

Journal of Teacher Education

<http://jte.sagepub.com/>

What Makes Good Teachers Good? A Cross-Case Analysis of the Connection Between Teacher Effectiveness and Student Achievement

James H. Stronge, Thomas J. Ward and Leslie W. Grant

Journal of Teacher Education 2011 62: 339

DOI: 10.1177/0022487111404241

The online version of this article can be found at:

<http://jte.sagepub.com/content/62/4/339>

Published by:



<http://www.sagepublications.com>

On behalf of:



[American Association of Colleges for Teacher Education \(AACTE\)](http://www.aacte.org)

Additional services and information for *Journal of Teacher Education* can be found at:

Email Alerts: <http://jte.sagepub.com/cgi/alerts>

Subscriptions: <http://jte.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>


Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://jte.sagepub.com/content/62/4/339.refs.html>

>> [Version of Record](#) - Sep 21, 2011

[What is This?](#)

What Makes Good Teachers Good? A Cross-Case Analysis of the Connection Between Teacher Effectiveness and Student Achievement

Journal of Teacher Education
62(4) 339–355
© 2011 American Association of
Colleges for Teacher Education
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/0022487111404241
<http://jte.sagepub.com>


James H. Stronge¹, Thomas J. Ward¹, and Leslie W. Grant¹

Abstract

This study examined classroom practices of effective versus less effective teachers (based on student achievement gain scores in reading and mathematics). In Phase I of the study, hierarchical linear modeling was used to assess the teacher effectiveness of 307 fifth-grade teachers in terms of student learning gains. In Phase II, 32 teachers (17 top quartile and 15 bottom quartile) participated in an in-depth cross-case analysis of their instructional and classroom management practices. Classroom observation findings (Phase II) were compared with teacher effectiveness data (Phase I) to determine the impact of selected teacher behaviors on the teachers' overall effectiveness drawn from a single year of value-added data.

Keywords

teacher effectiveness, teacher quality, value-added assessment, classroom management, instructional practices, student achievement, student learning

The question of what constitutes effective teaching has been researched for decades. However, changes in assessment strategies, the availability of newer statistical methodologies, and access to large databases of student achievement information, as well as the ability to manipulate these data, merit a careful review of how effective teachers are identified and how their work is examined. A better understanding of what constitutes teacher effectiveness has significant implications for decision making regarding the preparation, recruitment, compensation, inservice professional development, and evaluation of teachers. If an administrator seeks to hire effective or, at least, promising teachers, for example, she or he needs to understand what characterizes them. Recently, educators have begun to emphasize the importance of linking teacher effectiveness to various aspects of teacher education and district or school personnel administration, including

1. identifying the knowledge and skills preservice teachers need,
2. recruiting and inducting potentially effective teachers,
3. designing and implementing professional development,
4. conducting valid and credible evaluations of teachers, and
5. dismissing ineffective teachers while retaining effective ones (Darling-Hammond & Bransford, 2005;

Hanushek, 2008; National Academy of Education, 2008; Odden, 2004).

This type of alignment is receiving increasing attention as an important means for providing quality education to all students and improving school performance.

This study examined the measurable impact that individual teachers have on student achievement. Using residual student learning gains, the study investigated how effective teachers (i.e., teachers whose students experience high academic growth) differ from less effective teachers (teachers whose students experience less academic growth) in a single year. Classroom differences between effective and less effective teachers were examined in terms of both their teaching behaviors and their students' classroom behaviors. The purposes of this study were, first, to examine the impact that teachers had on student learning and, then, to examine the instructional practices and behaviors of effective teachers. In an effort to address these essential questions, we engaged in a two-phase study.

¹College of William and Mary, Williamsburg, VA, USA

Corresponding Author:

James H. Stronge, College of William and Mary, School of Education,
PO Box 8795, Williamsburg, VA 23187
Email: jhstro@wm.edu

Table 1. Teacher Effectiveness Dimensions

Dimensions of teacher effectiveness	Representative research base
Instructional delivery	
Instructional differentiation	Langer, 2001; Molnar et al., 1999; Randall, Sekulski, & Silberg, 2003; Weiss, Pasley, Smith, Banilower, & Heck, 2003
Instructional focus on learning	Darling-Hammond, 2000; Johnson, 1997; Molnar et al., 1999; National Center for Education Statistics, 1997; Wenglinsky, 2004; Zahorik, Halbach, Ehrle, & Molnar, 2003
Instructional clarity	Allington, 2002; Peart & Campbell, 1999; Zahorik et al., 2003
Instructional complexity	Molnar et al., 1999; Pressley, Wharton-McDonald, Allington, Block, & Morrow, 1998; Sternberg, 2003; Wenglinsky, 2000
Expectations for student learning	Peart & Campbell, 1999; Palardy & Rumberger, 2008; Wentzel, 2002
Use of technology	Cradler, McNabb, Freeman, & Burchett, 2002; Schacter, 1999; Wenglinsky, 1998
Questioning	Allington, 2002; Cawelti, 2004; Stronge, Ward, Tucker, & Hindman, 2008
Student assessment	
Assessment for understanding	P. Black, Harrison, Lee, Marshall, & William, 2004; Cotton, 2000; Yesseldyke & Bolt, 2007
Feedback	P. Black et al., 2004; Chappuis & Stiggins, 2002; Hattie & Timperley, 2007; Matsumura, Patthey-Chavez, Valeds, & Garnier, 2002
Learning environment	
Classroom management	Johnson, 1997; Taylor, Pearson, Clark, & Walpole, 2000; Wang, Haertel, & Walberg, 1993
Classroom organization	Marzano, Pickering, & McTighe, 1993; Zahorik et al., 2003
Behavioral expectations	Good & Brophy, 1997; Marzano, 2003; Palardy & Rumberger, 2008
Personal qualities	
Caring, positive relationships with students	Adams & Singh, 1998; Collinson, Killeavy, & Stephenson, 1999; National Association of Secondary School Principals, 1997
Fairness and respect	Agne, 1992; McBer, 2000
Encouragement of responsibility	Stronge, McColsky, Ward, & Tucker, 2005
Enthusiasm	Bain & Jacobs, 1990; Rowan, Chiang, & Miller, 1997

Phase I: To what degree do teachers have a positive, measurable effect on student achievement?

Phase II: How do instructional practices and behaviors differ between effective and less effective teachers based on student learning gains?

Background

Effectiveness is an elusive concept to define when we consider the complex task of teaching and the multitude of contexts in which teachers work. In discussing teacher preparation and the qualities of effective teachers, Lewis et al. (1999) aptly noted that “teacher quality is a complex phenomenon, and there is little consensus on what it is or how to measure it” (para. 3). In fact, there is considerable debate as to whether we should judge teacher effectiveness based on teacher inputs (e.g., qualifications), the teaching process (e.g., instructional practices), the product of teaching (e.g., effects on student learning), or a composite of these elements.

This study focused first on identifying those teachers who were successful in the product of teaching, namely, student achievement, and then it focused on an examination of the teaching process. Four dimensions that characterize teacher effectiveness synthesized from a meta-review of extant research and literature (Stronge, 2002, 2007) were used as

the conceptual framework for the study. The first two dimensions related to effective teaching practice, including instructional effectiveness and the use of assessment for student learning. The next two dimensions related to a positive learning environment, including the classroom environment itself and the personal qualities of the teacher.

Each of these dimensions focuses on a fundamental aspect of the teacher’s professional qualifications or responsibilities and is summarized below. It is important to note that the four primary dimensions and the subcomponents of each are not mutually exclusive. For example, *instructional clarity* is a dimension of *instructional delivery* but also can be viewed as a consequence of *learning environment*. This overlapping nature of teaching will always hold true when we attempt to deconstruct it into discrete categories.

Table 1 gives an overview of the dimensions of teacher effectiveness, including the representative research base of each.

Instructional Delivery

Instructional delivery includes the myriad teacher responsibilities that provide the connection between the curriculum and the student. Research on aspects of instructional delivery that lead to increased student learning can be examined

in terms of the following areas: instructional differentiation, focus on learning, instructional clarity, instructional complexity, expectations for student learning, the use of technology, and the use of questioning.

