ABSTRACT

Title of Document: MEASURING TEACHING PRACTICES: DOES A SELF-REPORT MEASURE OF INSTRUCTION PREDICT STUDENT ACHIEVEMENT?

Jill M. Berger, M.A., 2010

Directed By: Professor Gary D. Gottfredson, Department of Counseling and Personnel Services

Teachers affect student achievement. Measuring what makes teachers "effective" is a challenge without a clear definition of the construct or constructs involved. Self-reports cost little and allow for data collection from large samples, but the reliability and validity of self-report measures for studying teacher effectiveness have not been adequately examined. This study explored the utility of a self-report measure of instruction (Instructional Practices Scale). Hierarchical linear modeling was used to examine the effects of the scale on students' reading and math standardized test scores and report card grades. Although the scale showed small to moderate relationships with teacher characteristics, results suggested little predictive validity and little discriminant validity. Further, the effects of teacher-reported instruction on achievement were not dependent on students' entering level of achievement. When measuring loosely defined constructs such as "effective instruction," the cost of using a self-report measure may outweigh the benefits.

# MEASURING TEACHING PRACTICES: DOES A SELF-REPORT MEASURE OF INSTRUCTION PREDICT STUDENT ACHIEVEMENT?

By

Jill M. Berger

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2010

Advisory Committee:
Professor Gary D. Gottfredson, Chair
Professor Sylvia Rosenfield
Professor William Strein

Acknowledgements

Jill M. Berger, Counseling and Personnel Services, University of Maryland.

This research was supported in part by Institute of Education Sciences (IES) grant #R305F050051. Opinions expressed do not necessarily reflect those of the sponsor. This work would not have been possible without the contributions of the IC Teams research group at the University of Maryland (Gary Gottfredson, Sylvia Rosenfield, Kate Bruckman, Phuong Vu, Lauren Kaiser, Megan Vaganek, Deborah Nelson, Todd Gravois, and Eva Yiu). I would like to thank Gary Gottfredson and Sylvia Rosenfield in particular for their advice, edits, and support throughout the process. In addition, many thanks to Jennifer Cassata, Kathy Aux, and Prince William County schools for their help in collecting the data used for this study.

Correspondence concerning this article should be addressed to Jill Berger, Department of Counseling and Personnel Services, University of Maryland, College Park, MD 20742. E-mail: jmberger@umd.edu

# Table of Contents

List of Tables

# List of Figures

# Measuring Teaching Practices: Does A Self-Report Measure Of Instruction Predict Student Achievement?

Every school day, more than 3 million teachers instruct 49 million children in our public schools (National Center for Education Statistics, 2006). There is a great deal of variability among teachers; teachers differ not only in their certification, knowledge, and ability (Goldhaber & Brewer, 2000; Guarino, Hamilton, Lockwood, Rathbun, & Hausken, 2006) but also in how they manage (Brophy, 1983) and instruct the students in their classrooms (Brophy & Good, 1984). Unfortunately, students with academic difficulties are often placed with the least effective teachers in classrooms with other struggling students; thus they fall even further behind (Veldman & Brophy, 1974). Teacher effects may be largest in classrooms with many low-performing students compared to classrooms with many high-performing students, because students in high-achieving classrooms can learn from their peers as well as from the teacher (Veldman & Brophy, 1974). Teacher effects may also be more prominent in schools with low socioeconomic status (Nye, Konstantopoulos, & Hedges, 2004). Further, having an ineffective teacher for one year may retard academic gains for up to two years, even after later exposure to a highly effective teacher, according to at least one study (Sanders & Horn, 1998).

Much attention has been devoted to understanding the achievement gap, which refers to the lower achievement of poor and minority students compared to rest of the youth population, and finding ways to support groups falling behind with better educational practices (Haycock, 2001). Many states are working on ways to hold teachers accountable for student outcomes (Education Week, 2004; Haycock, 2001).

Defining and measuring teacher effectiveness has been a prominent topic in educational research for several decades. The consensus from the literature is that teachers play a critical role in the achievement outcomes of students in their classrooms. Across studies, there is evidence that teachers account for 15 to 20 percent of the variance in student achievement, with the remaining variance lying among schools (15-20%) and students (60-70%) (Scheerens & Bosker, 1997). Studies using different methods have found varying teacher effect sizes, but even by the most conservative estimates, teacher effects are significant predictors of student achievement (Peverly, 2009). Although we have evidence that student achievement varies as a function of teachers, in order to improve learning outcomes for all students and especially for disadvantaged students, we must understand the specific characteristics of classrooms and teaching practices that affect student achievement.

There are several challenges to studying teacher effects. First, we lack a concrete definition of quality teaching (LeTendre, 2009; Rosenshine, 1970). More research is needed to specify teaching practices that lead to the best outcomes for students of differing entry skill levels. Second, our methodology for measuring teacher effectiveness is crude (Mayer, 1999; Rosenshine, 1970). Researchers have not yet developed valid and reliable instruments for measuring features that make teachers and teaching effective. Third, many students, especially those at-risk, receive instruction not only from the classroom teacher but also from various support staff and other teachers within the school (Valli, Croninger, & Walters, 2007). This sharing of instructional responsibilities may benefit students, but it creates challenges for accountability and research on teacher effectiveness. Individual teachers are rarely the sole educators of at-risk students in their classroom, making it difficult to hold them accountable for test scores (Valli, Croninger, & Walters, 2007).

**Defining Teacher Effectiveness**

One gap in the literature on teacher effectiveness is the lack of reliable, valid tools for measuring teaching practices (Mayer, 1999). The primary barrier to creating useful measurement instruments is that we are unclear which, among the many classroom and teacher variables correlated with student success, are the most salient causes of academic achievement. Defining quality teaching in a more concrete and precise way should be a priority for education researchers and practitioners.

**Teacher qualifications.** Studies have examined teachers' qualifications, such as certification, educational background, knowledge, and ability, as possible characteristics that make teachers effective in the classroom (e.g., Goldhaber & Brewer, 2000; Guarino, Hamilton, Lockwood, Rathbun, & Hausken, 2006). Teacher education has received a great deal of attention in the last three decades. Federal initiatives have pushed for teaching certification via alternate routes in addition to the standard four-year undergraduate teacher preparation programs, in an effort to spend less time and resources on teaching training, recruit a greater number of teachers (Darling-Hammond, 2000), and potentially attract a larger variety of college students to the teaching profession. Stakeholders have debated the best methods for preparing today's teachers: traditional undergraduate programs to prepare teachers professionally or on-the-job training for teachers with solid content area knowledge (Darling-Hammond, 1990). In her speech on the history of teacher professionalism, Ravitch (2002) argued that teachers should be "not only well-trained but truly welleducated," suggesting that more teachers need degrees in academic fields rather than degrees in education only.

The effects of certification on teacher effectiveness have been difficult to isolate because many variables are in play: certification requirements differ among states; teachers with certification may have received their credentials from a traditional or an alternate program; and there is a serious issue of selection bias when comparing teachers (e.g., beginning teachers are more likely uncertified) (Whitehurst, 2000). Descriptive evidence has suggested that traditionally certified teachers report greater satisfaction with their preparation and are less likely to leave the profession than are teachers with alternate certification (Darling-Hammond, 2000). In a recent study, Constantine et al. (2009) randomly assigned elementary school students to teachers trained in traditional versus alternative certification programs. They found no significant between-group differences in students' reading and math achievement, and no significant differences between teachers trained traditionally or alternatively in terms of college entrance exam scores, undergraduate college selectivity, and educational attainment. Constantine et al. did not examine teacher persistence in teaching or job satisfaction.

Teachers' knowledge and ability have also been examined in an effort to understand teaching quality. What do teachers need to know in order to effectively teach the students in their classrooms? Shulman (1986) argued for the importance of subject matter knowledge in the mid 1980s when many states were focusing solely on the how well teachers could perform the actual process of teaching, such as organizing and presenting lesson plans, evaluating students, recognizing differences among students, understanding children, and managing classrooms. Shulman proposed three critical areas of teacher knowledge, to which Peverly (2009) added a fourth: subject matter knowledge, content pedagogical knowledge, general pedagogical knowledge, and Peverly’s general world knowledge. Subject matter knowledge explains how well the teachers know the content of the subject matter they are teaching to students. In his review, Peverly concluded that there is little experimental research on the effects of teacher subject matter knowledge on student outcomes, and there is some evidence that teachers’ knowledge of linguistics and mathematics is generally low. Content pedagogical knowledge refers to instructional practices used to deliver specific content to students, and general pedagogical knowledge describes general teaching practices used to facilitate learning in the classroom. Research on pedagogical knowledge suggests that teachers vary in their use of content and general instructional practices, and that these practices are related to student outcomes. Teachers’ general world knowledge might be described as a measure of teacher ability or intelligence and has been measured by test scores on certification and college admissions exams. Greenwald, Hedges, and Laine (1996) conducted a meta-analysis to study the effect of school resources on student achievement and found that teacher ability (i.e., Peverly’s general world knowledge), experience, and education had a consistently strong positive relationships with academic outcomes for students.

In sum, we can conceptualize and measure four domains of teacher knowledge, and correlational evidence suggests that teacher knowledge is related to student outcomes (Peverly, 2009). At the 2002 Whitehouse Conference on Preparing

Tomorrow's Teachers, Whitehurst, former director of the Institute of Education Sciences, drew two important conclusions from the research on teacher knowledge. First, subject matter knowledge may be an important predictor of student outcomes in high school specialized areas such as science and math but there is less evidence of its effects for students at the elementary level. Second, there is substantial evidence that teachers' verbal and cognitive ability, as measured by college admissions tests, is one of the most salient predictors of students' academic achievement.

**Teacher practices.** To be effective, teachers must not only have knowledge of content and effective instruction, but they must also be skilled in the application of these principles (Shulman, 1986). The school is a fast-paced environment, and teachers are faced with many competing demands in the classroom. In addition to delivering the content of lessons, teachers plan activities, prepare materials, grade student work, monitor the classroom, help students transition throughout the day, keep parents informed, work with other school staff, and engage in many other responsibilities. It takes a skillful teacher to be able to manage the day-to-day tasks while at the same time maximizing time for student learning and providing effective instruction to students with a wide range of achievement levels. When we examine student outcomes, it is apparent that the practices a teacher uses in the classroom may or may not create an environment that is conducive to student learning and socio-emotional development.

Process-product research is "the search for relations between classroom processes (teaching) and products (what students learn)" (Gage &Needels, 1989, pp. 254). Such research attempts to capture the essence of a quality classroom by looking at process variables that affect a certain product or outcome, which is generally student academic achievement. In the study of teacher effects, the process variables linked to student outcomes include teaching behaviors (e.g., instructional practices, classroom organization, behavior management) and classroom interactions (e.g., student-teacher relationships, peer interactions). The current study focuses on the instructional practices that teachers use in their classrooms. In a review of early experimental and correlational research on teaching practices, Rosenshine (1983) concluded that effective teachers:

> Structure the learning; proceed in small steps but at a brisk pace; give detailed and redundant instructions and explanations; provide many examples; ask a large number of questions and provide overt, active practice; provide feedback and corrections, particularly in the initial stages of learning new material; have a student success rate of 80% or higher in initial learning; divide seatwork assignments into smaller assignments; provide for continued student practice so that students have a success rate of 90-100% and become rapid, confident, and firm. (p. 337)

Rosenshine noted that these practices are particularly critical for younger and less skilled students. He discussed the notion of "overlearning" basic skills during the early instruction, a concept that is now generally referred to as mastery learning or learning to automaticity (Bloom, 1974). Mastery learning is achieved when teachers present new material in small chunks and then students practice these new skills repetitively until they become automatic.

The purpose of mastery learning is to build automaticity of basic skills so that limited cognitive capacity can be used for higher-level cognitive processing (Rosenshine, 1983). When students have not mastered basic skills that are prerequisites to more complex tasks, teaching complex tasks is ineffective (Bloom, 1974).

In their review of process-product research from the 1970s and early 1980s, Brophy and Good (1984) further identified aspects of teacher behavior that were correlated with student academic achievement. They examined teacher behaviors during basic skills instruction that were consistently, though not necessarily highly, correlated with student achievement. Of the behaviors studied by Brophy and Good (1984), variables associated with instructional quality and pacing were most highly correlated with student achievement. These variables included providing opportunities to learn; managing the classroom and allowing for student engaged time in activities; engaging in active teaching during which students were being instructed and supervised; and providing tasks that allowed students to experience consistent success. The last variable, consistent success, is a measure of maximizing academic learning time that involves creating a match between the task and the student's skill level. Creating this instructional match is particularly important for younger students to keep up classroom momentum and build foundational skills for more complex learning in later grades (Brophy & Good, 1984).

In order to provide effective instruction to students, the classroom teacher must be skillful in managing the many competing demands in the classroom. Effective teachers not only have content knowledge and knowledge of effective instruction, but they are also able to put their knowledge into action on a daily basis in the classroom. The advantage of studying teaching process variables is that it provides many possibilities for increasing the effectiveness of teachers already in the field. In contrast, the study of teacher characteristics, such as certification and knowledge, has implications for teacher selection and training but has had little additional application to improving student outcomes. If we can grasp the specific process variables that define quality teaching, we can provide professional development and constructive feedback to teachers to use with their current students to improve learning and achievement immediately.

**Measuring Teacher Effectiveness**

Researchers have used a variety of methods to measure process variables associated with student outcomes. Some promising approaches to measuring teaching quality include classroom observation, teacher logs, and self-report surveys. Qualitative research (e.g., interview studies) has also been conducted to examine teaching practices; unfortunately these studies have not been conducted within the framework of larger, quantitative studies and their small sample sizes limit generalizability (Rowan, Correnti, & Miller, 2002). Recently, qualitative research has been used to examine the validity of quantitative instruments (Desimone & Le Floch, 2004; Mullens & Kasprzyk, 1996). Focus groups, case studies, and interviews may be promising ways to help us shape instruments that can be used on a larger scale.

