (2010) from the General Observation Features rubric, Doabler, C., & Nelson-Walker, N. J. (2009).

APPENDIX B

MIDDLE SCHOOL INTERVENTION PROJECT

CLASSROOM OBSERVATION TOOL

Observation ID

MSIP CLASSROOM OBSERVATION TOOL

School: ____________ Instructor Name: ____________

School ID: Date: M M / D D / Y Y Observer Initials:

Number of Students: Number of adults providing instruction ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 Start Time: : Stop Time: :

Program Code	Class period and time			
	Curriculum Name		Curriculum Level	

Reliability	Item Agreements	Total Score	Percent Agreement Item	Total score
A - C. Fidelity Ratings	/ 7			
D. Quality Indicators	/ 11			

Teacher Totals		Student Totals		Response Type Totals		Teacher	Student	Response Type
(M) Mod		(G) Grp		(V) Vbl		(M)	(G)	(V)
(Q) Init		(I) Ind		(W) Writ		(Q)	(I)	(W)
(F) Fdbk		(N) Init		(R) Read		(F)	(N)	(R)
		(P) Peer					(P)	
Total		Total		Total		Total	Total	Total

For Data Coordinators Only

Checked in by: ____________

Choose One:
○ Primary (1)
○ Maintenance (2) ○ Checkout (3) ○ Practice (4) ○ Cross-district (5)

Paired Observation ID

5288472747

Observation ID

MSIP CLASSROOM OBSERVATION FIDELITY RATINGS

A. MATERIALS / ENVIRONMENT	
Item	*Level of Implementation*
1. Teacher's instructional materials are organized and available to present	① ② ③ ④
2. Program instructional materials are utilized during the lesson	① ② ③ ④
3. Curricula/program lesson plan is clearly driving the instruction	① ② ③ ④
4. Student materials are in use	① ② ③ ④
5. List student materials that are in use:	
B. PROGRAM DELIVERY	
1. Teacher is familiar and fluent with the lesson	① ② ③ ④
C. STUDENT RESPONSE	
1. Students are familiar with group routines	① ② ③ ④
2. Students are engaged in classroom activities	① ② ③ ④
D. QUALITY INDICATORS	
1. Classroom Climate (Rapport, Respect, Support, Sensitivity, Positive Attitudes)	① ② ③ ④
2. Organization (Preparation, Transitions, Routines, Accessibility, Task Completion, Scheduling, Use of Instructional Time)	① ② ③ ④
3. Relevance of Learning (Purpose, Prior Knowledge, Personal and Cultural Experiences)	① ② ③ ④
4. Classroom Management (Clear Expectation, Minimizes Non-Instructional Time, Addresses Appropriate Behavior)	① ② ③ ④
5. Representation of Content (Range of Examples, Range of Activities, Content Appropriateness)	① ② ③ ④
6. Clear and Consistent Delivery of Instruction (Teacher Demonstrations, Pacing, Consistent Language, Minimizes Student Confusion)	① ② ③ ④
7. Student Participation and Engagement (Active Involvement, Compliance, Completion of Work)	① ② ③ ④
8. Checks of Student Understanding (Timely Checks, Active Monitoring, Individual Response Opportunities)	① ② ③ ④
9. Instructional Adjustments (Student Response Time, Accommodates Learning Needs, Allows Independent Learning)	① ② ③ ④
10. Intellectual Challenge (Depth of Understanding, Analysis of Concepts, Inferential Thinking, Reading Discourse)	① ② ③ ④
11. Teaching for Reading Proficiency (Positive Outlook on Reading, Views Reading as Important, Confidence, Productive Disposition)	① ② ③ ④

Comments:

2097472742

Teacher ______________________ Curriculum Name & Level ______________________________