Instructional Differentiation. Studies that have examined the instructional practices of effective teachers have found that they use direct instruction (Pressley, Wharton-McDonald, Allington, Block, & Morrow, 1998), individualized instruction (Zahorik, Halbach, Ehrle, & Molnar, 2003), discovery methods, and hands-on learning (Wenglinsky, 2000), among other practices. Although these studies examined the efficacy of specific approaches to instructional delivery, researchers have found that effective teachers are adept at using a myriad of instructional strategies (Covino & Iwanicki, 1996; Langer, 2001; Molnar et al., 1999).

Instructional Focus on Learning. Effective teachers focus students on the central reason for schools to exist—learning. Although teachers stress both academic and personal learning goals with students, they focus on providing students with basic skills and critical thinking skills to be successful (Zahorik et al., 2003). In addition, effective teachers maximize instructional time (Taylor, Pearson, Clark, & Walpole, 1999) and spend more time on teaching than on classroom management (Molnar et al., 1999).

Instructional Clarity. Instructional clarity is related to a teacher's ability both to explain content clearly to students and to provide clear directions to students throughout instruction (Good & McCaslin, 1992; Peart & Campbell, 1999; Stronge, 2007). Indeed, one solid link between teacher skills and student achievement that has been supported by research over the past four decades is teachers' verbal ability, as measured by teacher performance on standardized assessments (Coleman et al., 1966; Strauss & Sawyer, 1986; Wenglinsky, 2000).

Instructional Complexity. Effective teachers recognize the complexities of the subject matter and focus on meaningful conceptualization of knowledge rather than on isolated facts, particularly in mathematics and reading (Mason, Schroeter, Combs, & Washington, 1992; Pressley et al., 1998; Wenglinsky, 2004). One study that examined elementary and middle school students' performance on academic achievement tests found that students who received instruction that emphasized both critical thinking and memorization performed better than those in classrooms where instruction emphasized critical thinking or memorization (Sternberg, 2003).

Expectations for Student Learning. The ability to communicate high expectations to students is directly associated with effective teaching (Stronge, 2007). Indeed, one indicator of student dropout rates is related to the teachers' expectations (Wahlage & Rutter, 1986). A study of middle school students found that teacher expectation was a significant predictor of student achievement (Wentzel, 2002). High expectations are communicated through the planning process in which teachers focus on complex as well as basic skills (Knapp, Shields, & Turnbull, 1992) and by expecting students to

complete their work (Bernard, 2003). A study of first-grade students found that reading achievement was lower for students whose teachers had low expectations (Palardy & Rumberger, 2008).

Use of Technology. The literature regarding the use of technology supports its inclusion as an effective practice in teaching. Schacter (1999) found that students made greater achievement gains when they had access to technology. Technology has a greater impact on student achievement when it is used to teach higher order thinking skills (Wenglinsky, 1998), and it has been associated with encouraging critical thinking in students (Cradler, McNabb, Freeman, & Burchett, 2002).

Student Assessment

Assessment is an ongoing process that occurs before, during, and after instruction is delivered. Effective teachers monitor student learning through the use of a variety of informal and formal assessments and offer meaningful feedback to students (Cotton, 2000; Good & Brophy, 1997; Hattie & Timperley, 2007; Peart & Campbell, 1999). Indeed, the well-designed use of formative assessment yields gains in student achievement equivalent to one or two grade levels (Assessment Reform Group, 1999), thus having a significant impact on student achievement (P. Black & Wiliam, 1998; Marzano, 2006). Effective teachers check for student understanding throughout the lesson and adjust instruction based on the feedback (Guskey, 1996).

Learning Environment

The importance of maintaining a positive and productive learning environment is noticeable when students are following routines and taking ownership of their learning (Covino & Iwanicki, 1996). Classroom management is based on respect, fairness, and trust, wherein a positive climate is cultivated and maintained (Tschannen-Moran, 2000). Effective teachers nurture a positive climate by setting and reinforcing clear expectations throughout the school year, but especially at its beginning (Cotton, 2000; Covino & Iwanicki, 1996; Emmer, Evertson, & Worsham, 2003). A productive and positive classroom is the result of the teacher's considering students' academic as well as social and personal needs.

Personal Qualities

One critical difference between more effective and less effective teachers is their affective skills (Emmer, Evertson, & Anderson, 1980). Teachers who convey that they care about students have higher levels of student achievement than teachers perceived by students as uncaring (Collinson, Killeavy, & Stephenson, 1999; Darling-Hammond, 2000; Hanushek, 1971; Wolk, 2002). These teachers establish

connections with students and are reflective practitioners dedicated to their students and to professional practice (R. Black & Howard-Jones, 2000; Stronge, 2007). In addition, more effective teachers encourage students to take responsibility for themselves (Stronge et al., 2005).

Purpose of the Study

If we are to understand how teachers' practices affect student learning, we must peer inside the black box of the classroom. As Rowan, Correnti, and Miller (2002) noted,

The time had come to move beyond variance decomposition models that estimate the random effects of schools and classrooms on student achievement. These analyses *treat the classroom as a black box* [italics added] and . . . do not tell us why some classrooms are more effective than others. (p. 1554)

Moreover, Goldhaber (2002) posed a question that epitomizes the purpose of this study: "What does the empirical evidence have to say about specific characteristics of teachers and their relationship to student achievement?" (p. 50).

This study focused on product (student achievement gain scores) first and on process (instructional practices of effective and less effective teachers) second. Then, we turned our attention to the relationship between the two. In some respects, we built on the tradition of process-product research (i.e., Berliner & Rosenshine, 1977; Good & Brophy, 1997) while accounting for contextual variables in teacher effectiveness. The advantage of this study over conventional studies employing value-added methods is its unique combination of the variance decomposition method with a more in-depth examination of the beliefs and practices of the high- and low-performing teachers. By comparing the findings from the observational phase of the study (Phase II) with the findings derived from the value-added assessment of teacher effectiveness (Phase I), our intent was to shed light on the elusive connection between teacher effects and teaching practices.¹ This study is based on one year's data and, thus, should be considered as one case study in which this critical question of teacher effectiveness-teacher behavior is explored.

Phase I: The Value-Added Impact of Teachers on Student Achievement

Phase I Sample Selection

Two years of student test scores in reading and math from 307 fifth-grade teachers from three public school districts in a state located in the southeastern United States were included in Phase I of the study. These three districts represented one large urban and suburban district (number of schools = 67) and two rural districts (number of schools = 43).

Although multiple years of student data would be desirable, the study was based upon data from a two-year set of lag

Table 2. Selected Characteristics of the Phase I Teacher Sample

Variable	
Mean years teaching	13.9
Percent female	90.6
Percent White	94.2
Percent with bachelor's degree only	48.7
Percent with master's degree	37.8
Percent with master's degree plus post-master's course work	13.5

data, providing for a preassessment and postassessment for each student. The primary reason for this restriction was the availability of data from the selected school districts, along with the districts' ability to extract those data. This limitation should be considered when interpreting the results presented later in the article.

To ensure that the students included in the fifth-grade data set could be properly tracked back to the teachers responsible for teaching them reading and mathematics, respectively, students were included only when there was a match between the classroom teacher and the teacher responsible for administering the end-of-year test. This matching process is similar to that used in other value-added studies (see, e.g., Rothstein, 2010).

The data provided by the three separate school districts were merged into a common data set. The final database contained the records of more than 4,600 fifth-grade students and 379 teachers. The data from all students were used in the student-level analyses, but achievement indices were calculated only for those teachers for whom there were data on 10 or more of their students and who had two years of data. Thus, the final number of teachers was reduced to 307 (as noted earlier). Selected characteristics of the teacher sample are presented in Table 2.

Phase I Method

The methodology for studying the relationship between teacher demographic characteristics and student achievement began with modeling fifth-grade students' math and reading achievement to obtain estimates of teacher effectiveness. These math and reading tests were designed to measure student performance on the grade-level competencies specified in the state's curriculum standards and, thus, were criterion-referenced assessments.