**Classroom observation.** Classroom observation can be an objective method to measure variables in the classroom by using external observers to rate the classroom and teacher on dimensions that are potentially associated with student outcomes. Rosenshine (1983) described how observation systems differ in the degree of inference the observer must make when using the system. Low-inference systems are preferred since the observer documents specific behaviors as they occur during the observation period thus making ratings more objective (Rosenshine, 1983).

The Vermont Classroom Observation Tool (VCOT), developed by the Vermont Institutes, measures teaching practices used during reading and math instruction (Constantine et al., 2009). The VCOT was adapted from an instrument designed to measure instruction in science and mathematics, the Science and Math Program Improvement (SAMPI) tool, developed by researchers at Western Michigan University. The VCOT underwent several revisions before it was adapted to measure literacy in addition to math instruction in the Constantine et al. (2009) study. The VCOT measures teaching practices in three domains: lesson implementation, lesson content, and classroom culture. The lesson implementation domain examines teacher's use of best practices, pacing, confidence, and student engagement. The domain of lesson content examines how well the teacher understands the lesson's concepts and content and links the lesson to real world applications and other subjects. The classroom culture domain measures the extent of routine in the classroom, appropriateness of behavior, and the teacher's sensitivity to student diversity. In the Constantine et al. study, observers scored indicators under each domain on a five-point scale during two math and two reading instruction periods over a course of four days. Two observers were selected based on their high level of agreement with an expert panel on videotaped classroom sessions during training. Reliabilities were high ($\alpha$=.88 to .98) for the sample, but they found no statistically significant positive correlations between the domain scores and achievement and an unexpected negative relationship between classroom culture in literacy instruction and reading scores. The Constantine et al. study is the only study that examined the relation between VCOT scores and student achievement.

Another observation instrument, the Classroom Assessment Scoring System (CLASS), measures classroom quality by examining relationships between students and teachers and among students in the classroom (Pianta, La Paro, & Hamre, 2008). The CLASS assesses “what teachers do with the materials they have and on the interactions they have with children” in preschool through third grade classrooms (La Paro, Pianta, & Stuhlman, 2004, p. 412). It measures classroom process variables that have been related to positive student outcomes in quasi-experimental studies. Dimensions measured include emotional climate, classroom management, and instructional support. The CLASS identifies specific behavioral markers to lower the inference that observers must make when coding these broad dimensions. Potential users of the CLASS must attend a two-day training and take a reliability test to complete training. In the technical manual, program developers reported interrater agreement on CLASS dimensions ranging from 78.8 to 96.9 percent. Program developers reported that the emotional support domain has been shown to predict preschool and first grade students’ literacy skills as assessed by standardized tests and that the instructional support domain has been associated with academic performance for preschoolers after adjusting for student background characteristics (Pianta, La Paro, & Hamre, 2008).

Ysseldyke and Christenson's (2002) Functional Assessment of Academic Behavior (FAAB), originally The Instructional Environment Scale (TIES), uses multiple methods for assessing a student's total learning environment, which is conceptualized as the student's instructional support, home support, and home-school support. The FAAB is a nine-step system designed to collect information about a teacher's concern, assess the student's instructional needs, assess the student's learning environment, and inform instructional interventions. After the teacher and parents have shared their perspectives on the student's instructional needs, data on the student's instructional environment are collected using at least two of four proposed methods: observation, recollection, record review, and testing. The instructional support domain, most relevant to this discussion, measures 12 dimensions in four areas: planning (instructional match, instructional expectations), management (classroom environment), delivery (instructional presentation, cognitive emphasis, motivational strategies, relevant practice, informed feedback), and evaluation (academic engaged time, adaptive instruction, progress evaluation, student understanding). Ysseldyke and Christenson recommended that observers use the Observation Record provided in the manual, which is a two-page form with four open-ended questions about the areas of instruction.

The FAAB in its entirety is a complicated system, and its observation method is less systematic and objective than the VCOT and the CLASS because observers take anecdotal notes on dimensions rather than making numerical ratings. The FAAB observations are not designed for use in research, and it would be difficult to translate the descriptive information into quantitative data for a research study. Quantitative observation systems, like the VCOT and the CLASS, are more applicable for research studies. Unfortunately, most of these existing observational methods are time consuming and costly—often too costly to use in large-scale studies. Another problem with using observation systems is that they only measure the specific behavioral markers that are observed. For example, the CLASS does not measure some of the instructional practices targeted by staff development interventions, such as teacher planning and materials provided to students. Finally, there is limited evidence suggesting that observation scores obtained from these systems are linked to student achievement. Correlational studies provide some indication that the CLASS scores are related to student outcomes (Pianta, La Paro, & Hamre, 2008). To date, there is no evidence that the VCOT scores are related to student outcomes (Constantine et al., 2009).

**Teacher logs.** Another measure of teaching practices that has been used to measure the outcomes of professional development programs is classroom logs. According to Mullens and Kasprzyk (1996), using daily or weekly logs to measure teaching practices has the advantage of obtaining information on a frequent and contemporaneous basis, compared to retrospective surveys which might ask teachers to make a summary report on their practices over long periods of time. Teachers might also be more willing to make accurate reports when they know that their practices are being reported and analyzed across multiple occasions (Mullens & Kasprzyk, 1996). When making a single report on their instruction, teachers might fail to take into account days when activities did not go as planned and instead report on their teaching under ideal circumstances.

Daily logs have been used to assess the outcomes of a professional development program on literacy instruction, the Study of Instructional Improvement (Correnti, 2007; Rowan, Camburn, & Correnti, 2004). Teachers were asked to complete logs for randomly identified target students documenting their frequency of topic coverage and the instructional strategies that they used in teaching comprehension, writing, and word analysis. There was a high response rate (89%) for the logs and high completion rate among those teachers who logged (90%). Teachers' log data were compared to data collected from external observers of the classroom. Researchers found that teacher logs could provide detailed information about instruction and frequency of instructional activities that corresponded to data collected from the observations. They concluded that using both log and observation data on instruction provided a more complete picture than relying on either data source alone (Camburn & Barnes, 2004).

**Self-report questionnaires.** Large-scale questionnaire survey research is a frequently used method for measuring many variables associated with teacher effectiveness. The popularity of questionnaires is not surprising given that they are easily administered to large samples, especially with web-based administration using survey development tools such as Survey Monkey and Zoomerang. In addition, self-report measures are not particularly cumbersome for participants and cost researchers relatively little (Mayer, 1999).

However, using questionnaires to measure teaching variables presents serious concerns since scholars are still in the process of refining the definition of teacher quality. Also, it is unclear what exactly teacher reports measure; survey items generally serve as proxy variables for certain constructs but they may not actually measure what was intended (Rowan, Correnti, & Miller, 2002). This is a considerable problem when measuring teaching process variables. For example, a self-report questionnaire of instruction asking teachers to report the frequency of certain behaviors used during instruction may actually measure teachers' perceptions of their own teaching or their ideal instead of actual behavior during lessons (Berger & Kaiser, 2007; Mayer, 1999). Although such measures are commonly used, relatively little is known about their reliability and validity as measures of teaching practices.

Mayer (1999) measured the reliability and validity of self-report questionnaires for assessing algebra teachers' instructional practices. Teachers were asked to fill out a 34-question self-report survey about the amount of time they spent using teaching practices recommended as standards by the National Council of Teachers of Mathematics (1989). With a sample of 19 algebra teachers, reliability was measured by administering the same instructional practices survey four months apart. The test-retest reliability of the composite score on the survey was .69, and there was little reliability between teachers' scores on individual items across the two survey administrations. To examine validity, classroom observations were used as a second measure of teachers' instructional practices for a stratified random sample of nine algebra teachers who had taken the self-report survey. Three observations of each teacher were spread out over several weeks; classroom observers were blind to the teachers' scores on the self-report survey. There was a .85 correlation between the composite score on the survey and a composite score based on the classroom observations. This study has a small sample size and is restricted to a specific standard of algebra instruction, but nonetheless demonstrates (a) the importance of validating self-report questionnaires of instruction before interpretations from the data can be made, and (b) that a high degree of convergence between self-reports and observations is possible.

In addition to issues with the reliability and validity of teacher self-reports, there is concern about how large-scale survey data are analyzed to assess teacher effects and how these findings are interpreted. Rowan et al. (2002) suggested that different analytical techniques for analyzing survey data have been a source of confusion over the magnitude of teacher effects. According to Rowan et al., variance decomposition models used in previous research have underestimated teacher effects by failing to take into account students' gains in achievement and instead simply measuring achievement status. The current study will use multilevel modeling to make covariate adjustments for potential confounds at the student and classroom level.

**Current Study**

The current study uses data collected from a teacher self-report questionnaire to estimate the effects of teaching practices on student achievement. This investigation fills a gap in our knowledge of teacher effectiveness in two ways. First, more research is needed to determine the reliability and validity of self-report measures of teaching practices. The current study is designed to assess the utility of a teacher self-report, the Instructional Practices Scale (IP Scale), for use in large-scale survey research.

Second, the IP Scale is currently being used as an outcome measure in experimental and quasi-experimental evaluations of Instructional Consultation Teams (IC Teams), a teacher support intervention (Rosenfield & Gravois, 1996). Kaiser

(2007) studied the relationship between IC Teams and teaching practices and found no significant differences on teachers' instructional practices between schools implementing IC Teams for 2 to 3 years, schools implementing IC Teams for 1 year, and schools without IC Teams. In an experimental evaluation of the effects of IC Teams on teacher outcomes after three years of implementation, Vu et al. (2010) found no significant effects of the intervention on teachers' reported instructional practices. There are two plausible explanations for the null findings that are relevant to the current study. First, it is possible that the IC Teams intervention has not affected teaching practices. A second explanation is that the self-reports do not provide a valid instrument for assessing teacher instructional practices that lead to student learning. The survey may fail to detect differences among teachers in behavior that are related to student achievement. The current study explores the utility of this survey instrument.

## The Instructional Practices Scale

**Background.** The Instructional Practices Scale (IP Scale) was created by the research team, consisting of Instructional Consultation Teams' (IC Teams) program developers and evaluators, as an outcome measure in a school-level experimental evaluation of the intervention. The IP Scale was designed to assess teaching practices that are the foci of IC Teams training. These teaching practices follow the effective principles of instruction reviewed in the literature (e.g., Bloom, 1974; Rosenshine, 1993; Brophy & Good, 1984; Shapiro, 2004), and are particularly important for use with students struggling in the classroom (see Appendix A for empirical support of items). These practices include maintaining appropriate instructional levels for students; assessing and activating students' prior knowledge when teaching new material; teaching amounts of new material that heeds the usual limits of students' working memory depending on age and skill level; providing repetition and corrective feedback to students; managing the classroom environment; and using individual behavior assessment and intervention techniques (Kaiser, 2007).

By following these practices, teachers individualize instructional strategies and interventions for students experiencing difficulties, as suggested by Shapiro (2004)[1]. Methods that improve the student's performance are continued; methods that are ineffective are discontinued, "but only for that individual [student]" (Shapiro, 2004, p. 165). In this model, teachers continuously assess the functional relationship between what they are doing (e.g., their instruction, strategies, interventions) and the student's performance and made modifications accordingly.

**Development.** The Instructional Practices Scale (IP Scale) was designed to measure the frequency with which teachers use these presumably effective instructional practices with students experiencing difficulties in their classroom. Initially, developers created five, four-item subscales tapping different components of instruction: planning, delivery, management, academic assessment, and behavioral assessment, similar to areas measured by Ysseldyke and Christenson's (1993) The Instructional Environment Scale-II (Kaiser & Rosenfield, 2006). The first 16 items of

---

[1] Shapiro (2004) described this approach as an examination of the functional relationship between individual students and their performance, as opposed to traditional aptitude-treatment interaction models that attempt to match learning styles with instructional methods (e.g., identify learning deficiencies and help student compensate for deficient processes; identify learning strengths and teach to those strengths). There is little empirical support for aptitude-treatment interaction models, although the use of assessment to identify underlying learning deficiencies and then match the student to an intervention is widely used (Shapiro, 2004). In contrast, teaching practices like those assessed with the IP Scale are designed for frequent progress monitoring and data collection about an individual student to inform instructional methods and interventions.

the IP Scale prompted respondents to answer based on the practices used with students in their classroom struggling academically. Response choices for these items were on a five-point Likert scale (1= Almost Never, 2= A Few Lessons a Week, 3= A Couple Lessons a Day, 4= Almost Every Lesson Per Day, 5=Every Lesson Per Day). The last 4 items prompted respondents to answer based on practices used with students in their classroom with behavioral difficulties. Response choices were on a different five-point Likert Scale (1= Never, 2= Rarely, 3= Sometimes, 4= Often, 5= Always). The survey was piloted with six elementary school teachers and revised based on feedback (Kaiser & Rosenfield, 2006).

This version of the survey was administered to samples of approximately 1,600 teachers in 2006 and 2007. Scale reliability was consistently high, with alpha coefficients ranging from 0.90 to 0.91 (Kaiser, 2007). However, after the second year of data collection, the research team decided to investigate the validity of the scale since teachers' ratings were higher than expected and developers wanted a measure that would be sensitive to the IC Teams intervention. Kaiser (2007) reported that scale items were negatively skewed (means = 3.76 to 4.11, SDs = .12 to .75), suggesting a possible ceiling effect.

Before the survey was administered for the third year of data collection, Berger and Kaiser (2008) conducted a small validity study using a cognitive interviewing method adapted from Karabenick et al. (2007). Nine teachers from schools both with and without IC Teams were asked to think aloud as they responded to survey items. Interviewers asked participants to respond to four prompts: which answer would you choose, what is this question trying to find out from you, can you explain why you chose this answer, and can you give me some examples? Berger and Kaiser identified three common themes from the teacher interviews that implied problems with item interpretation.

First, teachers often answered items with their entire class in mind, instead of focusing on what they did for their struggling students. For example, one teacher responded that she frequently engaged in the practice of assessing the student's prior knowledge and skills before teaching a lesson. When asked to provide an example, she explained that she always conducted a "question and answer" session with the entire class before starting a new lesson.