Observation ID: ☐☐☐☐ Pg ☐ of ☐

	Start Time	Column #	Structure	Content	Other Detail
1	:		I S O T L	1 2 3 4 5 6 7	
2	:		I S O T L	1 2 3 4 5 6 7	
3	:		I S O T L	1 2 3 4 5 6 7	
4	:		I S O T L	1 2 3 4 5 6 7	
5	:		I S O T L	1 2 3 4 5 6 7	
6	:		I S O T L	1 2 3 4 5 6 7	
7	:		I S O T L	1 2 3 4 5 6 7	
8	:		I S O T L	1 2 3 4 5 6 7	
9	:		I S O T L	1 2 3 4 5 6 7	
10	:		I S O T L	1 2 3 4 5 6 7	
Stop Time	:				

Structure Codes

- I Independent Work
- S Small Group, Ind (2-7)
- O 1:1 Instruction
- T Small Group, Tch Led (2-7)
- L Large Group (8+)

Content Codes

- 1 Reading words
- 2 Reading connected text
- 3 Reading comprehension
- 4 Writing
- 5 Vocabulary
- 6 Other literacy (add detail)
- 7 Other (add detail)

Teacher

- M Model
- Q Initiation
- F Feedback

Student

- G Group
- I Individual
- N Initiation
- P Peer

Response

- V Verbal
- W Written
- R Reading

		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	28	29	30	31	32	33	34	35
Teacher		M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
		Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
		F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Student		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
		I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
		N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
		P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Response Type		V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
		W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
		R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R

		36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Teacher		M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
		Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
		F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Student		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
		I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
		N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
		P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Response Type		V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V
		W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
		R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R

0170352514

MSIP CLASSROOM OBSERVATION TOOL: STICO 2011-2012

APPENDIX C

STUDENT ENGAGEMENT INSTRUMENT

Student Engagement Instrument

Student MSIP ID Number

Marking Instructions
Use a No. 2 pencil only. **Do not use ink, ball point, or felt tip pens** **Make solid marks that fill the response completely.** **Erase cleanly any marks you wish to change.** **Make no stray marks on this form.**
CORRECT: ● INCORRECT:

	Strongly Agree	Agree	Disagree	Strongly Disagree
1. My family or guardian(s) are there for me when I need them.	1	2	3	4
2. After finishing my schoolwork I check it over to see if it's correct.	1	2	3	4
3. My teachers are there for me when I need them.	1	2	3	4
4. Other students here like me the way I am.	1	2	3	4
5. Adults at my school listen to the students.	1	2	3	4
6. Other students at school care about me.	1	2	3	4
7. Students at my school are there for me when I need them.	1	2	3	4
8. My education will create many future opportunities for me.	1	2	3	4
9. Most of what is important to know you learn in school.	1	2	3	4
10. The school rules are fair.	1	2	3	4
11. Going to school after high school is important.	1	2	3	4
12. When something good happens at school, my family or guardian(s) want to know about it.	1	2	3	4
13. Most teachers at my school are interested in me as a person, not just as a student.	1	2	3	4
14. Students here respect what I have to say.	1	2	3	4
15. When I do schoolwork I check to see whether I understand what I'm doing.	1	2	3	4
16. Overall, my teachers are open and honest with me.	1	2	3	4

1 7 6 8 4 7

Student Engagement Instrument

	Strongly Agree	Agree	Disagree	Strongly Disagree
17. I plan to continue my education following high school.	1	2	3	4
18. I'll learn, but only if my teacher gives me a reward.	1	2	3	4
19. School is important for achieving my future goals.	1	2	3	4
20. When I have problems at school my family or guardian(s) are willing to help me.	1	2	3	4
21. Overall, adults at my school treat students fairly.	1	2	3	4
22. I enjoy talking to the teachers here.	1	2	3	4
23. I enjoy talking to the students here.	1	2	3	4
24. I have some friends at school.	1	2	3	4
25. When I do well in school it's because I work hard.	1	2	3	4
26. The tests in my classes do a good job of measuring what I'm able to do.	1	2	3	4
27. I feel safe at school.	1	2	3	4
28. I feel like I have a say about what happens to me at school.	1	2	3	4
29. My family or guardian(s) want me to keep trying when things are tough at school.	1	2	3	4
30. I am hopeful about my future.	1	2	3	4
31. At my school, teachers care about students.	1	2	3	4
32. I'll learn, but only if my family or guardian(s) give me a reward.	1	2	3	4
33. Learning is fun because I get better at something.	1	2	3	4
34. What I'm learning in my classes will be important in my future.	1	2	3	4
35. The grades in my classes do a good job of measuring what I'm able to do.	1	2	3	4