A regression-based methodology, hierarchical linear modeling (HLM), was used to estimate the growth for all students included in the sample in order to predict the expected achievement level for each child. In this HLM analysis, we used student-level variables as predictors of student performance at Stage 1 and classroom-level variables at Stage 2. The student-level variables included in the model were gender, ethnicity, free or reduced lunch status, English as a second language (ESL) programming, special education status, and

prior achievement as measured by the fourth-grade reading and mathematics scores. The classroom-level variables in the model were gender, ethnicity, free or reduced lunch status, eligible for special education services, English-language learner, and class size.

The equation below represents the student-level (Level 1) model:

$$Y_{ij} \text{ (math or reading scores)} = \beta_0_j + \beta_1_{j^*} \text{ (prior math achievement)} + \beta_2_{j^*} \text{ (prior reading achievement)} + \beta_3_{j^*} \text{ (Caucasian)} + \beta_4_{j^*} \text{ (free and reduced lunch)} + \beta_5_{j^*} \text{ (special education)} + \beta_6_{j^*} \text{ (English-language proficiency)} + e_{ij}$$

where β_0_j is the intercept of the outcome, whereas β_1_j , β_2_j , β_3_j , β_4_j , and β_5_j are the slopes or coefficients for each of the independent variables.

The classroom-level (or Level 2) model was

$$\begin{aligned} \beta_0_j &= \gamma_{00} + \gamma_{01^*} \text{ (class size)} + \gamma_{02^*} \text{ (percent minority)} + \gamma_{03^*} \text{ (percent male)} + \gamma_{04^*} \text{ (percent free and reduced lunch)} + \gamma_{05^*} \text{ (percent with identified disability)} + \gamma_{06^*} \text{ (percent English-language proficient)} + u_{0j} \\ \beta_1_j &= \gamma_{10} + \gamma_{11^*} \text{ (class size)} + \gamma_{12^*} \text{ (percent minority)} + \gamma_{13^*} \text{ (percent male)} + \gamma_{14^*} \text{ (percent free and reduced lunch)} + \gamma_{15^*} \text{ (percent with identified disability)} + u_{1j} \end{aligned}$$

Following the HLM analysis of the approximately 4,600 students' predicted and actual test scores on reading and math, estimates of teacher impact on achievement (referred to as the Teacher Achievement Indices, or TAI) were calculated by averaging all student residual gains for the 307 teachers. Similar value-added model (VAM) studies of student achievement have used averaging for the teacher (Bembry, Jordan, Gomez, Anderson, & Mendro, 1998; Jordan, Mendro, & Weerasinghe, 1997). To create the residuals, the estimated performance for each student from the model was compared to the student's actual fifth-grade performance. Finally, the TAI values were standardized (centered) on a *T* scale ($M = 50$, $SD = 10$) for ease of interpretation. The individual teachers were ranked on the TAI measures, and the listing was divided into quartiles to identify the teachers for analysis in Phase I and observation in Phase II of the study.

Phase I Results

Grand-mean centering² was used for the variables in this analysis—an approach recommended for this type of analysis (Goe, 2008). At the student level of analysis, special education status was a significant predictor for mathematics, whereas gender (favoring female students) was a significant contributor for reading. Prior mathematics and reading achievement were significant and, by far, the strongest predictors in both analyses. After controlling for other variables in the model, prior reading achievement accounted for 63% of the explained

Table 3. Hierarchical Linear Modeling Relationship to Dependent Measures (actual student achievement)

	Model relationship	Explained variance
Math	.87	.75
Reading	.81	.65

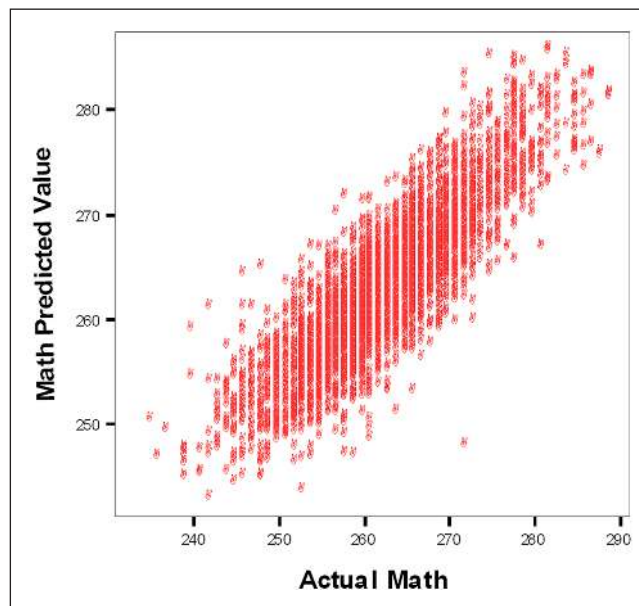


Figure 1. Scatterplot for fifth-grade student predicted versus actual mathematics achievement indices

variance in reading, and prior mathematics achievement accounted for 59% of the explained variance in math. It is interesting to note that free or reduced lunch status (i.e., socioeconomic status) had little explanatory power in the two analyses. None of the classroom-level measures in the HLM model were significant for mathematics; however, ESL students and ethnicity were significant for reading. Table 3 provides summary correlations depicting the relationships between the predictor and dependent variables. Figures 1 and 2 provide a graphical representation of the predicted and actual achievement scores of the approximately 4,600 fifth-grade students included in the analysis.

TAI distributions for reading and math were based on the mean residual student gain scores. The math TAIs ranged from 22 to 77 (Figure 3). The distribution had almost no skewness and only slight negative kurtosis. A test of normality indicated that math TAIs did not depart significantly from a normal distribution. The reading TAIs ranged from 13 to 78 (Figure 4). The distribution showed slight negative skewness and some positive kurtosis. A test of normality indicated that the reading TAIs did not depart significantly from a normal distribution.

Correlations were calculated between the reading and math TAIs and the identified teacher demographic variables.

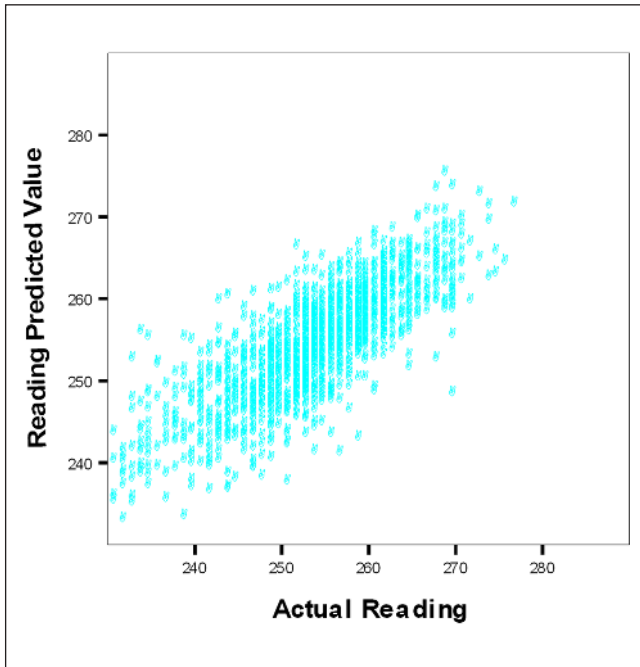


Figure 2. Scatterplot for fifth-grade student predicted versus actual reading achievement indices

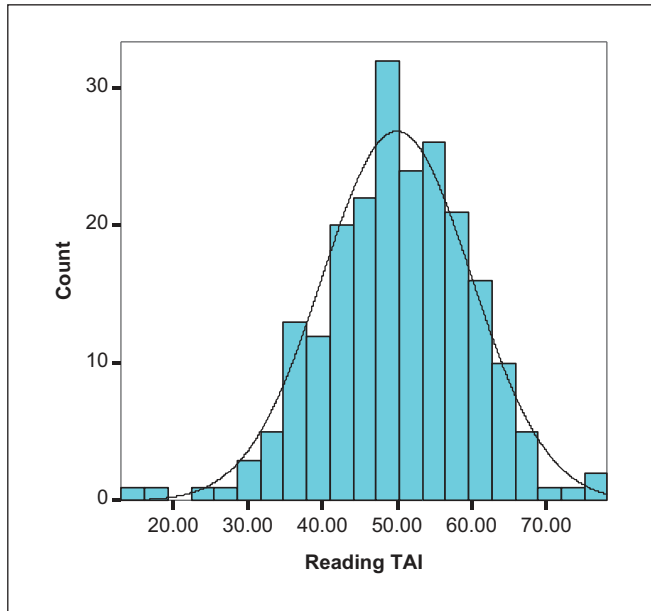


Figure 4. Teacher Effectiveness Indices (TAI) distribution for reading
N = 217.