Second, many teachers had difficulty with the response choices for the items about students struggling academically; the responses were on a five-point scale (1= Almost Never, 2 =A Few Lessons a Week, 3= A Couple Lessons a Day, 4= Almost Every Lesson Per Day, 5= Every Lesson Per Day). Teachers were often confused about how to choose a response when they engaged in practices for some subjects but not others.

Third, most teachers expressed a discrepancy between the ideal (i.e., what they would like to or plan to do) versus the actual (i.e., what they have time to do). For example, one teacher responded "I try [to develop my lesson so that I do not have the student work on too much unknown material at once]… that's my goal, but sometimes I get into the lesson and realize I have to switch gears." Teachers experiencing this discrepancy sometimes answered the item for the ideal and sometimes answered for the actual scenario.

Consequently, Berger and Kaiser (2008), in collaboration with the research team, made several revisions to the survey before it was administered in 2008 and 2009. The scale was shortened from 20 to 18 items by removing three items that were confusing or unnecessary ("I use classroom structures to free my time to work individually with students who need my help," "I have students work in pairs or small groups," "I have routines to ensure smooth transitions between tasks") and separating one existing item into two ("I set short term goals for this student" and "I collect data on this student to monitor progress toward short-term goals"). Response options for the questions about students struggling academically were changed to match the response options for items about students struggling behaviorally (1= Never, 2= Rarely, 3= Sometimes, 4= Often, 5= Always).

Question stems that separated items by category (i.e., planning lessons, delivering instruction, and assessing students) were either removed or, when necessary, incorporated into the item itself to improve the likelihood that teachers would read key components of each question. All items that were confusing to at least two of the nine teachers interviewed were revised by removing unnecessary words and phrases and/or clarifying the main idea of the question. To focus the teacher's attention on their struggling students, teachers were asked to think of a specific student in their classroom experiencing academic or behavior difficulties and to write the student's initials in a text boxbefore each section. Teachers were prompted throughout the survey to answer the questions with this student in mind.

To date, no follow-up interviews have been conducted to determine if teachers interpreted or responded to items on the revised scale differently. However, item means, variances, and inter-item correlations for the revised scale administered 2008 were similar to data collected using the original scale in 2007. Thus it is unlikely that survey revisions made a substantial impact on how teachers perceived items, although teachers may have had an easier time responding to the revised survey.

**Research.** Vu (2009) examined classroom characteristics related to student achievement and found that teaching practices measured by the IP Scale were significantly correlated with academic achievement for grades three through five (coefficients ranging from .07 to .08, $p < 0.05$). Vu also found that teachers reporting higher self-efficacy also reported more frequent use of good instructional practices with struggling students. Since Vu's cross-sectional study of the scale found a small, but significant relationship between teacher instructional practices and student achievement, a more thorough, rigorous investigation of the validity of the survey is warranted.

The current study examines the correlation between instruction and achievement with more recent data and attempts to make causal inferences about the effect of instruction on student achievement by establishing temporal precedence in a longitudinal design and ruling out alternative explanations with statistical controls. This study examines math and reading achievement outcomes specifically and in isolation to determine if instructional practices have different effects on these two critical content areas.

The current study also fills a gap in the literature by specifically examining how teaching practices are related to the academic achievement of lower performing students in the classroom. Much of the existing teacher effectiveness research has examined how all students or the average students in the classroom are influenced by teacher attributes and practices (e.g., D'Agostino, 2000; Guarino, et al., 2006; Nye, Konstantopoulos, & Hedges, 2004). Experimental data have suggested that the effects of high quality teaching are largest for students struggling academically (Sanders & Horn, 1998). There is also some evidence that a disproportionate number of minority students are assigned to less effective teachers (Sanders & Horn, 1998). As educational reform efforts work to decrease the achievement gap, it is crucial to determine which practices teachers are using with their struggling learners and how these practices are related to achievement outcomes.

**Research Questions**

The present study examines the validity and reliability of the Instructional Practices Scale, a teacher self-report measure of teaching practices used in the classroom. Four research questions are addressed:

1. Is the Instructional Practices Scale correlated with other teacher characteristics (i.e., education, experience, efficacy)?
2. Are teachers' scores on the Instructional Practices Scale stable over time?
3. Do teaching practices, as measured by the Instructional Practices Scale, affect students' academic achievement?
4. Do these teaching practices differentially affect the achievement of low versus high performing students?

# Method

## Participants

Data were obtained in a larger study, Experimental Evaluation of Instructional Consultation Teams, conducted by researchers at the University of Maryland. The study is a longitudinal randomized controlled trial investigating the effects of Instructional Consultation Teams on student, teacher, and school outcomes over four years. The present sample consists of two cohorts of classroom teachers of grades three through five and the students in their classrooms. The 2007 cohort included 595 classroom teachers and the 13,144 students in their classrooms. The 2008 cohort included 567 classroom teachers and the 13,238 students in their classrooms. All teacher and student data were obtained from 45 schools in Prince William County Public Schools in Virginia. Teachers of students in grades three, four, and five who taught more than 10 students were included.

Teachers were mostly Caucasian (85%) and female (88%), and were evenly spread among grade levels. Students were evenly spread across grade level and there were about equal numbers of boys and girls in the sample. Students were mostly Caucasian (43%), Hispanic (25%), and African American (21%). Approximately one-third of students participated in the free or reduced meal (FARM) program, English Speakers of Other Languages (ESOL) students made up almost a quarter (23%) of the sample, and 11% of students received Special Education services. Table 1 shows teacher and student participant characteristics by cohort.

## Procedures

A self-report survey about teacher beliefs, attitudes, and practices was administered online, using the SurveyMonkey website. Prior to survey administration, teachers received a small incentive gift (a notepad) and memo explaining the project. On the first day of data collection, teachers received an email and paper-copy letter with an invitation to participate and directions for completing the survey. Before survey questions were administered, teachers provided informed consent to participate and confirmed that their participation in the survey was voluntary. Every few weeks, teachers who had not yet completed the survey received an email reminder requesting their participation. Data were collected in the spring of 2006, 2007, and 2008. In the current sample, the survey response rate ranged from 87 to 92 percent. For the larger teacher survey, response rates ranged from 84 to 89 percent across the four years of data collection (Vu et al., 2009).

The school district provided data on teacher and student demographics and student achievement. In elementary schools, standardized (Standards of Learning or SOL) tests in reading and mathematics are administered to students in grades three, four, and five (Virginia Department of Education, n.d.). The school district provided data files containing students' scaled scores on the SOL tests and students' report card grades for 2006, 2007, and 2008.

## Measures

**Instructional Practices Scale (IP Scale).** The measure of teaching practices is a composite score based on the 18 to 20 items[2] on the Instructional Practices Scale

---

[2] The 2006 and 2007 versions of the IP Scale had 20 items; the 2008 version of the IP Scale had 18 items.

from the teacher survey. Specifically, the teacher's composite score on the scale over two or three years was used in the analyses (average of 2006 and 2007 scores for the 2007 cohort, average of 2006, 2007, and 2008 scores for the 2008 cohort). Since longitudinal data were available, the composite scores were used to provide the most reliable estimates of teacher-reported practices. Since the scale was revised in 2008 and may have influenced study results, I examined the correlation between the two IP Scale composites and their relationships with the dependent variables. There was a very high correlation between the two-year and three-year means of the IP Scale ($r$=.94). The correlations between the two-year and three-year means of the scale with each outcome measure were similar and in the same direction, thus, the 2008 scale revisions are unlikely to influence the results from this study.

The Instructional Practices Scale (IP Scale) is intended to measure the teaching practices of the teachers within the classroom (Kaiser, 2007). The scale specifically asks teachers to rate the frequency with which they use various instructional practices with students struggling academically or behaviorally in their classrooms. Sample items for this scale are "I develop my lesson so that I do not have the student work on too much unknown material at once," "I take the time to assess the student's prior knowledge and skills before teaching a lesson," and "I set and monitor progress towards short-term goals." The items are rated on a 5-point Likert-type scale (1= Almost never, 2 =A few lessons a week, 3= A couple lessons per day, 4= Almost every lesson per day, and 5= Every lesson per day). The development of the IP Scale is described in detail in the introduction of this paper, the rationale for each item is presented in Appendix A, and copy of the 2008 scale is presented in

Appendix B. The reliability of the IP Scale in the current sample was consistently high, with alphas of .91 each year. Alpha coefficients for the scale in this sample are almost identical to those reported by Vu et al. (2009).

**Student achievement test scores.** One dependent variable of interest is students' reading and math achievement as measured by student performance on the Standards of Learning (SOL) tests. The SOL tests are a set of standardized achievement tests used to assess student learning and achievement in Virginia public schools. SOL test scores were used as a measure of reading and math achievement hypothesized to be independent of teacher grading biases.

**Report card grades.** A second dependent variable is provided by report card grades. In grades three through five, reading and mathematics grades are provided by teachers for four nine-week grading periods and a final yearly grade. Teachers were asked to convert letter grades to numeric grades ranging from .00 to 4.00 (Prince William County Public Schools, 2004). The final yearly report card grade in numeric format was used to measure teacher-reported academic achievement in reading and math.

**Student prior achievement.** Prior achievement was used as a student-level and classroom-level covariate. When the outcome measure was performance on the Standards of Learning (SOL) tests, students' test scores in the same subject area from the previous year were used as the covariate. When the outcome measure was the final yearly report card grade, students' report card grades in the same subject area from the previous year were used as the covariate. Since third grade students in the outcome year had no prior SOL test scores, the only outcome measures used for third grade students were report card grades in reading and math.

Average prior achievement at the classroom level was also used as a covariate in the level 2 model. For each classroom in the outcome year, students' achievement from the prior year was aggregated to control for the average incoming achievement level for that classroom. This aggregate is designed to be a measure of the classroom's level of achievement when entering the classroom for the outcome year. The incoming achievement level across students in the classroom likely influences student achievement that year and may also affect the teacher's use of instructional practices.

**Student academic risk factors.** To capture demographic variables that may place a student at risk for academic problems, a student academic risk variable was created from three dichotomous variables: ethnic minority status (0=White or Asian, 1=African American, Hispanic, American Indian, Hawaiian), eligibility for free or reduced meals (0=Not eligible, 1=Eligible), and English as a second language status (0=English is not a second language, 1=English is a second language). Special education status was explored as a possible fourth risk factor, but since it was not highly correlated with the other risk variables, it was not included in the final academic risk variable. Academic risk ranged from zero (i.e., no risk factors) to three (i.e., all three risk factors).

An aggregate of student academic risk at the classroom level was also included as a covariate in the classroom-level model.

**Teacher experience and education.** Teaching experience and education level were examined as possible controls at the classroom level. Kaiser (2007) found that teacher experience was related to higher reported use of instructional practices. Teachers' experience and education may also be related to student achievement. On the Teacher Self Report survey, teachers were asked to provide demographic information including years of teaching experience and highest educational degree. Teaching experience was measured by teachers' response to "How many years have you been a teacher?" and responses were recorded on an ordinal scale from 1 to 5 (1= 1 year or less, 2= 2-5 years, 3= 6-10 years , 4= 11-20 years, 5= More than 20 years). Teacher education was measured by teachers' response to "What is the highest level of education you have earned?" and responses were also recorded on an ordinal scale from 1 to 5 (1= Bachelor's degree, 2= Bachelor's degree and additional coursework, 3= Master's degree,4= Master's degree and additional coursework, 5= Doctorate).

**Teaching Efficacy.** Vu (2009) found that teacher efficacy was related to use of instructional practices. Teacher efficacy may also be related to student achievement, thus it was explored as a possible covariate in the classroom-level model. Teacher efficacy was self reported on the teacher survey. It refers to the extent to which a teacher believes in the efficacy of his or her teaching to overcome student learning or behavioral problems, and it indicates to what extent a teacher judges his or her capabilities to bring about desired outcomes of student engagement and learning. Vu et al. (2009) found that IC Teams significantly increased teacher self-efficacy. Sample items for this 16-item scale include, "How much can you do to control disruptive behavior in the classroom," "How much can you do to adjust your lessons to the proper level for individual students," and "How much can you do to increase the achievement of a student who has a specific learning disability?" The items are rated on a 5 point Likert-type scale (1= Nothing/Not at all, 2= Very little, 3= Some, 4= Quite a bit, 5= A Great Deal). Teachers' composite efficacy score from the prior year was used as a covariate at the classroom level. The reliability of the Teaching Efficacy Scalefor the current sample was consistently high, with alphas ranging from .91 to .92. This scale was also used to examine the discriminant validity of the Instructional Practices Scale.

Descriptive statistics of student and classroom variables by grade level and cohort are reported in Appendix C.

**Data Analysis**

**Analysis for Question 1: Is the Instructional Practices Scale correlated with other teacher characteristics?** The first research question examines the correlation of the Instructional Practices Scale (IP Scale) with other teacher characteristics. I addressed this question by creating a correlation matrix showing the correlation coefficients ($r$) between teachers' composite score on the Instructional Practices Scale and years of teaching experience, educational level, and teacher self-efficacy.

**Analysis for Question 2: Are teachers' scores on the IP Scale stable over time?** The second research question addresses the stability of teachers' ratings on the Instructional Practices scale over three years. I answered this question by calculating the correlation coefficients between composite scores from each year of the study. I also contrasted these correlations with the correlation between the IP Scale and another self-reported scale from the same survey that measures Teaching Efficacy. This allowed me to examine both convergent and discriminant validity.

**Analysis for Research Question 3: Do instructional practices affect students' academic achievement?** In order to address the impact of instructional practices on low achieving students, I first explored a possible causal relationship between instructional practices and student achievement for all students, regardless of achievement level. Path analysis (Figure 1) was used to explore how instructional practices affected student achievement while statistically controlling for student characteristics (prior achievement and academic risk), teacher characteristics (experience, education, and efficacy), and classroom characteristics (mean prior achievement and mean academic risk). Teacher education level was excluded to simplify the models since it did not have a consistent, significant relationship with student achievement when statistical controls were in place at level 1.