176847

APPENDIX D

RELIABILITY CRITERIA FOR THE MSIP-COT

Measure	Standard	Description
Fidelity (sections A-C): Item agreement	.90	Percent of items data collectors rated the same, +/- 1.
Fidelity (sections A-C): Total score agreement	.85	Sum point values for each item; Smaller sum divided by larger sum (across observers)
Fidelity (section D): Item agreement	.90	Percent of items data collectors rated the same, +/- 1.
Fidelity (section D): Total score agreement	.85	Sum point values for each item; Smaller sum divided by larger sum (across observers)
Teacher Behavior	.80	Count total number of teacher behaviors; Divide smaller frequency by larger frequency (across observers)
Student Behavior	.85	Count total number of student behaviors; Divide smaller frequency by larger frequency (across observers)
Student Response Type	.85	Count total number of student response codes; Divide smaller frequency by larger frequency (across observers)

APPENDIX E

STUDENT ENGAGEMENT INSTRUMENT ADMINISTRATION DIRECTIONS

Procedures:

- Distribute questionnaires by student name.
- Read questionnaire items aloud with 3- to 5-second pauses between items depending on the reading levels within the class.
- Items should be read with brief pauses between the general text and parenthetical sections to aid in understanding, e.g., "extracurricular (after school) activities."
- Plural versions should be used for items with a plural option, e.g., "parent/guardian(s)."
- Choices (i.e., "strongly agree" to "strongly disagree") are described during the introduction. Following the introduction, the questions can be read without the choices.
- Cover sheets are only required for distribution. Once the form has been given to the student, the cover sheet can be discarded.
- Students in the class who are not provided a survey with a cover sheet/MSIP ID can be given an extra form to fill out with the rest of the class.

Note: If students ask, they may work ahead if the administrator's pace of reading is too slow for them.

Collection:

Collect completed questionnaires and:

- Arrange them so they all face the same way
- Make sure that half sheets and paper clips are removed
- Place them back in the envelope and give them to MSIP contact person

Surveys that were not completed will be placed in a separate envelope and given to the MSIP contact.

What to Say to Students:

1. "Today we have a questionnaire to learn about your experiences while attending this school. Individual answers to these questions will not be shared with your teachers or school administrators, but scores for each section may be used by your school to make instructional decisions. Your name and ID will NOT be linked to any of your answers or scores in any reports outside of the school district. Only averages for groups of students will be reported outside of the school district. Your honest answers will be used to help the school serve you and other students better.
2. "Do not begin marking answers until we discuss the directions and I begin to read the questionnaire items aloud.
3. "First, remove the half sheet with your name on it and set it aside.
4. "For each of the questionnaire items you will be choosing if you agree with the statement by selecting from 'strongly agree,' 'agree,' 'disagree,' or 'strongly disagree.

5. “For each item mark only one answer by filling in the circle completely with a pencil. If you make a mistake or change your mind, erase your old answer entirely and fill in your new answer.
6. “If you like, you can use a piece of paper or notebook to cover your answers.
7. “I’ll be reading the items and I can respond to any questions you might have right away.
8. “If you have any questions about the items I’m reading or if you need a bit more time with an item be sure to let me know.” *(Read items as direction in the Administration Procedures)*
9. “Thank you for your time and opinions.”