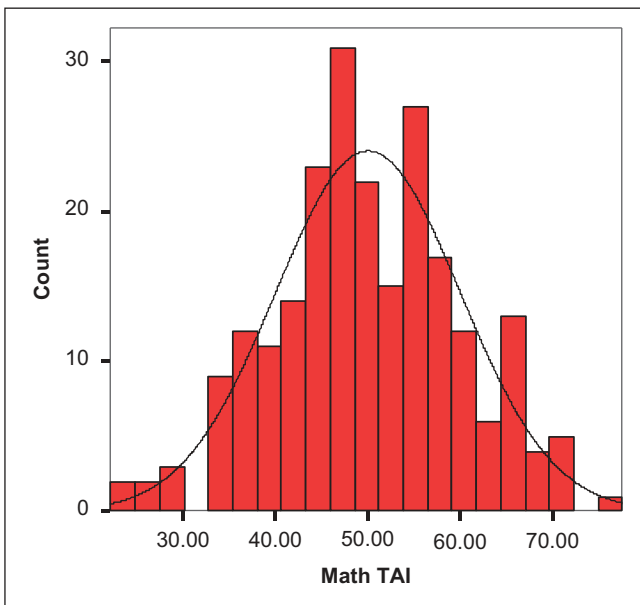


Figure 3. Teacher Effectiveness Indices (TAI) distribution for mathematics
N = 229.

Years of service, ethnicity, and pay grade were the variables used in this analysis. The correlations, reported in Table 4, indicate that there were no significant relationships among the teacher demographics and the TAIs. When we divided the sample by experience level and considered only those teachers with less experience (10 years or less), we still found no

Table 4. Correlations Between Teacher Demographics and Teacher Achievement Indices (TAI)

	Years of service	Ethnicity	Pay grade
TAI math	.10	.08	.05
TAI reading	.10	-.03	.05

For all correlations, $p > .05$.

significant relationships. In addition, when we checked for a potential curvilinear relationship with years of experience, again, none was evident.

In follow-up analyses, we considered the implications of having a top- versus bottom-quartile teacher in terms of student gains. In both reading and math, there were no statistically significant differences in student achievement levels on the end-of-course fourth-grade tests (which served as pretests in the study) at the beginning of the school year between the top- and bottom-quartile teachers' classes. When we considered end-of-course fifth-grade scores and calculated gains, the differences were striking. For reading, the difference in gains was 0.59 standard deviations in one year. In practical terms, students taught by bottom-quartile teachers could expect to score, on average, at the 21st percentile on the state's reading assessment, whereas students taught by the top-quartile teachers could expect to score at approximately the 54th percentile.³ This difference, more than 30 percentile points, can be attributed to the quality of teaching occurring in the classrooms during one academic year.

We found similar results for mathematics, with a difference in gain scores of 0.45 standard deviations. When translated

Table 5. Pre- and Posttest Data for Students' Reading and Mathematics Scores

	Reading percentile scores		Mathematics percentile scores	
	Pre	Post	Pre	Post
Students in bottom-quartile teachers' classes	43	21*	48	38*
Students in top-quartile teachers' classes	43	54*	51	70*

N = 1,984 (931 students in bottom-quartile classrooms, 1,053 in top-quartile classes).

**p* < .05, for posttest comparisons for both reading and mathematics

Table 6. Teachers Invited and Agreeing to Participate by District and Group

District	Number of top-quartile teachers invited	Number of top-quartile teachers who agreed to participate	Number of bottom-quartile teachers invited	Number of bottom-quartile teachers who agreed to participate
Urban/suburban 1	29	9	34	7
Rural 1	13	3	12	1
Rural 2	12	5	14	7
Total	54	17 (31%)	60	15 (25%)

N = 32.

into percentile scores, the students in the bottom-quartile teachers' classrooms scored, on average, at the 38th percentile; students in the top-quartile teachers' classrooms scored at the 70th percentile. This translates into more than a 30 percentile difference in achievement based on one year's teaching and learning experience (Table 5).

Phase II: Comparison of Teaching Practices Between Top- and Bottom-Quartile Teachers

Phase II of the study focused on case studies of selected teachers from Phase I to answer the following question: *How do teaching practices differ between effective and less effective teachers?* Effective teachers were defined as those with TAIs in the top quartile; less effective teachers were defined as those with TAIs in the bottom quartile.

Phase II Sample Selection: Recruitment of Classroom Teachers

In this phase of the study, we conducted a cross-case analysis with 17 top- and 15 bottom-quartile teachers as identified by an analysis of student achievement gain score residuals. Table 6 summarizes the selection of the teachers from the three participating school districts. The average number of students in each classroom was 21.1, with a standard deviation of 3.98.

Phase II Instrumentation

Teacher Beliefs. A teacher's sense of efficacy is based on a set of beliefs in his or her ability to make a difference in student learning, including the ability to reach difficult or unmotivated students, and was assessed in the study using the short form of the Teacher Sense of Efficacy Scale (TSES; Tschannen-Moran & Hoy, 2001). Reliabilities for the teacher efficacy subscales were .91 for Instructional Strategies, .90 for Classroom Management, and .87 for Student Engagement. Intercorrelations between the subscales of Instructional Strategies, Classroom Management, and Student Engagement were .60, .70, and .58, respectively (*p* < .001). Means for the three subscales ranged from 6.71 to 7.27. In a validation study by Tschannen-Moran and Hoy (2001) for the short form, the strongest correlations between the TSES and other measures are with scales that assess personal teaching efficacy.

Questioning Techniques Analysis Chart. This instrument was designed to categorize the types of questions asked by teachers and their students. In this study, all instructional questions asked by the teacher, orally and in writing, were recorded for a one-hour period during the language arts lesson. In addition, student-generated questions that were not procedural in nature were recorded. Questions were categorized based on low, intermediate, and high cognitive demand (Good & Brophy, 1997). Cognitive demand is the level of cognitive complexity (Anderson & Krathwol, 2001; Bloom, 1956). The observers also documented three examples of

each question type on the Questioning Techniques Analysis Chart and tallied the number of questions asked by teachers and students at each level. Percentages were calculated for total questions asked at each cognitive level. A guide for categorizing questions based on Bloom's taxonomy was provided as a reference for observers to ensure consistency in coding. Low cognitive demand included knowledge, intermediate cognitive demand included comprehension and application, and higher cognitive demand included analysis, evaluation, and synthesis.

Student Time-on-Task Chart. This instrument was designed to record student engagement in the teaching-learning process at regular five-minute intervals. In addition, comments regarding off-task behavior and teacher responses were recorded. This was a modified version of an instrument from an earlier study related to the efficacy of national board-certified teachers (Bond, Smith, Baker, & Hattie, 2000).

Teacher Effectiveness Summary Rating Form. This is a behaviorally anchored rating scale of dimensions of effective teaching as identified through prior studies and is based on Stronge's (2002, 2007) analysis of research on effective teaching. The instrument is designed to capture both the *types* of behaviors and the *degree* to which the participating classroom teachers exhibit those behaviors. For each classroom visit, two observers completed the Teacher Effectiveness Summary Rating Form using the four-point scale (Teacher Effectiveness Behavior Scale) to guide their judgments about teacher effectiveness and coding on each dimension (Stronge et al., 2008). After the observation, observers' individual ratings for each dimension were recorded along with their rationales for each. Once the two observers had completed all the instruments, they compared and discussed their respective ratings on the Teacher Effectiveness Summary Rating Form and reached consensus on the most accurate coding for each dimension in those instances in which their initial ratings differed.