Hierarchical linear modeling (HLM) is useful for studying contextual effects of educational settings such as the school or classroom (Lee, 2000). HLM specifies models that statistically control for variables at different levels, allowing for integration of student-level and classroom-level data (Osborne, 2000). A two-level hierarchical linear model in which students are nested within classrooms tested the effects of instructional practices on student achievement. The level-1 (student-level) model is:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1_{ij}} - \overline{X}_{1..}) + \beta_{2j}(X_{2_{ij}} - \overline{X}_{2..}) + r_{ij} \qquad (1)$$

where $Y_{ij}$ is the student achievement outcome for the $i^{th}$ student in the $j^{th}$ classroom,

$\beta_{0j}$ is the intercept or the average covariate-adjusted student achievement outcome in the $j^{th}$ classroom,

$\beta_{1j}$ is the slope of the regression of current achievement on prior achievement in the $j^{th}$ classroom,

$\beta_{2j}$ is the slope of the regression of current achievement on academic risk in the $j^{th}$ classroom,

$X_{1ij}$ is the grand-mean centered value of prior achievement (grades or test scores) for individual $i$ in classroom $j$,

$X_{2ij}$ is the grand-mean centered value of the academic risk for student $i$ in classroom $j$ for the baseline year, and

$r_{ij}$ is residual error for student $i$ in classroom $j$.

For each outcome variable, the level-1 covariates are the student's score for a comparable achievement measure administered the prior year and the student's academic risk level at baseline. First, I tested the homogeneity of the slopes of the covariates (prior achievement and academic risk) on the outcome measures (variables were classroom centered for this test). When covariates had significantly varying slopes, I left the slopes free to vary in future analyses.

The level-2 model (classroom-level model) consisted of three equations, with equations 3 and 4 depending on whether the slopes were fixed or random for a particular outcome:

$$\beta_{0j} = \gamma_0 + \gamma_1 W_{1_j} + \gamma_2 W_{2_j} + \gamma_3 W_{3_j} + \gamma_4 W_{4_j} + \gamma_5 W_{5_j} + u_{ij} \qquad (2)$$

(3a)

$$\beta_{1j} = \gamma_0 + \gamma_1 W_{1_j} + \gamma_2 W_{2_j} + \gamma_3 W_{3_j} + \gamma_4 W_{4_j} + \gamma_5 W_{5_j} + u_{ij} \qquad (3b)$$

$$\beta_{1j} = \gamma_0 \qquad (4a)$$

$$\beta_{2j} = \gamma_0 + \gamma_1 W_{1_j} + \gamma_2 W_{2_j} + \gamma_3 W_{3_j} + \gamma_4 W_{4_j} + \gamma_5 W_{5_j} + u_{ij} \qquad (4b)$$

where $\beta_{0j}$ is the average covariate-adjusted mean for the $j^{th}$ classroom,

$\gamma_0$ is the grand mean student outcome for all teachers,

$\gamma_1$ is the estimated effect of instructional practices on student outcome,

$W_{1j}$ is the teachers' composite score on the Instructional Practices Scale,

$W_{2j}$ is the baseline mean achievement of the current classroom,

$\gamma_2$ represents the effect of classroom mean prior achievement on the student outcome,

$W_{3j}$ is the current year covariate percentage of students with disadvantaged status,

$\gamma_3$ represents the effect of classroom-level disadvantaged status on the student outcome,

$W_{4j}$ is the teacher's years of experience reported at baseline,

$\gamma_4$ represents the effect of teacher experience on the student outcome,

$W_{5j}$ is the composite of teacher reported efficacy at baseline,

$\gamma_5$ represents the effect of teacher efficacy on the student outcome,

$u_{ij}$ is residual error for the $j^{th}$ classroom.

I tested the hypothesis that the intercept in equation 1 ($\beta_{0j}$) can be predicted with equation 2. That is, student achievement is a function of teacher instruction when prior achievement, academic risk, and other teacher and classroom characteristics are statistically controlled for. Teacher education was dropped from the models because it was not significantly related to student achievement when statistical controls were in place at the student level.

Equations 3 and 4 model the coefficients of the covariates in the level-1 model (equation 1) and were used when covariates had significantly varying slopes. Equations 3a and 4a were used when statistical testing suggested that the slopes can defensibly be fixed. In these cases, I did not further investigate an interaction between current and prior achievement and variables were grand-mean centered at level 1. Equations 3b and 4b were used to predict the coefficients of the covariates in the level-1 model when the slopes were random. In these cases, I grand-mean centered level-1 variables to test the main effect of instruction on achievement, and group-centered level-1 variables to examine cross-level interaction(s).

**Research Question 4: Do teachers' instructional practices differentially affect the achievement of low versus high performing students?** I addressed a possible interaction effect by examining the effect of instructional practices on the slope of the relationship between students' prior and current achievement. An interaction term was added at level 2 to explore the hypothesis that the effect of instruction on achievement would be most pronounced for students with incoming low achievement scores.

## Results

### Results for Question 1: Is the Instructional Practices Scale correlated with other teacher characteristics?

The correlations between teacher instructional practices and years of experience, education level, and self-reported teaching efficacy were explored. The correlation coefficients are reported in Table 2. There was a significant positive relationship between years of experience and instructional practices (*r*=.11-.13) and between education level and instruction (*r*=.10-.12). Self-reported teaching efficacy also had a significant positive relationship with instructional practices (*r*=.56-.57).

**Results for Question 2: Are teachers' scores on the Instructional Practices Scale stable over time?**

There was a significant positive relationship between teacher instructional practices as measured in 2006, 2007, and 2008. Correlation coefficients ranged from .57 to .64 (Table 3). There was also a significant positive relationship between teacher efficacy as measured in 2006, 2007, and 2008, with correlation coefficients ranging from .57 to .66.

When the correlations between scales were examined, there was a significant positive relationship between teacher instructional practices as measured in each year with teacher efficacy as measured in each year. Correlation coefficients between instructional practices and efficacy ranged from .42 to .57 (Table 3).

**Results for Research Question 3: Do instructional practices affect students' academic achievement?**

The effects of instruction on reading and math grades were examined for third, fourth, and fifth grade students, and the effects of instruction on reading and math standardized tests were only examined for fourth and fifth graders. Results are reported by grade level and outcome year in Tables 4 to 9.

Model 1 only includes student-level predictors. Prior achievement and academic risk were significant predictors of student achievement in the hypothesized direction for each grade level, outcome measure, and cohort. In Model 2, Instructional Practices was added at the classroom level without other level-2 covariates. In Model 3, other classroom-level covariates were added, including mean prior achievement, mean academic risk, teacher years of experience, and teacher efficacy. Significant results from the final models (Model 3) are discussed in the following paragraphs.

Tables 7 and 8 show that for reading achievement outcomes, Instructional Practices was a significant positive predictor of reading grade and standardized test outcomes for fifth graders in the 2007 cohort only. For reading report card grades, coefficients ranged from .03 to .10 with a median of .08. An interpretation of the median coefficient is that a one-point increase on the Instructional Practices Scale yields less than a tenth of a grade point improvement in reading. In other words, a 1-point increase on the teacher's Instructional Practices (which is more than 2 standard deviations on the scale) only improves a student's grade in reading by approximately 10 percent of a standard deviation. For reading standardized test scores, coefficients ranged from -6.87 to 7.33 with a median of 3.60 (the negative coefficient was not significant).An interpretation of the median coefficient is that a one-point increase on the Instructional Practices Scale yields an improvement of less than 4 scaled score points on the standardized test in reading. In other words, a 1-point increase on the teacher's Instructional Practices only improves a student's reading test score by approximately 5 percent of a standard deviation.

Tables 4 and 5 show that for math achievement outcomes, instruction was a significant positive predictor of math grades for third graders in the 2007 cohort only. Instruction did not predict math standardized test scores for any grade in either cohort. For math report card grades, coefficients ranged from -.06 to .12 with a median of .06 (the negative coefficient was not significant). An interpretation of the median coefficient is that a one-point increase on the Instructional Practices Scale yields less than a tenth of a grade point improvement in math. In other words, a 1-point increase on the teacher's Instructional Practices only improves a student's reading grade by approximately 8 percent of a standard deviation. For math standardized test scores, coefficients ranged from -5.05 to .88 with a median of -2.15. An interpretation of the median coefficient is that a one-point increase on the Instructional Practices Scale decreases the math standardized test score by approximately two scaled score points. Two-scaled score points is less than 3 percent of a standard deviation on the math standardized test.

**Research Question 4: Do teachers' instructional practices differentially affect the achievement of low versus high performing students?**

I found no interaction effects—providing no evidence in support of my hypothesis that high instructional practices would improve student achievement most when students' prior achievement is low.

I explored teachers' missing data on their instructional practices from year to year. I found no significant differences among teachers with and without missing data in terms of ethnicity, gender, and years of experience.

## Discussion

In the present study, I examined the validity and reliability of the Instructional Practices Scale (IP Scale), a self-report measure designed to capture effective teaching practices used in the general education classroom. This validation study is part of a larger experimental evaluation, and its findings are useful in the interpretation of results of the evaluation. On a broader level, this study contributes to the discussion around the utility of self-report measures in educational research. While self-reports are easy to administer and used widely in large-scale research, the results of this study suggest that when measuring loosely defined constructs such as “effective instruction,” the cost of using a self-report measure may outweigh these benefits.

Correlations between the Instructional Practices Scale and education, years of experience, and Teaching Efficacy were in the expected direction and statistically significant. Between instruction and education and years of experience, coefficients were small ($r$=.11 and $r$=.13 respectively), suggesting a positive but weak relationship.

Teacher education level was not a significant predictor of student achievement for outcomes when student-level covariates were included, and it was dropped from the analyses. While our measure of education was an ordinal scale ranging from a bachelor’s degree to a doctorate (rather than distinguishing between traditional and alternative teacher training programs), the results are similar to Constantine et al. (2009) in suggesting that other teacher factors seem more important to student outcomes than teacher education.

There was a positive, moderate correlation between self-reported instructional practices and teaching efficacy expectations (*r*=.56). However, when the correlation of Instruction Practice Scale scores were examined over three years (*r*'s ranged from .57 to .64) and compared to correlation coefficients between the IP Scale scores and scores on the Teaching Efficacy Scale (*r*'s ranged from .42 to .57), the data provided little support for the discriminant validity of these two sets of self-reports. It appears that the Instructional Practices Scale and the Teaching Efficacy Scale are not measuring distinct constructs. It is possible that the Instructional Practices Scale also captures beliefs about teaching and a teacher's sense of ideal practices rather than specifically measuring how often they use these effective practices in the classroom.

This study also examined the predictive validity of the Instructional Practices Scale by looking at how instruction predicted math and reading achievement for third, fourth, and fifth grade students. Analyses were conducted using statistical software that allows for hierarchical data in which students were nested within classrooms. With statistical controls for prior achievement and academic disadvantage in place at both levels of the model and additional teacher characteristics at the classroom level, instruction was only a significant predictor for 3 out of 20 possible outcomes, with no clear pattern among the results.

Instruction significantly predicted better reading grades and standardized test scores for fifth graders in 2007, but the result was not replicated with the 2008 cohort. Instruction was a significant positive predictor of math grades for third grade students in 2007, but the effect was not significant with the 2008 cohort. There were no significant results for the effect of instruction on math standardized test scores, and coefficients were primarily negative. Even when effects were significant and in the hypothesized direction, coefficients were small and suggested that large increases in teacher's self-reported instruction yielded very small improvements in student's achievement. For some outcomes, multicollinearity among level 2 variables was a potential problem since the effect of instruction appeared to increase when covariates were added.

There was no interaction between instruction and prior achievement for any grade, outcome measure, or outcome year. The lack of interaction effects suggests that self-reported instruction did not differentially affect the current achievement of students who came into the classroom with differing levels of prior achievement. This is particularly problematic since the primary goal of the scale was to measure the teaching practices specifically used with struggling learners.

These findings suggest that the Instructional Practices Scale does not consistently predict student achievement, regardless of entering achievement level. The scale may be slightly more useful in predicting reading achievement, which is not surprising given that some of the items seem more geared toward reading than math instruction.

**Limitations**

There are a number of limitations to this study, including missing data, fallibility of measures, imperfect research design, and issues with generalizability. Due to the longitudinal nature of this study, there were missing data at both the teacher and student level each year of the study. An exploration of missing data suggested no apparent demographic differences between the population that remained in the sample and the population that left the sample. However, there may have been unknown differences between the populations that were not measured and influenced the results.

It is possible that the choice of variables used in this study made it difficult to discern effects. Teachers tended to rate themselves highly on Instructional Practices, suggesting a possible ceiling effect of the independent variable. In addition, the Instructional Practices Scale was revised during the study so that teachers completed a different version of the scale in 2008. It is possible that the changes made in 2008 impacted results, though it is unlikely given that instruction for the 2008 cohort was a composite measure that averaged 2006, 2007, and 2008 scores. The composite measure would also have the advantage of decreasing the negative skew.

Student outcomes, the dependent variables, were measured by report card grades and standardized test scores, both having limitations. Grades are designated by teachers, thus are inherently subjective and biased by teacher beliefs and opinions (Bruckman, 2009). Scores on standardized tests are more objective, but they may not measure what student are learning in the classroom, despite the intention of the school district to measure the results of instruction. Report card grades and standardized test scores were both negatively skewed, suggesting a potential problem with ceiling effects. While grades and test scores were the best available achievement data in this study, it is possible that neither measure is sensitive enough to detect small but potentially important improvements for students, especially those struggling in the classroom.

Research methodology used in this study was imperfect. The Pearson correlation coefficient was used to examine relationships between variables. This statistic does not take clustering of variables into account and may have overestimated the significance of the relationship between variables studied, so significance levels reported are nominal. While I attempted to statistically control for confounding variables in the two-level models, it is possible that important constructs were not included—though omitted variables would have led to an overestimate of effects. For example, special education status was initially explored as an academic risk variable but it was excluded due to low correlations with the other risk variables under study. However, it may have been informative to study the interaction between student eligibility for special education services and teacher-reported instructional practices. Future research might measure teaching practices used with different groups of students, including both students with varying levels of incoming achievement and students eligible for special education services.