REFERENCES CITED

Alonzo, J., Park, B. J., & Tindal, G. (2008). *The development of middle school passage reading fluency measures for use in a progress monitoring assessment system* (No. 46). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *easyCBM online progress monitoring assessment system*. Eugene, OR: Center for Educational Assessment Accountability.

Anderson, D. A., Alonzo, J., & Tindal, G. (2010). *easyCBM reading criterion related validity evidence: Oregon state test 2009-2010* (No. 1103). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*(5), 369–386. doi:10.1002/pits

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. doi:10.1016/j.jsp.2006.04.002

Asparouhov, T., & Muthén, B. O. (2010). *Simple second order chi-square correction*.

Baker, S. K., Fien, H., & Crone, D. A. (2009). *Middle School Intervention Project (MSIP): Evaluating the implementation of a comprehensive middle school intervention*. *[Grant 84.305E proposal to the Institute of Education Sciences, Evaluation of State and Local Education Programs and Policies]*.

Betts, J. E., Appleton, J. J., Reschly, A. L., Christenson, S. L., & Huebner, E. S. (2010). A study of the factorial invariance of the Student Engagement Instrument (SEI): Results from middle and high school students. *School Psychology Quarterly*, *25*(2), 84–93. doi:10.1037/a0020259

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural models*. Newbury Park, CA: SAGE Publications.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Cardiff Software. (2009). Cardiff TeleForm. Solana Beach, CA.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Decker, D. M., Dona, D. P., & Christenson, S. L. (2007). Behaviorally at-risk African American students: The importance of student–teacher relationships for student outcomes. *Journal of School Psychology*, *45*(1), 83–109. doi:10.1016/j.jsp.2006.09.004

Doabler, C., & Nelson-Walker, N. J. (2009). *Ratings of Classroom Management and Instructional Support*. Center on Teaching and Learning, University of Oregon.

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. *Journal of Youth and Adolescence*, *40*(12), 1649–1660. doi:10.1007/s10964-011-9647-5

Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology*, *101*(2), 509–519. doi:10.1037/a0013987

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children’s risk for school problems contribute to children's behavioral engagement in learning? *School Psychology Review*, *36*(3), 413–432.

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128.

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Mac Iver, D., & Feldlaufer, H. (1993). Negative effects of traditional middle schools on students’ motivation. *The Elementary School Journal*, *93*(5), 553–574.

Fielding, A., & Goldstein, H. (2006). *Cross-classified and multiple membership structures in multilevel models: An introduction and review* (No. RR791). Birmingham, UK: University of Birmingham, Department for Education and Skills.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117. doi:10.2307/1170412

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. doi:10.3102/00346543074001059

Fredricks, J. A., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments. (Issues & Answers Report, REL 2011–No. 098)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2010). Rethinking response to intervention at middle and high school. *School Psychology Review*, *39*(1), 22–28.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148–162. doi:10.1037/0022-0663.95.1.148

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D. O., Raudenbush, S. W., & Whitehurst, G. J. (2010). *Evaluating teachers: The important role of value-added*. Brown Center on Education Policy at Brookings.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145. doi:10.3102/01623737022002129

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wychoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English Language Arts and teachers' value-added scores* (No. 16015). Cambridge, MA.

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. doi:10.1111/j.1467-8624.2005.00889.x

Hamre, B. K., Pianta, R. C., Downer, J. T., Decoster, J., Mashburn, Andrew, J., Jones, S. M., … Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, *113*(4), 461–487.

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, *61*, 280–289.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 1051–1078). Amsterdam: Elsevier B.V. doi:10.1016/S1574-0692(06)02018-6

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267–271. doi:10.1257/aer.100.2.267

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness of fit indices. *Sociological Methods & Research*, *11*, 325–344. doi:10.1177/0049124183011003003

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.