Phase II Method

Selection and Training of Observers. Graduate students from a southeastern U.S. university and retired educators were recruited to serve as observers for this phase of the study. After completing an application and interview process, eligible candidates were invited to attend a one-day training session focusing on skills in conducting classroom observations using the instruments developed for the study. The training session included an overview of the study, training on the use of each protocol, and instructions on synthesizing the data for the overall rating of the observation. Each participant practiced using the various observation instruments while viewing videotapes of one reading teacher and one mathematics teacher instructing their classes. Finally, participants scored the videos using the Teacher Effectiveness Summary Rating Form. Scoring was completed individually and was followed by a large-group discussion to establish a

common understanding of the rubric. The training session culminated with a performance assessment that simulated the actual data collection process.

Five members of the research team used the same practice videotape to establish a target set of scores for assessing observers' performance with the rubric. Scores of potential observers were compared to the target scores for each dimension of the rubric. All participants who scored the videotaped performance of the teaching episode with an 80% or above agreement with the target scores were selected to be observers. Those with between 70% and 79% agreement were invited to return for additional training and assessment in an effort to achieve a minimum of 80% agreement. Those with less than 70% agreement were not selected to serve as observers.

Observation Procedure. Two observers visited each participating teacher's classroom for three hours, which typically encompassed both language arts and math instruction. Neither the observers nor the teachers selected for observation knew which teachers were high or low performing.

Phase II Data Analysis

Phase II analyses included 32 teachers from the two identified groups: teachers in the top quartile in student achievement gain ($n = 17$) and teachers in the bottom-quartile ($n = 15$). The top-quartile and bottom-quartile teachers were identified based on the TAIs generated in Phase I of this study. Due to the relatively small numbers in the case study sample sizes for each of the groups and the nature of the data collected, more conservative Mann-Whitney U tests were used in subsequent analyses.

Phase II Results

Teacher Beliefs. The two groups of teachers completed the TSES, which assessed teacher beliefs about their capabilities concerning instructional strategies, student engagement, and classroom management (Table 7). The Mann-Whitney U test comparing the teacher groups ($n = 31$; $U = 56.5$, $p = .25$)⁴ did not indicate significant differences between the groups.

Questioning Activity. The observers noted questions asked by the teachers and their students. The questions were recorded according to low, intermediate, or high cognitive demand levels. Low cognitive demand questions included fact-based questions. Intermediate and high cognitive demand questions followed the patterns established in Bloom's taxonomy. The raw data were standardized to questions per minute for each question level. Two additional variables, student questions and teacher questions, were calculated as the total number of questions per hour. As indicated in Table 7, the analyses indicated no differences between the teacher groups.

Time on Task. During a one-hour period, the observers gathered five-minute samplings of the number of students visibly disengaged from the lesson and the number of students who

Table 7. Analyses for Selected Teacher Dimensions in Relation to Teachers' Impact on Student Achievement

Dimension	Top-quartile mean rank	Bottom-quartile mean rank	Mann-Whitney results	Significance
Teacher beliefs	11.35	14.79	56.5	.25
Teacher questioning	16.38	16.63	125.5	.94
Student questioning	17.97	14.83	102.5	.35
Time on task/disruptive behavior	12.91	20.57	66.5	.02*
Time on task/visibly disengaged	13.59	19.80	78.0	.06

$p < .05$

Table 8. Analysis of Teacher Effectiveness Dimensions by Top- and Bottom-Quartile Teachers

Teacher effectiveness dimension	Top-quartile mean rank	Bottom-quartile mean rank	Mann-Whitney results	Significance
I1 Instructional differentiation	16.88	14.93	104.0	.57
I2 Instructional focus on learning	18.65	12.79	74.0	.08
I3 Instructional clarity	17.76	13.86	89.0	.25
I4 Instructional complexity	17.76	13.86	89.0	.25
I5 Expectations for student learning	17.44	14.29	94.5	.34
I6 Use of technology	14.79	14.21	94.0	.87
A1 Assessment for understanding	17.47	14.21	94.0	.34
A2 Quality of verbal feedback	18.09	13.46	83.5	.16
M1 Classroom management	19.76	11.43	55.0	.01*
M2 Classroom organization	19.41	11.86	61.0	.02*
P1 Caring	17.44	14.29	94.5	.34
P2 Fairness and respect	18.53	12.93	76.0	.09
P3 Positive relationships	19.24	12.07	64.0	.03*
P4 Encouragement of responsibility	19.62	11.61	57.5	.01*
P5 Enthusiasm	18.09	13.46	83.5	.16

For the dimensions, I = instructional delivery; A = student assessment; M = learning environment/classroom management; P = personal qualities. Top quartile $n = 17$; bottom quartile $n = 14$.

* $p < .05$.

initiated disruptive activities. Observers also recorded comments regarding off-task behaviors and teacher responses, such as "Student A talking to Student B during lesson and being corrected by Teacher," "Student C going to pencil sharpener multiple times," and "Student D playing with pencil holder unobserved by teacher." As shown in Table 7, the results indicated that there were no statistically significant group differences at a .05 level for disengaged students. However, the results did show a significant difference in terms of disruptive behavior, with the classrooms of the bottom-quartile teachers experiencing more disruptions. In terms of actual disruptive episodes per hour, the bottom-quartile teachers' classes had three times as many disruptive events as compared to the top-quartile teachers' classes.

Teacher Effectiveness Ratings. Data on the effectiveness of the teachers in their classrooms included ratings by the observers using the Teacher Effectiveness Rating Form. The observers individually rated the effectiveness of the teachers in the four broad areas of instructional skills, assessment skills, classroom management, and personal qualities. Next, the observers compared and discussed their respective ratings

and reached consensus on the most accurate rating for each item when differences occurred. Table 8 presents the descriptive data for the 15 teacher effectiveness dimensions using the consensus ratings.

The Mann-Whitney U test results indicated there were statistically significant differences favoring the effective teachers on four of the 15 variables: classroom management (M1), better organized (M2), more positive relationships with their students (P3), and greater student responsibility (P4).

Discussion of Findings

A central purpose of the study was to determine if the teaching practices of effective (top-quartile) and less effective (bottom-quartile) teachers differed in any discernable ways. Obviously, student achievement is just one educational outcome measure. It does not address the extent to which high- versus low-performing teachers might differ in their instructional practices, use of questioning, and classroom management practices. It measures the outcome, a crucial

consideration in effective teaching, but does not measure the process, or instructional practices, that result in increased student achievement.

Comparison of Student Achievement Among Teachers

Although numerous studies have analyzed the value-added impact of teachers on student achievement gain scores (Mendro, 1998; Nye, Konstantopoulos, & Hedges, 2004; Palardy & Rumberger, 2008; Sanders & Horn, 1994; Wright, Horn, & Sanders, 1997), few empirical studies have addressed the matter of what high-performing versus low-performing teachers do differently. In one such study, Stronge et al. (2008) not only examined the measurable impact that teachers have on student learning but also further explored the practices of effective versus less effective teachers. Although the studies that examine the value-added impact that teachers have on student learning explore the practices of effective teachers differently, one common finding emerges: Teachers have a measurable impact on student learning. Palardy and Rumberger (2008) concluded that "a string of highly effective or ineffective teachers will have an enormous impact on a child's learning trajectory during the course of Grades K-12" (p. 127). The results of the current study support this statement in that the differences in student achievement in mathematics and reading for effective teachers and less effective teachers were more than 30 percentile points.

Comparison of Teaching Practices Between Top- and Bottom-Quartile Teachers

The purpose of Phase II was to determine if fifth-grade teachers with high and low student achievement were measurably different based on selected classroom practices. Because of the small sample sizes between the two groups in the case studies, the statistical power of many of the comparisons was weakened. In addition, having to depend on invited teachers who agreed to participate in Phase II may have introduced an equalizing force across the two groups in that only the more confident and articulate teachers in each group may have agreed to participate.

Student Disruptive Behavior

As noted in Table 7, the disruptive behavior of students in the top- and bottom-quartile classes was significantly different. On average, bottom-quartile teachers had disruptions in their classrooms every 20 minutes, whereas top-quartile teachers had disruptions once an hour. This finding is consistent with previous research that found that top-quartile teachers experienced one-half disruption per hour versus five disruptions per hour for low-quartile teachers (Stronge et al., 2008).