While the results of this study suggest that the Instructional Practices Scale is not a great measure of the effective teaching practices that improve student achievement, it does not invalidate other self-report questionnaires designed with a similar purpose. It is not clear whether the problem lies in the methodology, the particular measure used, or in the teaching practices emphasized in the program the scale was designed to evaluate. A different self-report of teaching practices could provide a better measure of what teachers are doing in the classroom to enhance student achievement, or teaching practices might best be measured using other approaches such as observation or teacher logs. It is also possible that the current study would have yielded different results if a different sample of the population was used. Results of the current study may not generalize to teachers and students in another locale.

**Implications and Virtues of Current Research**

Despite these limitations, there are a number of strengths to this research. The use of multiple outcomes, at three grade levels, for two cohorts allowed me to comprehensively explore the utility of the Instructional Practices Scale being used in the experimental evaluation of IC Teams. Separating the analyses by grade level made it possible to determine if the effects of instruction might have been influenced by the student's age, the curriculum, or another factor particular to that grade level. With two cohorts for each grade level, I was able to replicate the same model for each grade level to see if significant effects held. Further, with a longitudinal design, I was able to statistically control for variables prior to the outcome year and also to use a measure of instruction that was more reliable since it was averaged across years.

The current study is important to understanding results from a large-scale, experimental evaluation of IC Teams and to the discussion of appropriate methodology for educational research. As results from the IC Teams evaluation are interpreted, it isimportant to keep in mind that the Instructional Practices Scale, the only measure of teacher's instruction used in the study, has limited validity. An effect or lack of effect of IC Teams on scores on the Instructional Practices Scale will tell us little about mechanisms for mediated effects on student achievement.

While this validation study only examined a single self-report measure of instruction, it a critical step in improving research on teacher effectiveness. Self-reports are widely used in this type of research despite their limitations. Self-report measures should continue to be refined as we develop a better understanding of effective instruction. Since there is a dearth of existing, reliable, valid measures of instruction, researchers may wish to develop their own measures. Prior to using a newly developed self-report instrument in a large-scale study, researchers should pilot the instrument, revise items based on feedback and results, and re-evaluate the measure. Further, when having a precise measure of instruction is critical to study outcomes, it should be best practice to use multiple methodologies instead of relying solely on self report (Camburn & Barnes, 2004).

Table 1
*Participant Characteristics*

| | 2007 Cohort | | 2008 Cohort | |
|---|---|---|---|---|
| Characteristics | Teachers (*N*=595) | Students (N=13,144) | Teachers (*N*=567) | Students (*N*=13,238) |
| Gender | | | | |
| Female | 88 | 51 | 88 | 50 |
| Male | 12 | 49 | 12 | 50 |
| Ethnicity | | | | |
| Caucasian | 84 | 43 | 85 | 43 |
| African American | 8 | 20 | 9 | 21 |
| Hispanic | 2 | 25 | 2 | 24 |
| Asian | 1 | 7 | 1 | 7 |
| Other | 4 | 4 | 3 | 5 |
| Grade Level | | | | |
| 3rd grade | 33 | 34 | 36 | 34 |
| 4th grade | 31 | 34 | 33 | 33 |
| 5th grade | 37 | 32 | 31 | 33 |
| FARM | - | 31 | - | 32 |
| ESOL | - | 23 | - | 22 |
| Special Education | - | 11 | - | 11 |

*Note.* Numbers shown indicate percentages. FARM indicates students receiving free and/or reduced meals. ESOL indicates students who are English Speakers of Other Languages.

Table 2

*Correlations between the Instructional Practices Scale and other teacher variables.*

| Characteristics | 2007 Cohort Instructional Practices (2006 & 2007) | 2008 Cohort Instructional Practices (2006, 2007, & 2008) |
|---|---|---|
| Teacher | | |
| Efficacy (prior year) | 0.54 | 0.58 |
| Level of education (outcome year) | 0.11 | 0.10 |
| Years of experience (outcome year) | 0.14 | 0.15 |

*Note*. For the 2007 Cohort, Instructional Practices refers to 2-year mean of teachers' composite score on the scale from 2006 and 2007. For the 2008 Cohort, Instructional Practices refers to 3-year mean from 2006, 2007, and 2008. Efficacy was measured in the year prior (i.e., 2006 efficacy for 2007 cohort, 2007 efficacy for 2008 cohort). Level of education and years of experience were measured in the same year as the outcomes for each cohort (i.e., in 2007 for 2007 cohort, in 2008 for 2008 cohort).

Table 3
*Correlations between Instructional Practices and Teaching Efficacy Scales.*

| Scale and Year | 1. | 2. | 3. | 4. | 5. |
|---|---|---|---|---|---|
| 1. Instruction 2006 | - | | | | |
| 2. Instruction 2007 | **0.64** | **-** | | | |
| 3. Instruction 2008 | **0.57** | **0.61** | - | | |
| 4. Efficacy 2006 | 0.55 | 0.44 | 0.42 | - | |
| 5. Efficacy 2007 | 0.47 | 0.56 | 0.50 | **0.62** | - |
| 6. Efficacy 2008 | 0.47 | 0.45 | 0.57 | **0.57** | **0.66** |

*Note.* Bolded entries show convergent correlations for the two scales. Unbolded entries show discriminant correlations between the two scales.

Table 4

*Summary of Analyses Predicting 2007 3rd Grade Report Card Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2006) | 0.47 | 0.02 | 0.00 | 0.47 | 0.02 | 0.00 | 0.47 | 0.02 | 0.00 |
| Academic risk | -0.13 | 0.01 | 0.00 | -0.13 | 0.01 | 0.00 | -0.1 | 0.02 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.04 | 0.05 | 0.43 | 0.09 | 0.05 | 0.08 |
| Mean prior achievement | | | | | | | -0.14 | 0.11 | 0.20 |
| Mean academic risk factor | | | | | | | -0.11 | 0.05 | 0.04 |
| Teacher years of experience | | | | | | | -0.03 | 0.02 | 0.11 |
| Teacher efficacy | | | | | | | -0.07 | 0.05 | 0.18 |
| Math Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2006) | 0.56 | 0.02 | 0.00 | 0.56 | 0.02 | 0.00 | 0.56 | 0.02 | 0.00 |
| Academic risk | -0.12 | 0.01 | 0.00 | -0.11 | 0.01 | 0.00 | -0.11 | 0.02 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.08 | 0.05 | 0.07 | 0.12 | 0.05 | 0.01 |
| Mean prior achievement | | | | | | | -0.25 | 0.11 | 0.03 |
| Mean academic risk factor | | | | | | | -0.14 | 0.05 | 0.01 |
| Teacher years of experience | | | | | | | -0.03 | 0.02 | 0.08 |
| Teacher efficacy | | | | | | | -0.06 | 0.04 | 0.21 |

*Note.* *N*=132 teachers, *N*=2456-2458 students. There was no interaction effect of instruction on the slope of prior achievement and current achievement. Instructional practices is the mean of teacher-reported practices from 2005-06 and 2006-07. Variables are not standardized, and they are grand centered.

Table 5

*Summary of Analyses Predicting 2008 3rd Grade Report Card Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2007) | 0.54 | 0.02 | 0.00 | 0.54 | 0.02 | 0.00 | 0.54 | 0.02 | 0.00 |
| Academic risk | -0.15 | 0.01 | 0.00 | -0.15 | 0.01 | 0.00 | -0.1 | 0.02 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.02 | 0.05 | 0.72 | 0.10 | 0.06 | 0.10 |
| Mean prior achievement | | | | | | | -0.09 | 0.11 | 0.39 |
| Mean academic risk factor | | | | | | | -0.09 | 0.05 | 0.05 |
| Teaching experience | | | | | | | -0.03 | 0.02 | 0.11 |
| Teacher efficacy | | | | | | | -0.10 | 0.05 | 0.05 |
| Math Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2007) | 0.57 | 0.02 | 0.00 | 0.57 | 0.02 | 0.00 | 0.57 | 0.02 | 0.00 |
| Academic risk | -0.14 | 0.01 | 0.00 | -0.14 | 0.01 | 0.00 | -0.14 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.04 | 0.05 | 0.40 | 0.07 | 0.06 | 0.31 |
| Mean prior achievement | | | | | | | 0.10 | 0.16 | 0.52 |
| Mean academic risk factor | | | | | | | 0.02 | 0.05 | 0.77 |
| Teaching experience | | | | | | | -0.02 | 0.02 | 0.31 |
| Teacher efficacy | | | | | | | -0.02 | 0.06 | 0.8 |

*Note.* *N*=138 teachers, *N*=2578-2579 students. There was no interaction effect of instruction on the slope of prior achievement and current achievement for any outcomes. Instructional practices is the mean of teacher-reported practices from 2005-06, 2006-07, and 2007-08. Variables are not standardized, and they are grand centered.

Table 6

*Summary of Analyses Predicting 2007 4th Grade SOLs and Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading SOL 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading SOL 2006) | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 |
| Academic risk | -5.52 | 1.20 | 0.00 | -5.40 | 1.17 | 0.00 | .4.79 | 1.36 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | -4.88 | 3.03 | 0.11 | -6.87 | 4.86 | 0.16 |
| Mean prior achievement | | | | | | | -0.04 | 0.07 | 0.55 |
| Mean academic risk factor | | | | | | | -3.09 | 3.11 | 0.32 |
| Teacher years of experience | | | | | | | 1.13 | 1.11 | 0.31 |
| Teacher efficacy | | | | | | | 4.49 | 5.00 | 0.37 |
| Math SOL 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math SOL 2006) | 0.76 | 0.02 | 0.00 | 0.76 | 0.01 | 0.00 | 0.76 | 0.02 | 0.00 |
| Academic risk | -7.84 | 1.26 | 0.00 | -7.86 | 1.25 | 0.00 | -8.59 | 1.36 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 2.09 | 4.30 | 0.63 | 0.88 | 5.83 | 0.88 |
| Mean prior achievement | | | | | | | -0.14 | 0.09 | 0.14 |
| Mean academic risk factor | | | | | | | 1.71 | 4.12 | 0.68 |
| Teacher years of experience | | | | | | | -2.81 | 1.64 | 0.09 |
| Teacher efficacy | | | | | | | 0.04 | 5.98 | 0.99 |
| Reading Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2006) | 0.55 | 0.02 | 0.00 | 0.55 | 0.02 | 0.00 | 0.55 | 0.02 | 0.00 |
| Academic risk | -0.08 | 0.01 | 0.00 | -0.08 | 0.01 | 0.00 | -0.07 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.07 | 0.05 | 0.19 | 0.03 | 0.05 | 0.5 |
| Mean prior achievement | | | | | | | -0.33 | 0.10 | 0.002 |
| Mean academic risk factor | | | | | | | -0.19 | 0.05 | 0.00 |
| Teacher years of experience | | | | | | | -0.01 | 0.02 | 0.38 |
| Teacher efficacy | | | | | | | 0.06 | 0.06 | 0.34 |
| Math Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2006) | 0.62 | 0.02 | 0.00 | 0.62 | 0.02 | 0.00 | 0.62 | 0.02 | 0.00 |
| Academic risk | -0.06 | 0.02 | 0.00 | -0.06 | 0.02 | 0.00 | -0.06 | 0.02 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.05 | 0.04 | 0.23 | 0.04 | 0.05 | 0.45 |
| Mean prior achievement | | | | | | | -0.20 | 0.08 | 0.02 |
| Mean academic risk factor | | | | | | | -0.12 | 0.04 | 0.01 |
| Teacher years of experience | | | | | | | -0.04 | 0.01 | 0.00 |
| Teacher efficacy | | | | | | | 0.01 | 0.05 | 0.83 |

*Note. N*=120 teachers, *N*=2232-2425 students. There was no interaction effect of instruction on the slope of prior achievement and current achievement for any outcomes. Instructional practices is the mean of teacher-reported practices from 2005-06 and 2006-07. Variables are not standardized, and they are grand centered.

Table 7

*Summary of Analyses Predicting 2008 4th Grade SOLs and Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading SOL 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading SOL 2007) | 0.58 | 0.02 | 0.00 | 0.58 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 |
| Academic risk | -5.99 | 1.13 | 0.00 | -5.98 | 1.13 | 0.00 | -5.34 | 1.37 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | -0.26 | 3.12 | 0.93 | 4.81 | 3.44 | 0.17 |
| Mean prior achievement | | | | | | | -0.03 | 0.06 | 0.56 |
| Mean academic risk factor | | | | | | | -3.35 | 3.23 | 0.3 |
| Teaching experience | | | | | | | -0.18 | 1.13 | 0.88 |
| Teacher efficacy | | | | | | | -9.74 | 3.50 | 0.01 |
| Math SOL 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math SOL 2007) | 0.69 | 0.02 | 0.00 | 0.69 | 0.02 | 0.00 | 0.69 | 0.02 | 0.00 |
| Academic risk | -5.56 | 1.19 | 0.00 | -5.54 | 1.19 | 0.00 | -5.72 | 1.22 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | -3.94 | 5.42 | 0.47 | -2.88 | 5.54 | 0.6 |
| Mean prior achievement | | | | | | | -0.25 | 0.12 | 0.03 |
| Mean academic risk factor | | | | | | | -3.41 | 4.82 | 0.48 |
| Teaching experience | | | | | | | -2.21 | 2.10 | 0.3 |
| Teacher efficacy | | | | | | | -2.63 | 5.79 | 0.65 |
| Reading Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2007) | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 |
| Academic risk | -0.07 | 0.01 | 0.00 | -0.07 | 0.01 | 0.00 | -0.07 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.04 | 0.05 | 0.41 | 0.06 | 0.06 | 0.27 |
| Mean prior achievement | | | | | | | -0.15 | 0.09 | 0.09 |
| Mean academic risk factor | | | | | | | -0.10 | 0.04 | 0.03 |
| Teaching experience | | | | | | | -0.03 | 0.02 | 0.09 |
| Teacher efficacy | | | | | | | -0.03 | 0.05 | 0.56 |
| Math Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2007) | 0.64 | 0.02 | 0.00 | 0.64 | 0.02 | 0.00 | 0.64 | 0.02 | 0.00 |
| Academic risk | -0.06 | 0.01 | 0.00 | -0.07 | 0.01 | 0.00 | -0.06 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.12 | 0.05 | 0.02 | 0.10 | 0.06 | 0.1 |
| Mean prior achievement | | | | | | | -0.13 | 0.12 | 0.3 |
| Mean academic risk factor | | | | | | | -0.10 | 0.06 | 0.09 |
| Teaching experience | | | | | | | -0.04 | 0.02 | 0.04 |
| Teacher efficacy | | | | | | | 0.06 | 0.06 | 0.34 |

*Note.* $N$=113 teachers, $N$=2295-2311 students. There was no interaction effect of instruction on the slope of prior achievement and current achievement for any outcomes Instructional practices is the mean of teacher-reported practices from 2005-06, 2006-07, and 2007-08. Variables are not standardized, and they are grand centered.