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology*, *100*(1), 1–14. doi:10.1037/0022-0663.100.1.1

IBM Corporation. (2011). IBM SPSS Statistics.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment*. Seattle, WA: Bill and Melinda Gates Foundation.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631. doi:10.1016/j.econedurev.2007.05.005

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*, 587–613.

Kennedy, P. C., & Nelson-Walker, N. J. (2013). [Inter-rater reliability and stability of the MSIP-COT]. Unpublished data.

Klem, A. M., & Connell, J. P. (2004). Linking teacher support to student engagement and achievement. *Journal of School Health*, *74*(7), 262–274.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.

Lewis, L., Parsad, B., Carey, N., Bartfai, N., Farris, E., & Smerdon, B. (1998). Teacher quality: A report on the preparation and qualifications of public school teachers. *Education Statistics Quarterly*, *1*(1), 1–5.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490.

Maehr, M. L., & Meyer, H. A. (1997). Understanding motivation and schooling: Where we've been, where we are, and where we need to go. *Educational Psychology Review*, *9*(4), 371–409.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *33*, 181–220.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341.

McDonald, R. P. (2010). Structural Models and the Art of Approximation. *Perspectives on Psychological Science*, *5*(6), 675–686. doi:10.1177/1745691610388766

Measures of Effective Teaching [MET] Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Bill and Melinda Gates Foundation.

Muthén, L. K., & Muthén, B. O. (2012). Mplus user's guide. Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2011). Common core of data: Public elementary/secondary school universe survey [Data file and code book]. Retrieved from http://nces.ed.gov/ccd/pubschuniv.asp.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, *102*(5), 367–387. doi:10.1086/499709

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal*, *105*(3), 305–323. doi:10.1086/428746

Nelson-Walker, N. J., Kennedy, P. C., Cohen, J., & Crone, D. A. (2011). MSIP-COT: Middle School Intervention Project Classroom Observation Tool [Classroom observation instrument]. Eugene, OR: Center on Teaching and Learning, University of Oregon.

Nelson-Walker, N. J., Turtura, J., & Cohen, J. (2012). *Middle School Intervention Project: Classroom Observation Tool (MSIP-COT) technical report*. Eugene, OR: Center on Teaching and Learning, University of Oregon.

Oregon Department of Education Office of Assessment and Information Services. (2007). *The Oregon statewide assessment system annual technical report (Vol. 4: Reliability and validity)* (Vol. 4). Salem, OR: Author.

Oregon Department of Education Office of Assessment and Information Services. (2011a). *The Oregon statewide assessment system reading/literature test specifications and blueprints 2011-2012: Grade 6*. Salem, OR.

Oregon Department of Education Office of Assessment and Information Services. (2011b). *The Oregon statewide assessment system reading/literature test specifications and blueprints 2011-2012: Grade 7*. Salem, OR: Author.

Oregon Department of Education Office of Assessment and Information Services. (2012). *The Oregon statewide assessment system annual technical report (Vol. 1: Annual report)*. Salem, OR: Author.

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology*, *99*(1), 83–98. doi:10.1037/0022-0663.99.1.83

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109–119. doi:10.3102/0013189X09332374

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

Pianta, R. C., la Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, *102*(3), 225–238.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. doi:10.1037/0022-0663.95.4.667

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

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

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175–215.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education*, *14*(4), 329–339.

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student–Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, *27*(2), 316–328. doi:10.1016/j.ecresq.2011.09.004

Strong, M. (2011). *The highly qualified teacher: What is teacher quality and how do you measure it?* New York: Teachers College Press.

Strong, M., Gargani, J., & Hacifazlioglu, O. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, *62*(4), 367–382. doi:10.1177/0022487110390221

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, *62*(4), 339–355. doi:10.1177/0022487111404241

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2008). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, *20*(3-4), 165–184. doi:10.1007/s11092-008-9053-z

Wang, M.-T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, *47*(3), 633–662. doi:10.3102/0002831209361209