Teacher Effectiveness Variables

To illustrate the differences between the high- and low-performing teacher groups, a graphical representation of the measures for the 15 teacher effectiveness dimensions was constructed. All of the variables in the chart were standardized so that all variables could be observed along a common metric (Figure 5).

The 15 teacher effectiveness dimensions rated by the observers were categorized into four domains of practice. Differences were found between the two groups of teachers in the areas of classroom management and personal qualities but not in the areas of instruction or assessment.

As noted in Table 8, the top-quartile teachers had significantly higher ratings in four of the teacher effectiveness dimensions. Although not statistically significant, as depicted in Figure 5, the top-quartile group of teachers had higher mean ratings than the bottom-quartile teachers on all dimensions. Based on these results, one hypothesis is that teachers who are "effective" in terms of their student achievement results have some particular set of attitudes, approaches, strategies, or connections with students that manifest themselves in nonacademic ways (positive relationships, encouragement of responsibility, classroom management, and organization) and that lead to higher achievement.

Classroom management. Top-quartile teachers scored significantly higher in the two dimensions related to classroom management. One dimension related to managing the classroom: establishing routines, monitoring student behavior, and using time efficiently and effectively. The other related to classroom organization, which includes ensuring availability of necessary materials for student use, physical layout of the classroom, and using space effectively. This finding is supported by a previous study that found that teachers who were more effective, in terms of student achievement, were more organized, used routines and procedures with greater efficiency, and held higher expectations of their students' behavior (Stronge et al., 2008). Similarly, a study that linked second- and third-grade reading achievement to teacher behaviors found that in teachers' classrooms where students made greater achievement gains, the teacher maintained control over the classroom (Fidler, 2002).

Personal qualities. Two dimensions in personal qualities indicated a significant difference between effective and less effective teachers. Top-quartile teachers scored higher in fairness and respect as well as in having positive relationships with students. An earlier exploratory study that examined practices of more and less effective teachers found similar results (Stronge et al., 2008).

Phase II of this study compared the key qualities of teachers identified as effective or ineffective by the achievement gains they facilitated with their students. However, we recognize that the instruments used did not provide a holistic portrait of teacher effectiveness. Similar to the limitation of traditional

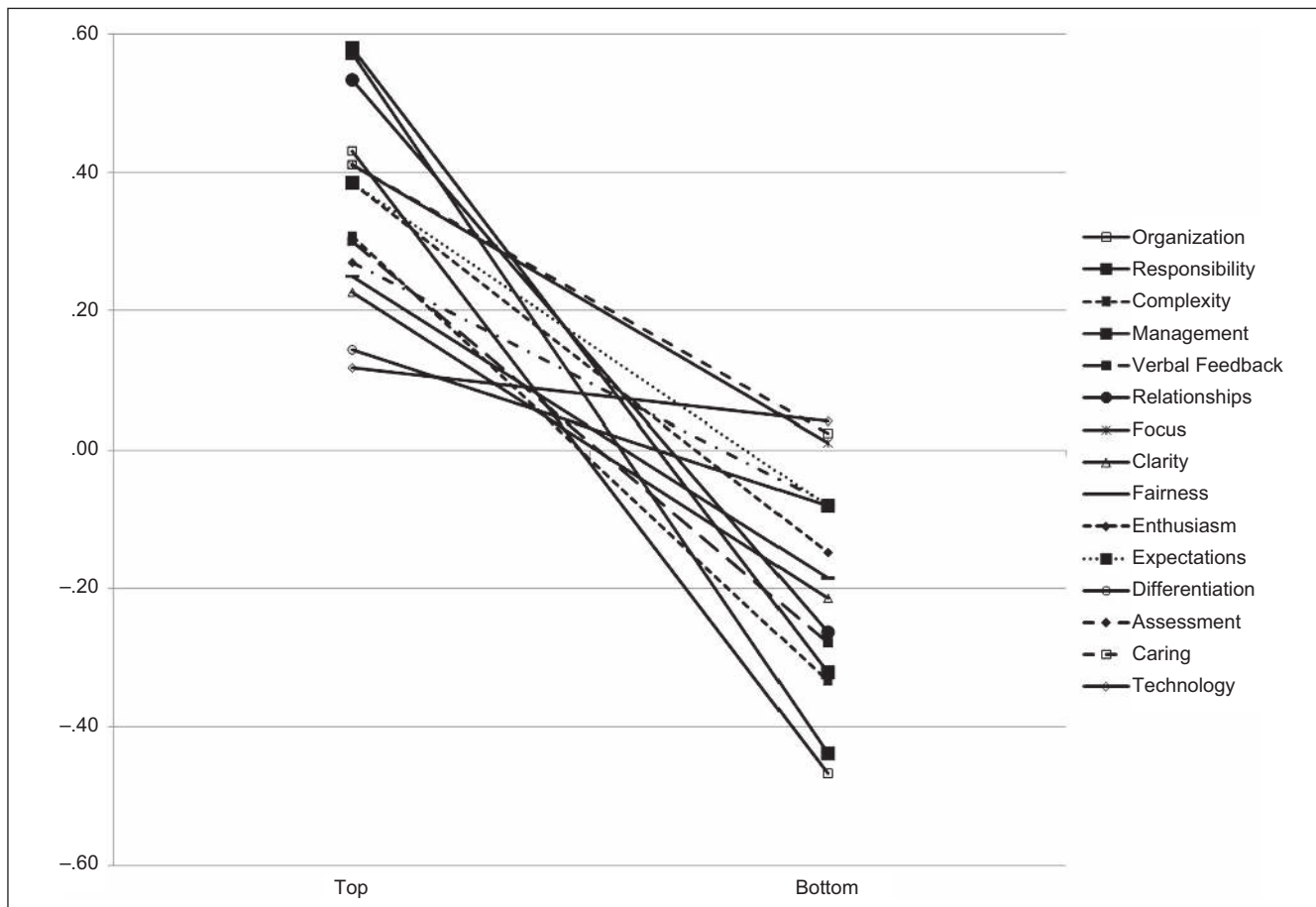


Figure 5. Teacher effectiveness variables

product–process research, this study did not find silver-bullet practices that would lead to higher levels of teacher effectiveness for all teachers (Campbell, Kyriakides, Muijs, & Robinson, 2004). We acknowledge that effective teaching involves a dynamic interplay among content, pedagogical methods, characteristics of learners, and the contexts in which the learning will occur (Schalock, Schalock, Cowart, & Myton, 1993). Consequently, we recommend continued research to explore the nuances, settings, complexities, and interdependencies of teachers teaching in their natural classroom settings.

How Effective and Ineffective Teachers Differ: Synthesis of Findings

Considerations and Cautions in Interpreting and Applying Study Findings

This study found that top-quartile teachers had fewer classroom disruptions, better classroom management skills, and better relationships with their students than did bottom-

quartile teachers. One interpretation of this finding is that differences in personalities and dispositions of students can better explain the differences found among the teachers. Perhaps this one year, the higher quartile teachers had students who had less difficulty behaving in school. Although this is certainly a possibility, we doubt that the differences in students are wholly responsible for the differences in teachers. A study of first-grade achievement in reading and mathematics did find that the composition of a class was a strong predictor of student achievement gains, hence the need to control for student characteristics (Palardy & Rumberger, 2008). We did control for multiple characteristics that might affect student achievement and we still found differences in student achievement. To further bolster our interpretation of the results, study after study has shown that teacher effectiveness varies from classroom to classroom. A study of student achievement in mathematics and reading found that socioeconomic status was not as strong a predictor of student achievement as was the teacher (Nye et al., 2004). Further, studies support that effective teachers are adept at classroom management (Cotton, 2000).

Although we did not find significant differences between effective and ineffective teachers on the dimensions of instructional delivery and assessment, in no way are we suggesting that these teacher skill areas are unimportant. To the contrary, we recognize the deep research base supporting them, and indeed, it would be counterintuitive to suggest otherwise. No doubt, the lack of statistical findings for instructional delivery and assessment, at least partially, is an artifact of the small sample size with which we were working.