Table 8

*Summary of Analyses Predicting 2007 5th Grade SOLs and Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading SOL 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading SOL 2006) | 0.62 | 0.02 | 0.00 | 0.62 | 0.02 | 0.00 | 0.63 | 0.02 | 0.00 |
| Academic risk | -9.25 | 1.07 | 0.00 | -9.29 | 1.06 | 0.00 | -8.28 | 1.22 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 6.03 | 2.22 | 0.01 | 7.33 | 2.47 | 0.004 |
| Mean prior achievement | | | | | | | -0.05 | 0.07 | 0.49 |
| Mean academic risk factor | | | | | | | -5.73 | 3.19 | 0.07 |
| Teacher years of experience | | | | | | | 0.70 | 0.99 | 0.48 |
| Teacher efficacy | | | | | | | -3.44 | 3.22 | 0.29 |
| Math SOL 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math SOL 2006) | 0.76 | 0.02 | 0.00 | 0.76 | 0.02 | 0.00 | 0.76 | 0.02 | 0.00 |
| Academic risk | -2.90 | 1.15 | 0.01 | -2.90 | 1.15 | 0.01 | -2.36 | 1.20 | 0.05 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | -6.59 | 3.24 | 0.04 | -5.05 | 3.57 | 0.16 |
| Mean prior achievement | | | | | | | -0.23 | 0.09 | 0.01 |
| Mean academic risk factor | | | | | | | -10.75 | 4.86 | 0.03 |
| Teacher years of experience | | | | | | | 1.10 | 1.27 | 0.39 |
| Teacher efficacy | | | | | | | -1.57 | 4.33 | 0.72 |
| Reading Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2006) | 0.54 | 0.02 | 0.00 | 0.54 | 0.02 | 0.00 | 0.54 | 0.02 | 0.00 |
| Academic risk | -0.11 | 0.01 | 0.00 | -0.11 | 0.01 | 0.00 | -0.1 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.08 | 0.03 | 0.008 | 0.08 | 0.04 | 0.04 |
| Mean prior achievement | | | | | | | -0.01 | 0.10 | 0.92 |
| Mean academic risk factor | | | | | | | -0.08 | 0.05 | 0.11 |
| Teacher years of experience | | | | | | | -0.02 | 0.01 | 0.29 |
| Teacher efficacy | | | | | | | -0.02 | 0.05 | 0.71 |
| Math Grades 2007 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2006) | 0.65 | 0.02 | 0.00 | 0.65 | 0.02 | 0.00 | 0.65 | 0.02 | 0.00 |
| Academic risk | -0.06 | 0.01 | 0.00 | -0.06 | 0.01 | 0.00 | -0.06 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.06 | 0.04 | 0.10 | 0.04 | 0.05 | 0.45 |
| Mean prior achievement | | | | | | | -0.14 | 0.11 | 0.21 |
| Mean academic risk factor | | | | | | | -0.08 | 0.06 | 0.17 |
| Teacher years of experience | | | | | | | 0.01 | 0.02 | 0.74 |
| Teacher efficacy | | | | | | | 0.07 | 0.06 | 0.21 |

*Note.* $N$=135 teachers, $N$=2650-2815 students. There was no interaction effect of instruction on the slope of prior achievement and current achievement for any outcomes. Instructional practices is the mean of teacher-reported practices from 2005-06 and 2006-07. Variables are not standardized, and they are grand centerd.

Table 9

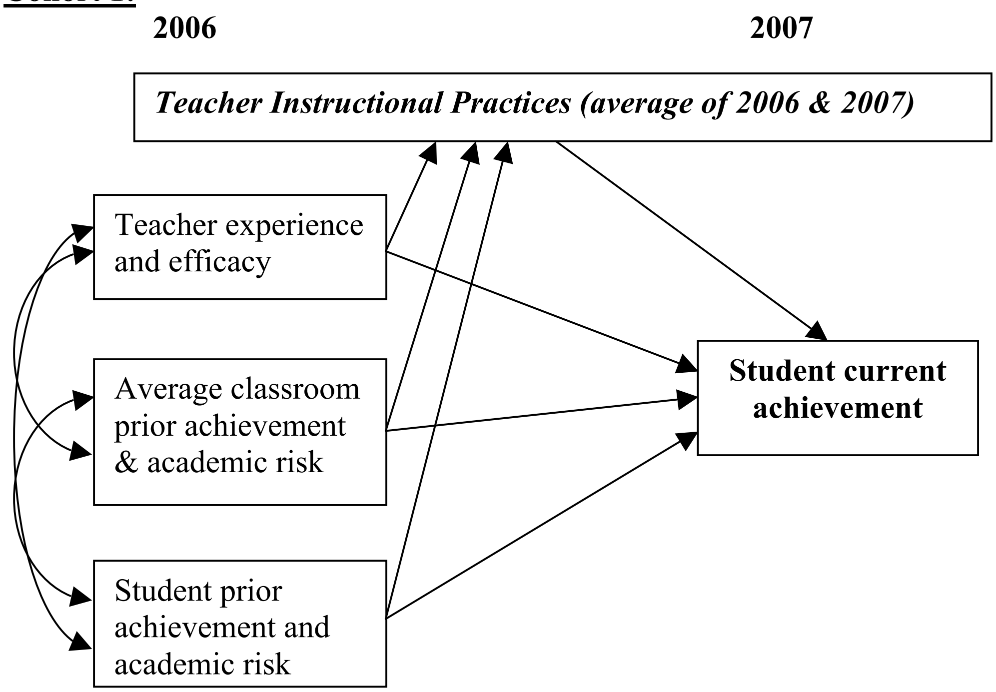
*Summary of Analyses Predicting 2008 5th Grade SOLs and Grades from Teacher Instructional Practices*

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Variables | Coef. | *SE* | *p* | Coef. | *SE* | *p* | Coef. | *SE* | *p* |
| Reading SOL 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading SOL 2007) | 0.60 | 0.02 | 0.00 | 0.60 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 |
| Academic risk | -9.49 | 1.15 | 0.00 | -9.48 | 1.14 | 0.00 | -9.60 | 1.30 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 3.50 | 2.10 | 0.10 | 2.38 | 2.91 | 0.42 |
| Mean prior achievement | | | | | | | 0.09 | 0.06 | 0.14 |
| Mean academic risk factor | | | | | | | 3.06 | 2.72 | 0.26 |
| Teacher years of experience | | | | | | | 1.23 | 1.11 | 0.27 |
| Teacher efficacy | | | | | | | 1.43 | 2.86 | 0.62 |
| Math SOL 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math SOL 2007) | 0.70 | 0.02 | 0.00 | 0.70 | 0.02 | 0.00 | 0.70 | 0.02 | 0.00 |
| Academic risk | -4.55 | 1.19 | 0.00 | -4.55 | 1.19 | 0.00 | -4.58 | 1.24 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.06 | 3.21 | 0.99 | -1.42 | 3.95 | 0.72 |
| Mean prior achievement | | | | | | | -0.14 | 0.08 | 0.09 |
| Mean academic risk factor | | | | | | | -3.64 | 4.69 | 0.44 |
| Teacher years of experience | | | | | | | -0.21 | 1.58 | 0.89 |
| Teacher efficacy | | | | | | | 3.28 | 4.18 | 0.44 |
| Reading Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Reading Grades 2007) | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 | 0.59 | 0.02 | 0.00 |
| Academic risk | -0.08 | 0.01 | 0.00 | -0.08 | 0.01 | 0.00 | -0.07 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.07 | 0.04 | 0.11 | 0.08 | 0.05 | 0.08 |
| Mean prior achievement | | | | | | | 0.12 | 0.10 | 0.23 |
| Mean academic risk factor | | | | | | | 0.02 | 0.05 | 0.72 |
| Teacher years of experience | | | | | | | -0.01 | 0.02 | 0.7 |
| Teacher efficacy | | | | | | | -0.02 | 0.05 | 0.72 |
| Math Grades 2008 (outcome) | | | | | | | | | |
| Student-level | | | | | | | | | |
| Prior achievement (Math Grades 2007) | 0.68 | 0.02 | 0.00 | 0.68 | 0.02 | 0.00 | 0.67 | 0.02 | 0.00 |
| Academic risk | -0.06 | 0.01 | 0.00 | -0.06 | 0.01 | 0.00 | -0.05 | 0.01 | 0.00 |
| Classroom-level | | | | | | | | | |
| Instructional Practices | | | | 0.07 | 0.03 | 0.04 | -0.06 | 0.04 | 0.12 |
| Mean prior achievement | | | | | | | 0.06 | 0.10 | 0.57 |
| Mean academic risk factor | | | | | | | 0.00 | 0.05 | 0.97 |
| Teacher years of experience | | | | | | | 0.00 | 0.02 | 0.79 |
| Teacher efficacy | | | | | | | 0.00 | 0.05 | 0.92 |

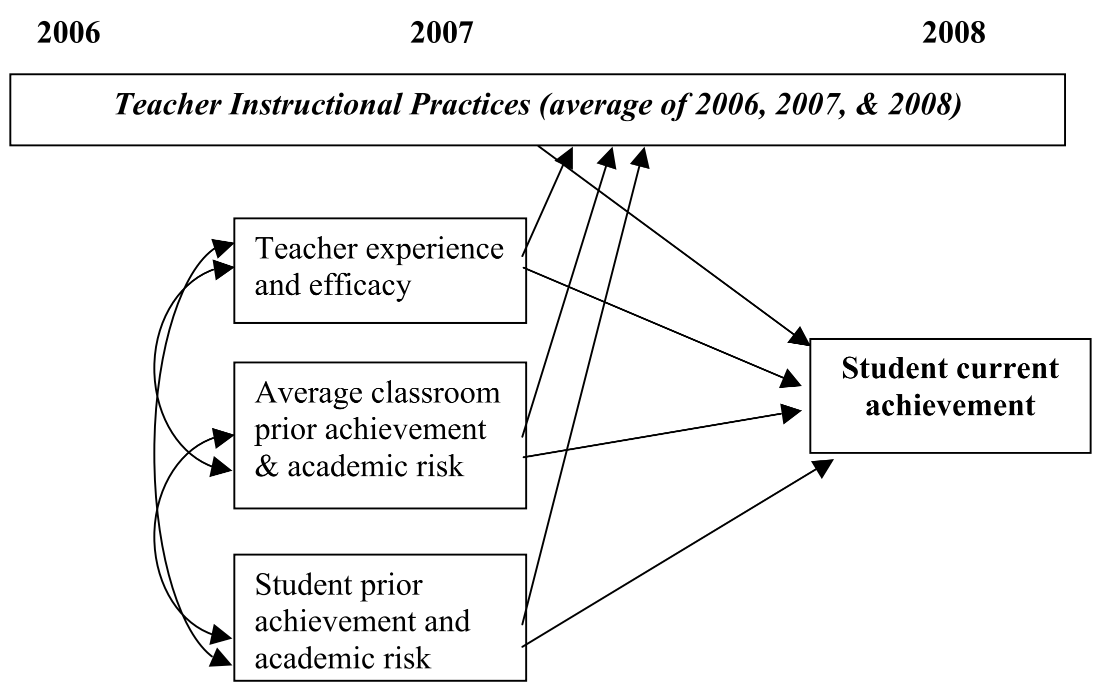
*Note.* *N*=112 teachers, *N*=2332-2345 students. No interaction of instruction on the slope of prior achievement and current achievement. Instruction is the 3-year mean of teacher-reported practices. Variables are not standardized and are grand centered.

*Figure 1.* Path analysis modeling how instructional practices are hypothesized to predict student achievement (level 1) while controlling for teacher characteristics (level 2), classroom characteristics (level 2), and student characteristics (level 1).

**Cohort 1:**




**Cohort 2:**



# Appendix A

## Items from the Instructional Practices Scale and their Rationale from the Literature

| Items about Students Struggling Academically | Support from the Literature |
|---|---|
| 1. I assess the level of challenge an academic task will provide this student. | Ysseldyke and Christenson (2002) recommended conducting a task analysis to examine how well components of the task will match the student's current skill level. Brophy and Good (1984) found that providing tasks allowing students to experience consistent success was highly correlated with student achievement. Consistent success involves creating a match between the task and the student's skill level. Creating this instructional match is particularly important for younger and less-skilled students to keep up classroom momentum and build foundational skills for more complex learning in later grades (Brophy & Good, 1984). |
| 2. I take the time to assess this student's prior knowledge and skills before teaching a lesson. | Glaser (1984) discussed how schema theory and evidence from studies of children's problem-solving suggest that "people continually try to understand and think about the new in terms of what they already know" (p. 100). Glaser suggested that activating prior knowledge before teaching new material helps the learner organize information efficiently. To make this connection, "one must understand an individual's current state of knowledge in a domain related to the subject matter to be learned" (Glaser, 1984, p. 101).<br><br>Peterson, Carpenter, and Fennema (1989) discussed how teachers' understanding of students' problem-solving knowledge may help them be more effective in mathematics instruction. Peterson et al. suggested that teachers knowledgeable about how children learn might be more successful with students "because they assessed their children's knowledge and skills at regular intervals... [and because] these teachers might have adapted some of their instruction to build on children's informal and formal knowledge of mathematics" (p. 559). |
| 3. I preview reading materials to ensure that this student will be able to read text with at least 93% level of accuracy. | In a repeated measures design, Gickling and Armstrong (1978) compared students' reading performance at three levels of task difficulty: frustrational (less than 90% known), instructional (93 to 97% known), and independent (more than 97% known). They found that students at instructional level had consistently high task comprehension, task completion and on-task behavior compared to the other conditions. Gickling and Armstrong concluded that teaching at instructional level creates optimal conditions for learning. |
| 4. I monitor the student's understanding or the content of a skill during activities and make adjustments accordingly. | Ysseldyke and Christenson (2002) recommended frequent informal assessment of student learning and modification of instruction and/or tasks as needed.<br>Peterson, Carpenter, and Fennema (1989) suggested that knowledgeable and effective teachers "assessed their children's knowledge and skills at regular intervals and...they planned instruction to take into account the wide range of knowledge and skills that existed in the classroom at any point in time" (p. 559). |

| | | |
|---|---|---|
| 5. | I make adjustments during lessons based on this student's understanding of the content or skill. | From a case study of two elementary school teachers' mathematics instruction, Peterson, Carpenter, and Fennema (1989) suggested that assessment used specifically to guide instruction for individual students is an important component of effective teaching. |
| 6. | I walk around to give immediate and specific feedback to this student while he or she is practicing a new skill. | Rosenshine (1983) found that effective teachers gave feedback and made corrections on student work, especially when students were learning new material. In their review of the literature on classroom assessment, Black & William (1998) concluded that "several studies show firm evidence that innovations designed to strengthen the frequent feedback that students receive about their learning yield substantial learning gains" (p. 7). |
| 7. | I prepare practice exercises for this student so that he or she knows at least 75% of the material before starting the task. | Rosenshine (1983) found that effective teachers ensured a student success rate of 80 percent or higher in initial learning tasks. Also, Brophy and Good (1984) found that instructional quality and pacing were highly correlated with student achievement. One of these variables included providing tasks that allowed students to experience consistent success. |
| 8. | For critical skills, I ensure that this student's practice is continued to the point of mastery. | Rosenshine (1983) found that effective teachers "provide for continued student practice so that students have a success rate of 90-100% and become rapid, confident, and firm" (p. 337). He discussed the notion of "overlearning" basic skills during the early instruction, a concept that is now generally referred to as mastery learning or learning to automaticity (Bloom, 1974). Mastery learning is achieved when teachers present new material in small chunks and then students practice these new skills repetitively until they become automatic. The purpose of mastery learning is to build automaticity of basic skills so that limited cognitive capacity can be used for higher-level cognitive processing (Rosenshine, 1983). When students have not mastered basic skills that are prerequisites to more complex tasks, teaching complex tasks is ineffective (Bloom, 1974). |
| 9. | I ensure that this student's engagement is high during independent work activities. | Brophy and Good (1984) found that instructional quality and pacing were highly correlated with student achievement. These variables included providing opportunities to learn; managing the classroom and allowing for student engaged time in activities; engaging in active teaching during which students were being instructed and supervised; and providing tasks that allowed students to experience consistent success. |
| 10. | I do more than the school system and curriculum requires to assess this student's performance on classroom tasks. | Shapiro (2004) outlined a model for assessing the academic skills of struggling students. The first step of the model involves evaluating the instructional environment using a variety of methods (e.g., interviews, rating scales, observation, review of student work). Ysseldyke and Christenson (2002) also recommended a thorough assessment of the instructional environment, including a review of student products and records. In an experimental study, Fuchs, Deno, and Mirkin (1984) found that special education teachers who were instructed to conduct frequent curriculum-based measurement and evaluation had students with higher reading achievement than teachers who conducted assessment in the conventional manner. |

| | | |
|---|---|---|
| 11. | I assess this student to pinpoint the most important instructional needs. | The second step in Shapiro's (2004) model of academic assessment is to determine the student's current instructional level in the classroom. Assessment of instructional needs is also a critical component in Ysseldyke and Christenson's (2002) nine-step Functional Assessment of Academic Behavior (FAAB). Ysseldyke and Christenson recommended gathering information from the teacher and parents on the student's instructional needs and collecting data on the instructional environment using a variety of methods (e.g., observation, record review, testing). |
| 12. | I set short-term goals for this student. | The third step in Shapiro's (2004) model of academic assessment is to make instructional modifications and set short-term goals. Locke and Latham (2002) reviewed more than three decades of research on goal-setting theory. According to Locke and Latham, goal-setting is effective through four mechanisms. First, having a goal directs behavior toward the goal and away from competing behaviors. Second, goals can energize people toward achieving them, especially when goals are set high but attainable. Third, having a goal to reach improves persistence. Finally, goals promote the activation of task-relevant knowledge and existing strategies (Locke & Latham, 2002). Goal-setting by students and teachers has been an effective way to increase achievement (Shapiro, 2004). |
| 13. | I collect data on this student to monitor progress toward short-term goals. | Shapiro's (2004) model also requires continuous assessment of progress toward short-term goals based on work by Fuchs, Deno, and Mirkin (1984) and others. Fuchs, Deno, and Mirkin conducted an experimental study in which special education teachers were randomly assigned either to specifically use repeated curriculum-based measurement and evaluation or to conduct assessment as usual. After 18 weeks of implementation, teachers in the repeated measurement and evaluation group had students with higher reading achievement and students who were more aware of their own goals and progress. In addition, these teachers were more realistic about and responsive to student progress and had improved instructional methods. |
| 14. | I flexibly group this student with other students by skill or objective. | Slavin (1987) reviewed the research on the effects of ability grouping on student achievement in elementary schools. Based on his best-evidence synthesis, Slavin recommended heterogeneous classes in which ability groups are used only for subjects (e.g., math, reading) that require greater homogeneity in student skill level for instruction. Slavin advised that "grouping plans should frequently reassess student placements and should be flexible enough to allow for easy reassignments after initial placement" (p. 37). |

| Items about Students Struggling Behaviorally | Support from the Literature |
|---|---|
| 15. I assess this student's academic skills in the subject areas in which the behaviors are occurring. | Gickling and Armstrong (1978) found a relationship between academic assignment difficulty and off-task behavior. Students working on tasks that were too hard (frustrational level) or too easy (independent level) engaged in high percentages of off-task behavior compared to students working at instructional level. Gickling and Armstrong suggested that the task given to students was an antecedent to their behavior though they cautioned against a definitive causal interpretation of this finding. Treptow, Burns, & McComas (2007) replicated Gickling and Armstrong's study and found similar, though less robust, findings. Specifically, Treptow et al. found that time on-task and comprehension were highest at instructional level compared to frustrational level. Roberts, Marshall, Nelson, & Albers (2001) examined the role of avoidance of academic tasks in classroom off-task behavior with general educations students. Roberts et al. manipulated academic task difficulty for three elementary school students and found a functional relationship between task difficulty and off-task behavior for all three students. They concluded that "assessing and remediating off-task classroom behaviors should include the simultaneous examination of both academic and behavior problems in a general education classroom" (p. 274). |
| 16. I define this student's behavior in specific and observable terms. | A key characteristic of the behavioral approach is its emphasis on observable behavior (Kazdin, 1984). Sulzer-Azaroff and Mayer (1991) further explained the importance of a precise definition of the behavior targeted for change; "this permits change to be more effectively programmed and monitored, and communication to be clearer" (p. 24). In the behavior approach, ambiguous descriptions of student problems should first be operationally defined so that specific goals can be set and progress toward these goals monitored (Sulzer-Azaroff & Mayer, 1991). |
| 17. I analyze what happens immediately before and after this student's behavior. | Functional behavioral assessments are used to identify, interpret, and modify student behavior by examining antecedents and consequences of the behavior. Teachers' role in functional assessment has typically been peripheral since data collection procedures are often complex and time-consuming. Symons, McDonald, & Wehby (1998) studied how special education teachers could collect observational data within their own classrooms. Teachers collecting their own data reported greater awareness of times when problem behaviors occurred and that using a scatter plot to chart these data helped them with student planning. Teachers also reported that data collection procedures were relatively easy; reliability checks with the researcher's observations showed a high degree of consistency (93%) between the teacher and researcher's observations. A study conducted in general education classrooms by Roberts, Marshall, Nelson, & Alberts (2001) implied that off-task behavior serves as an escape mechanism for students struggling with academic tasks. Roberts et al. suggested that analyzing task difficulty as an antecedent and avoidance of the task as a consequence of off-task behavior is critical to the assessment and intervention of problem behaviors in the classroom. |

| | | |
|---|---|---|
| 18. | I graph data about this student's increase in appropriate behaviors. | Poteet (1973) listed five reasons for teachers to observe, record, and measure behavior using charts or graphs: "1. Support your opinions about a situation or behavior which should be altered, 2. Reveal a pattern of behavior which might suggest ways to alter the behavior, 3. Reveal discrepancies between your perception of the situation and the situation as it exists, 4. Visually indicate if your program was successful, how successful it was, and when the significant change occurred, 5. Describe interaction for purposes of pinpointing target behavior for further programming" (p. 13). Poteet also recommended that teachers focus on increasing positive behaviors rather than decreasing negative behaviors: "the psychological effect of increasing a positively viewed behavior would seem closely related to developing a more positive self-concept of the student [and]... it would seem to lend support to a more positive classroom environment" (p. 9). |

# Appendix B

## Instructional Practices Scale 2008

*1 = Never, 2 = Rarely, 3 = Sometimes , 4 = Often, 5 = Always*

*1. I assess the level of challenge an academic task will provide this student.

*2. I take the time to assess this student's prior knowledge and skills before teaching a lesson.

*3. I preview reading materials to ensure that this student will be able to read text with at least 93% level of accuracy.

*4. I monitor the student's understanding or the content of a skill during activities and make adjustments accordingly.

*5. I make adjustments during lessons based on this student's understanding of the content or skill.

*6. I walk around to give immediate and specific feedback to this student while he or she is practicing a new skill.

*7. I prepare practice exercises for this student so that he or she knows at least 75% of the material before starting the task.

*8. For critical skills, I ensure that this student's practice is continued to the point of mastery.

*9. I ensure that this student's engagement is high during independent work activities.

*10. I do more than the school system and curriculum requires to assess this student's performance on classroom tasks.

*11. I assess this student to pinpoint the most important instructional needs.

*12. I set short-term goals for this student.

*13. I collect data on this student to monitor progress toward short-term goals.

*14. I flexibly group this student with other students by skill or objective.

**15. I assess this student's academic skills in the subject areas in which the behaviors are occurring.

**16. I define this student's behavior in specific and observable terms.

**17. I analyze what happens immediately before and after this student’s behavior.

**18. I graph data about this student’s increase in appropriate behaviors.

---

* Items about a student struggling academically

** Items about a student struggling behaviorally

# Appendix C

## Descriptive Statistics by Grade Level

*Descriptive Statistics for Third-Grade Student- and Classroom-Level Variables*

| Variables | 2007 | | | | 2008 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| Student-Level | | | | | | | | |
| Reading grade | 3.06 | 0.76 | 0.00 | 4.00 | 3.08 | 0.76 | 0.00 | 4.00 |
| Prior reading grade | 3.19 | 0.71 | 0.00 | 4.00 | 3.17 | 0.68 | 0.00 | 4.00 |
| Math grade | 3.12 | 0.77 | 0.00 | 4.00 | 3.14 | 0.76 | 0.00 | 4.00 |
| Prior math grade | 3.20 | 0.70 | 0.00 | 4.00 | 3.10 | 0.64 | 0.00 | 4.00 |
| Academic risk | 1.05 | 1.08 | 0.00 | 3.00 | 1.06 | 1.09 | 0.00 | 3.00 |
| Classroom-Level | | | | | | | | |
| Instructional Practices | 3.83 | 0.47 | 2.31 | 4.97 | 3.82 | 0.44 | 2.56 | 4.98 |
| Class prior reading grade | 3.17 | 0.29 | 2.10 | 3.82 | 3.16 | 0.27 | 2.43 | 3.87 |
| Class prior math grade | 3.18 | 0.29 | 2.20 | 3.74 | 3.09 | 0.24 | 2.29 | 3.64 |
| Class academic risk | 1.10 | 0.66 | 0.00 | 2.67 | 1.12 | 0.67 | 0.05 | 2.59 |
| Teaching experience | 3.01 | 1.34 | 1.00 | 5.00 | 3.03 | 1.15 | 1.00 | 5.00 |
| Teaching efficacy | 4.05 | 0.49 | 2.95 | 5.00 | 3.88 | 0.50 | 2.75 | 5.00 |

*Note.* Grades were final marking period report card grades. Student academic risk was the sum of 3 dichotomous variables: ethnic minority status (0=White or Asian, 1=Other ethnicity), eligibility for free or reduced meals (0=Not eligible, 1=Eligible), and English as a second language status (0=English is not a second language, 1=English is a second language). Instructional Practices was the average of the teacher's composite score on the Instructional Practices Scale from 2006 and 2007 for 2007 outcome year and from 2006, 2007, and 2008 for the 2008 outcome year (1 = infrequent use of practices, 5 = frequent use of practices). Teaching experience was self-reported years of experience (1 = 1 year or less, 2 = 2 to 5 years, 3 = 6 to 10 years , 4 = 11 to 20 years, 5 = More than 20 years). Teaching efficacy was the teacher's composite score on the Efficacy scale from the prior year (l = low efficacy, 5 = high efficacy).