Issues Concerning Small Sample Size

This study was composed of two distinct phases. In Phase I, a fairly large sample of students and classrooms was used. A major consideration in interpreting and applying the results of this phase is the representativeness of the sample. Our sample represented diversity but was limited to four school districts in one state. In Phase II, the sample was much more modest in that we used a cross-case analysis approach. Certainly, the statistical power of the analyses in Phase II is reduced. In addition, it should be noted that in Phase II classroom observations were restricted to three-hour classroom visits by teams of trained observers. Although this protocol allowed for verification of observational findings through interrater agreement, it is a relatively small observation sample. We did not distinguish between observations of math lessons or reading lessons. As a result, generalizability is more limited.

Issues Concerning Value-Added Model Data

VAM data are particularly useful for determining the value-added effects of schools or teachers. By using sophisticated models with longitudinal data on student achievement, researchers have been able to estimate the powerful effects that teachers have on student achievement by controlling for those factors that have been shown to affect student achievement (Palardy & Rumberger, 2008; Rowan et al., 2002). Without controlling for prior achievement and socioeconomic status, the effects of teachers on student achievement can be masked by student-level variables.

Issues Concerning the Influence of Teacher Experience on Value-Added Modeling

Rockoff (2004) found that teaching experience significantly raised student test scores for both reading and math computation (but not math concepts) at the elementary level, particularly in reading subject areas. On average, reading test scores differed by approximately 0.17 standard deviations between beginning teachers and teachers with 10 or more years of experience. For mathematics subject areas, the effects of experience were smaller. The first two years of teaching experience appear to raise scores significantly in

math computation. However, in this study, subsequent years of experience appeared to have a negative impact on test scores (a counterintuitive finding).

Rivkin, Hanushek, and Kain (2005) noted that at the elementary level, teacher effectiveness increased during the first year or two but leveled off after the third year. They also found that differences between new and experienced teachers account for only 10% of the teacher quality variance in mathematics and somewhere between 5% and 20% of the variance in reading. Hanushek, Kain, O'Brien, and Rivkin (2005) linked student achievement in fourth- through eighth-grade mathematics with various teacher characteristics, including teaching experience. They found teacher experience associated positively with student achievement gains but only for the first few years.

Contrasting with the above findings, a study by Munoz and Chang (2007), which used HLM to estimate the effects of teacher characteristics in high school reading achievement gains in a large urban district, found that teaching experience is not predictive of student growth rates in high school reading. At the elementary school level, a VAM study by Heistad (1999) also found no significant correlation between teacher experience and student achievement; thus, the effectiveness of second-grade reading teachers was not dependent on their years of service. In summary, the extant research generally supports the impact of teaching experience on student learning only for the first few years.

In this study, we checked for the influence of teaching experience on teacher effectiveness as explained by their students' value-added gain scores. We did not find any relationship between teacher experience and effectiveness. In addition, we compared the achievement results of teachers with less than five years, five to 10 years, and more than 10 years of experience. We found no significant differences among these groups. Nonetheless, this issue should be further explored.

Issues Concerning Stability of Teacher Rankings in Value-Added Modeling

One reason that some policy makers and researchers are skeptical about using VAM is that teachers' performances as measured by VAMs can fluctuate over time and can be unevenly distributed across districts (Viadero, 2008a). Part of the problem is that value-added calculations operate on various assumptions that may not hold. To illustrate, results might be biased if it turns out that a school's students are not randomly assigned to teachers or that significant amounts of student data are missing (Viadero, 2008b). However, other researchers (Goldhaber & Hansen, 2008, 2010; Sanders & Horn, 1994; Sass, 2008) have found that estimates of school and teacher effects tend to be consistent from year to year and that they are dependable even with some gaps in data.

Issues Concerning the Use of Value-Added Model Data in Teacher Evaluation

A critical consideration in the application of studies such as this that attempt to connect teacher effectiveness with student achievement is, How should we use the results? More specifically, should teachers be evaluated based on measures of student achievement? This intriguing question has received ample attention in recent years, as numerous states and individual school districts have experimented with VAM applications for teacher evaluation, performance pay, or merely monitoring teacher and school effectiveness in value-added terms. With the Obama administration's focus on gain scores and teacher evaluation as a key component in its Race to the Top initiative, the public policy debate on the connection between teacher effectiveness and student academic growth is likely to intensify.

This study was not explicitly designed to address teacher evaluation or performance pay issues. However, we would like to offer a word of caution. When VAM (or any other method, for that matter) is used for high-stakes purposes, extraordinary care and diligence must be applied to ensure adherence to the premises of fair, ethical, and defensible treatment.

If VAM data are to be employed in high-stakes teacher decisions, then one key consideration is not to rely on data derived from a single year as the predictor of teacher effectiveness. As noted previously, we based our TAI scores on a single year of student residual gain scores. This was a matter of necessity, due to restrictions in data availability, and not one of policy recommendation. Indeed, basing teacher evaluation or teacher pay on a single year of student growth evidence is woefully inadequate.

A second key consideration is that teacher evaluation, teacher pay, or any other teacher-specific decision should never rely on a single source of evidence. Despite the daunting technical challenges of adequately measuring the influence of teachers on student achievement, measures of student learning do provide the ultimate accountability for teacher success (Tucker & Stronge, 2006). However, just as observation-only teacher evaluation systems are systematically flawed (Peterson, 2000; Stronge & Tucker, 2003), so too would be systems that put all their eggs in a VAM basket. No single data source is valid or feasible for all teachers in all school districts (Peterson, 2006). Thus, our recommendation is that if VAM-generated evidence is to be considered in high-stakes applications, it be used as one source in a multifaceted review of teacher effectiveness (Stronge, 2006; Stronge & Tucker, 2003).

Conclusions

One conclusion regarding effective teachers is abundantly clear: The common denominator in school improvement and

student success *is* the teacher. Richard Riley (1998), former U.S. secretary of education, captured the essence of the importance of teachers in his discussion about teacher excellence and diversity:

Providing quality education means that we should invest in higher standards for all children, improved curricula, tests to measure student achievement, safe schools, and increased use of technology—but *the most critical investment we can make is in well-qualified, caring, and committed teachers* [italics added]. Without good teachers to implement them, no educational reforms will succeed at helping all students learn to their full potential. (p. 18)

Reforming American education is about enhancing learning opportunities and results for students. In the final analysis, as Carroll (1994) stated, “nothing, absolutely nothing has happened in education until it has happened to a student” (p. 87). The education challenge facing the United States and other countries around the world is not that our schools are not as good as they once were. It is that schools must help the vast majority of young people reach levels of skill and competence that were once thought to be within the reach of only a few (Darling-Hammond, 1996). Although various educational policy initiatives may offer the promise of improving education, nothing is more fundamentally important to improving our schools than improving the teaching that occurs every day in every classroom. To make a difference in the quality of education, we must be able to provide ready and well-founded answers to the question, What do good teachers do that enhances student learning? Although we strongly encourage readers to consider the limitations of this study and to apply caution when interpreting and inferring from its results, we hope this study moves us closer to answers to this fundamental question.

Acknowledgment

Thank you to the anonymous reviewers of earlier versions of this manuscript. Their insightful comments were instrumental in our attempts to bring greater focus and clarity to the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors received funding from the National Board for Professional Teaching Standards through SERVE at the University of North Carolina, Greensboro, for an earlier study from which the data in this article were drawn.

Notes

1. This study is a replication of a similar study conducted by Stronge, Ward, Tucker, and Hindman (2008). As with research in other fields (e.g., the medical profession), replication is a vital aspect for verifying and validating purported findings prior to broad-based acceptance and implementation. Nonetheless, this study differs from Stronge et al.'s study in that the present study (a) is based upon a substantially larger sample of teachers and students, particularly in Phase I; (b) was conducted in a different state with different school districts (in this case, districts with urban, rural, and suburban characteristics); and (c) was conducted with a different grade level as the target population (intermediate grade in the present study vs. primary grade in the earlier study).
2. We considered three options for centering data: noncentering, group centering, and grand mean centering. We chose grand mean centering based on the recommendations by Goe (2008) as appropriate for value-added data analysis and our objective of examining teacher effect.
3. Both groups scored at approximately the 43rd percentile on the fourth-grade end-of-course reading test that served as the fifth-grade pretest. The resulting differences on the posttests show a loss of approximately 22 percentile points for the bottom-quartile teachers' classes and a gain of approximately 11 percentile points for the students in the top-quartile teachers' classes.
4. The total sample size for the case study teachers was 32; however, there were missing data for one teacher for this analysis.