*Descriptive Statistics for Fourth-Grade Student- and Classroom-Level Variables*

| Variables | 2007 | | | | 2008 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| Student-Level | | | | | | | | |
| Reading SOL score | 484 | 72.7 | 240 | 600 | 494 | 71.9 | 228 | 600 |
| Prior reading SOL score | 468 | 72.6 | 231 | 600 | 466 | 80.6 | 207 | 600 |
| Math SOL score | 475 | 79.8 | 241 | 600 | 486 | 74.6 | 204 | 600 |
| Prior math SOL score | 501 | 73 | 257 | 600 | 497 | 73.8 | 255 | 600 |
| Reading grade | 3.06 | 0.74 | 0.00 | 4.00 | 3.13 | 0.73 | 0.00 | 4.00 |
| Prior reading grade | 3.07 | 0.75 | 0.00 | 4.00 | 3.09 | 0.75 | 0.00 | 4.00 |
| Math grade | 3.08 | 0.77 | 0.00 | 4.00 | 3.11 | 0.77 | 0.00 | 4.00 |
| Prior math grade | 3.14 | 0.76 | 0.00 | 4.00 | 3.15 | 0.76 | 0.00 | 4.00 |
| Academic risk | 1.05 | 1.08 | 0.00 | 3.00 | 1.02 | 1.06 | 0.00 | 3.00 |
| Classroom-Level | | | | | | | | |
| Instructional practices | 3.81 | 0.51 | 2.60 | 4.90 | 3.86 | 0.48 | 2.72 | 4.89 |
| Class prior reading SOL score | 466 | 28.7 | 376 | 540 | 463 | 34.4 | 365 | 539 |
| Class prior math SOL score | 501 | 30.7 | 401 | 556 | 495 | 30.1 | 412 | 565 |
| Class prior reading grade | 3.05 | 0.33 | 2.24 | 3.72 | 3.07 | 0.32 | 1.98 | 3.81 |
| Class prior math grade | 3.12 | 0.33 | 2.26 | 3.70 | 3.12 | 0.32 | 2.17 | 3.67 |
| Class academic risk | 1.07 | 0.66 | 0.04 | 2.58 | 1.04 | 0.63 | 0.00 | 2.24 |
| Teaching experience | 3.15 | 1.32 | 1.00 | 5.00 | 3.20 | 1.20 | 1.00 | 5.00 |
| Teaching efficacy | 4.11 | 0.43 | 2.95 | 5.00 | 3.89 | 0.46 | 2.81 | 5.00 |

*Note.* SOL score is the scaled score earned on the Standards of Learning, the statewide standardized test administered in Virginia public schools. Grades were final marking period report card grades. Student academic risk was the sum of 3 dichotomous variables: ethnic minority status (0=White or Asian, 1=Other ethnicity), eligibility for free or reduced meals (0=Not eligible, 1=Eligible), and English as a second language status (0=English is not a second language, 1=English is a second language). Instructional Practices was the average of the teacher's composite score on the Instructional Practices Scale from 2006 and 2007 for 2007 outcome year and from 2006, 2007, and 2008 for the 2008 outcome year (1 = infrequent use of practices, 5 = frequent use of practices). Teaching experience was self-reported years of experience (1 = 1 year or less, 2 = 2 to 5 years, 3 = 6 to 10 years , 4 = 11 to 20 years, 5 = More than 20 years). Teaching efficacy was the teacher's composite score on the Efficacy scale from the prior year (l = low efficacy, 5 = high efficacy).

# Appendix D

## Correlation Tables between All Variables at Level 1 by Year

*Correlations between all variables at Level 1 for 2007.*

| Variables | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Instructional Practices | - | | | | | | | | | | | |
| 2. Reading test score | -0.01 | - | | | | | | | | | | |
| 3. Math test score | **-0.03** | **0.64** | - | | | | | | | | | |
| 4. Reading grade | **0.03** | **0.54** | **0.68** | - | | | | | | | | |
| 5. Math grade | **0.03** | **0.44** | **0.55** | **0.71** | - | | | | | | | |
| 6. Reading test score (prior year) | -0.02 | **0.65** | **0.59** | **0.50** | **0.53** | - | | | | | | |
| 7. Math test score (prior year) | 0.02 | **0.62** | **0.68** | **0.64** | **0.52** | **0.61** | - | | | | | |
| 8. Reading grade (prior year) | 0.00 | **0.56** | **0.55** | **0.55** | **0.57** | **0.55** | **0.53** | - | | | | |
| 9. Math grade (prior year) | -0.01 | **0.54** | **0.63** | **0.63** | **0.54** | **0.52** | **0.65** | **0.68** | - | | | |
| 10. Academic risk | 0.01 | **-0.40** | **-0.37** | **-0.35** | **-0.37** | **-0.31** | **-0.37** | **-0.37** | **-0.38** | - | | |
| 11. Teacher efficacy (prior year) | **0.54** | 0.02 | -0.01 | **0.04** | **0.04** | -0.01 | **0.03** | **0.02** | 0.00 | **-0.04** | - | |
| 12. Teacher education | **0.10** | **0.03** | -0.01 | **0.03** | **0.06** | 0.01 | **0.04** | **0.03** | 0.02 | **-0.05** | **0.10** | - |
| 13. Teacher experience | **0.11** | **0.05** | 0.02 | -0.01 | -0.01 | **0.04** | **0.07** | 0.01 | 0.02 | **0.07** | **0.17** | **0.27** |

*Note.* Bolded entries indicate $p<.05$. Variables from 2007 unless otherwise noted. Instructional Practices is composite score on the scale averaged across 2006 and 2007.

*Correlations between all variables at Level 1 for 2008.*

| Variables | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. | 12. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Instructional Practices | - | | | | | | | | | | | |
| 2. Reading test score | 0.02 | - | | | | | | | | | | |
| 3. Math test score | **-0.03** | **0.59** | - | | | | | | | | | |
| 4. Reading grade | 0.02 | **0.55** | **0.53** | - | | | | | | | | |
| 5. Math grade | 0.01 | **0.52** | **0.65** | **0.71** | - | | | | | | | |
| 6. Reading test score (prior year) | **-0.04** | **0.66** | **0.58** | **0.52** | **0.50** | - | | | | | | |
| 7. Math test score (prior year) | 0.00 | **0.59** | **0.68** | **0.53** | **0.64** | **0.63** | - | | | | | |
| 8. Reading grade (prior year) | -0.01 | **0.56** | **0.54** | **0.60** | **0.56** | **0.58** | **0.56** | - | | | | |
| 9. Math grade (prior year) | **-0.02** | **0.52** | **0.64** | **0.55** | **0.63** | **0.55** | **0.68** | **0.68** | - | | | |
| 10. Academic risk | **0.02** | **-0.36** | **-0.32** | **-0.35** | **-0.33** | **-0.41** | **-0.37** | **-0.37** | -0.36 | - | | |
| 11. Teacher efficacy (prior year) | **0.54** | **-0.03** | 0.01 | -0.01 | 0.00 | -0.02 | -0.03 | -0.03 | -0.01 | **0.03** | - | |
| 12. Teacher education | **0.02** | **0.03** | 0.01 | -0.02 | 0.00 | 0.02 | 0.02 | 0.01 | **0.02** | **-0.02** | **0.06** | - |
| 13. Teacher experience | **0.19** | **0.05** | 0.01 | -0.01 | **-0.03** | **0.03** | **0.04** | 0.00 | 0.01 | **-0.05** | **0.11** | **0.19** |

*Note.* Bolded entries indicate $p<.05$. Variables from 2008 unless otherwise noted. Instructional Practices is composite score on the scale averaged across 2006, 2007, and 2008.

## References

Berger, J., & Kaiser, L. (2008, August). *Assessment of teacher classroom instructional practices.* Poster presented at the annual meeting of the American Psychological Association, Boston, MA.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5,* 7-75.

Bloom, B. (1974). Time and learning. *American Psychologist, 29,* 682-688.

Brophy, J. (1983). Classroom organization and management. *The Elementary School Journal, 83,* 264-285.

Brophy, J. & Good, T. (1984). *Teacher behavior and student achievement.* The Institute for Research on Teaching: East Lansing, MI.

Bruckman, K. (2009). *Academic ability, problem behavior, and teacher tolerance on student report card grades.* Unpublished master's thesis, University of Maryland, College Park.

Camburn, E., & Barnes, C. (2004). Assessing the validity of a language arts instruction log triangulation. *The Elementary School Journal, 105,* 49-73.

Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., and Deke, J. (2009). *An evaluation of teachers trained through different routes to certification, final report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis, 29,* 262-295.

D'Agostino, J. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievement. *School Effectiveness and School Improvement, 11*, 197-235.

Darling-Hammond, Linda. (1990). Teaching and knowledge: Policy issues posed by alternate certification for teachers. *Peabody Journal of Education, 67,* 123-154.

Darling-Hammond, Linda. (2000). How teacher education matters. *Journal of Teacher Education, 51,* 166-173.

Desimone, L., & Le Floch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis, 26,* 1-22.

Education Week. (2004, September). *Accountability*. Retrieved from www.edweek.org.

Fuchs, L., Deno, S., & Mirkin, P. (1984). The effects of frequent curriculum-based measures and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21,* 449-460.

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

Gickling, E., & Armstrong, D. (1978). Levels of instructional difficulty as related to on task behavior, task completion, and comprehension. *Journal of Learning Disabilities, 11,* 559-566.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

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

Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66,* 361-396.

Guarino, C., Hamilton, L., Lockwood, J., Rathbun, A., & Hausken, E. (2006). Teacher qualifications, instructional practices, and reading and mathematics gains of kindergarteners (NCES 2006-031). U.S. Department of Education. Washington, DC: National Center of Education Statistics.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58,* 6-11.

National Center for Education Statistics. (2006). *State non-fiscal survey of public elementary/secondary Education*. Retrieved February 5, 2009 from http://nces.ed.gov/programs/digest/d07/tables/dt07_062.asp?referrer=list

Kaiser, L. (2007). *The effect of Instructional Consultation Teams on teachers' reported instructional practices*. Unpublished doctoral dissertation, University of Maryland, College Park. Retrieved June 4, 2008 from Digital Repository at the University of Maryland from http://hdl.handle.net/1903/7701

Kaiser, L., & Rosenfield, S. (2006, August). *Measurement of instructional practices: Survey design and psychometric properties*. Poster session presented at the Annual American Psychological Association Convention, New Orleans, LA.

Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R., DeGroot, E., Gilbert, M.C., & Musu, L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42(3),* 139-151.

Kazdin, A. (1984). *Behavior modification in applied settings.* The Dorsey Press: Chicago, IL.

Lee, V. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35,* 125-141.

LeTendre, G. (2009). Teacher Quality: National Trends in Teacher Policy. *Child Research Net.* Accessed on May 17, 2010 at www.childresearch.net.

Locke, E., & Latham, G. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57,* 705-717.

Mayer, D. (1999). Measuring instructional practices: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*, 29-45.

Mullens, J., & Kasprzyk, D. (1996). *Using qualitative methods to validate quantitative survey instruments: Proceedings of the section on survey research methods.* Paper presented at the annual meeting of the American Statistical Association, Alexandria, VA.

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

Osborne, J. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation, 7*. Retrieved April 5, 2004 from http://PAREonline.net/getvn.asp?v=7&n=1.

Peterson, P., Carpenter, T., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *Journal of Educational Psychology, 81,* 558-569.

Peverly, S. (2009). Beyond the monitoring of students' progress in classrooms: The assessment of students, curricula, and teachers. In S. Rosenfield& V. Berninger (eds.). Implementing evidence-based academic interventions in school settings (pp. 575-600). New York, NY: Oxford University Press.

Pianta, R., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System Manual, K-3*. Baltimore, Maryland: Paul H. Brookes Publishing Co., Inc.

Poteet, J. (1973). *Behavior modification: A practical guide for teachers.* Burgess Publishing Company: Minneapolis, MN.

Prince William County Public Schools. (2004). *Grading, Elementary School* (Regulation 661-1).

Ravitch, D. (2002, March). *A brief history of teacher professionalism.* Speech delivered at the White House Conference on Preparing Tomorrow's Teachers, Washington, D.C.

Roberts, M., Marshall, J., Nelson, J., & Albers, C. (2001). Curriculum-based assessment procedures embedded within functional behavioral assessments: Identifying escape-motivated behaviors in a general education classroom. *School Psychology Review, 30,* 264-277.

Rosenfield, S., & Gravois, T. (1996). *Instructional Consultation Teams*. New York, NY: Guilford.

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

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83,* 335-351.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal, 105,* 75-101.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*, 1525-1567.

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

Scheerens, J. & Bosker, R. (1997). *The Foundations of Educational Effectiveness*. New York, NY: Pergamon.

Shapiro, E. (2004). *Academic skills problems: Direct assessment and intervention*. New York, NY: The Guildford Press.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 16,* 4-14.

Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57,* 293-336.

Sulzer-Azaroff, B., & Mayer, G. (1991). *Behavioral analysis for lasting change.* Wadsworth/Thomson Learning: Belmont, CA.

Symons, F., McDonald, L., & Wehby, J. (1998). Functional assessment and teacher collected data. *Education & Treatment of Children, 21,* 135-150.

Valli, L., Croninger, R., & Walters, K. (2007). Who (else) is the teacher? Cautionary notes on teacher accountability systems. *American Journal of Education, 113,* 635-662.

Virginia Department of Education (n.d.). Standards of Learning currently in effect for Virginia public schools. Retrieved February 5, 2009, from http://www.doe.virginia.gov/go/Sols/home.shtml

Veldman, D., & Brophy, J. (1974). Measuring teacher effects on pupil achievement. *Journal of Educational Psychology, 66,* 319-324.

Vu, P. (2009). *The influences of classroom characteristics and teacher-student relations on student academic achievement*. Unpublished master's thesis, University of Maryland, College Park.

Vu, P., Bruckman, K., Rosenfield, S., Nelson, D., Gravois, T., Berger, J., Kaiser, L., Vaganek, M., & Gottfredson, G. (2008). *The effects of Instructional Consultation Teams on teacher beliefs and instructional practices.* Manuscript submitted for publication.

Whitehurst, Grover. (2002, March). *Research on teacher preparation and professional development.* Speech delivered at the White House Conference on Preparing Tomorrow's Teachers, Washington, D.C.

Ysseldyke, J., & Christenson, S. (1993). *The Instructional Environment System-II: A system to identify a student's instructional needs.* Longmont, Colorado: Sopris West.

Ysseldyke, J., & Christenson, S. (2002). *Functional Assessment of Academic*

*Behavior (FAAB): Creating successful learning environments.* Longmont, Colorado: Sopris West.