References

- Adams, C., & Singh, K. (1998). Direct and indirect effects of school learning variables on the academic achievement of African American 10th graders. *Journal of Negro Education, 67*(1), 48-66.
- Agne, K. J. (1992). Caring: The expert teacher's edge. *Educational Horizons, 70*(3), 120-124.
- Allington, R. L. (2002). What I've learned about effective reading instruction. *Phi Delta Kappan, 83*, 740-747.
- Anderson, L. W., & Krathwol, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge School of Education.
- Bain, H. P., & Jacobs, R. (1990, September). The case for smaller classes and better teachers. *Streamlined Seminar—National Association of Elementary School Principals, 9*(1).
- Bembry, K. L., Jordan, H. R., Gomez, E., Anderson, M. C., & Mendro, R. L. (1998, April). *Policy implications of long-term teacher effects on student achievement*. Paper presented at the 1998 Annual Meeting of the American Educational Research Association, San Diego, CA.
- Berliner, D. C., & Rosenshine, B. V. (1977). The acquisition of knowledge in the classroom. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 375-396). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bernard, B. (2003). Turnaround teachers and schools. In B. Williams (Ed.), *Closing the achievement gap: A vision for changing beliefs and practices* (pp. 115-137). Alexandria, VA: Association for Supervision and Curriculum Development.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 9-21.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.
- Black, R., & Howard-Jones, A. (2000). Reflections on best and worst teachers: An experiential perspective of teaching. *Journal of Research and Development in Education, 34*(1), 1-13.
- Bloom, B. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I. Cognitive domain*. New York: Longmans Green.
- Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro: University of North Carolina at Greensboro, Center for Educational Research and Evaluation.
- Campbell, J., Kyriakides, L., Muijs, D., & Robinson, W. (2004). *Assessing teacher effectiveness: Developing a differential model*. New York: RoutledgeFalmer.
- Carroll, J. M., (1994). *The Copernican plan evaluated: The evolution of a revolution*. Topsfield, MA: Copernican Associates.
- Cawelti, G. (Ed.). (2004). *Handbook of research on improving student achievement* (2nd ed.). Arlington, VA: Educational Research Service.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-43.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.
- Collinson, V., Killeavy, M., & Stephenson, H. J. (1999). Exemplary teachers: Practicing an ethic of care in England, Ireland, and the United States. *Journal for a Just and Caring Education, 5*(4), 349-366.
- Cotton, K. (2000). *The schooling practices that matter most*. Portland, OR, and Alexandria, VA: Northwest Regional Educational Laboratory and the Association for Supervision and Curriculum Development.
- Covino, E. A., & Iwanicki, E. (1996). Experienced teachers: Their constructs on effective teaching. *Journal of Personnel Evaluation in Education, 11*, 325-363.
- Cradler, J., McNabb, M., Freeman, M., & Burchett, R. (2002). How does technology influence student learning? *Learning and Leading With Technology, 29*(8), 46-50.
- Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. *Phi Delta Kappan, 78*, 193-200.

- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archive*, 8(1). Retrieved from <http://olam.ed.asu.edu/epaa/v8n1>
- Darling-Hammond, L., & Bransford, J. (2005). (Eds.). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass.
- Emmer, E. T., Evertson, C. M., & Anderson, L. M. (1980). Effective classroom management at the beginning of the school year. *Elementary School Journal*, 80(5), 219-231.
- Emmer, E. T., Evertson, C. M., & Worsham, M. E. (2003). *Classroom management for secondary teachers*. Boston: Allyn & Bacon.
- Fidler, P. (2002). *The relationship between teacher instructional techniques and characteristics and student achievement in reduced size classes*. Los Angeles: Los Angeles Unified School District.
- Goe, L. (2008, May). *Key issue: Using value-added models to identify and support highly effective teachers*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50-55. Retrieved from http://www.nuac.org/articles/pdf/mystery_goodteaching.pdf
- Goldhaber, D., & Hansen, M. (2008). *Is it just a bad class? Assessing the stability of measured teacher performance* (CRPE Working Paper No. 2008_5). Retrieved from http://www.crpe.org/cs/crpe/download/csr_files/wp_crpe5_badclass_nov08.pdf
- Goldhaber, D., & Hansen, M. (2010). *Using performance on the job to inform teacher tenure decisions: Brief 10*. Washington, DC: Urban Institute, National Center for Analysis of Longitudinal Data in Education Research.
- Good, T. L., & Brophy, J. E. (1997). *Looking in classrooms* (7th ed.). New York: Addison-Wesley.
- Good, T. L., & McCaslin, M. M. (1992). Teacher licensure and certification. In M.C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 1352-1388). New York: Macmillan.
- Guskey, T. R. (Ed.). (1996). *Communicating student learning: 1996 yearbook of the Association for Supervision and Curriculum Development*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, 61(2), 280-288.
- Hanushek, E. A. (2008, May). *Teacher deselection*. Retrieved from http://www.stanfordalumni.org/leadingmatters/san_francisco/documents/Teacher_Deselection-Hanushek.pdf
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w11154.pdf>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heistad, D. (1999, April). *Teachers who beat the odds: Value-added reading instruction in Minneapolis 2nd grade*. Paper presented at the Annual American Educational Research Association Conference, Montreal, Canada.
- Johnson, B. L. (1997). An organizational analysis of multiple perspectives of effective teaching: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 69-87.
- Jordan, H. R., Mendro, R. L., & Weerasinghe, D. (1997, July). *Teacher effect on longitudinal student achievement*. Paper presented at the CREATE Annual Meeting, Indianapolis, IN.
- Knapp, M. S., Shields, P. M., & Turnbull, B. J. (1992). *Academic challenge for the children of poverty: Summary report*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.
- Langer, J.A. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal*, 38(4), 837-880.
- Lewis, L., Parsad, B., Carey, N., Bartfai, N., Farris, E., & Smerdon, B. (1999). Teacher quality: A report on the preparation and qualifications of public school teachers. *Education Statistics Quarterly*, 1(1). Retrieved from http://nces.ed.gov/programs/quarterly/Vol_1/1_1/2-esq11-a.asp
- Marzano, R. J. (2003). *What works in schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J. (2006). *Assessment and grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mason, D. A., Schroeter, D. D., Combs, R. K., & Washington, K. (1992). Assigning average-achieving eighth graders to advanced mathematics classes in an urban junior high. *Elementary School Journal*, 92(5), 587-599.
- Matsumura, L. C., Patthey-Chavez, G. G., Valeds, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal*, 103, 3-26.
- McBer, H. (2000). *Research into teacher effectiveness: A model of teacher effectiveness* (Research Report No. 216). Nottingham, UK: Department for Education and Employment.
- Mendro, R. L. (1998). Student achievement and school and teacher accountability. *Journal of Personnel Evaluation in Education*, 12, 257-267.
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 21(2), 165-178.
- Munoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education*, 20, 147-164.
- National Academy of Education. (2008). *Teacher quality: Education policy white paper*. Washington, DC: Author. Retrieved from http://www.naeducation.org/Teacher_Quality_White_Paper.pdf

- National Association of Secondary School Principals. (1997). Students say: What makes a good teacher? *Schools in the Middle*, 6(5), 15-17.
- National Center for Education Statistics. (1997). *Time spent teaching core academic subjects in elementary schools: Comparisons across community, school, teacher, and student characteristics*. Washington, DC: U.S. Department of Education.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257.
- Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education*, 79(4), 126-137.
- Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, 30(2), 111-140.
- Peart, N. A., & Campbell, F. A. (1999). At-risk students' perceptions of teacher effectiveness. *Journal for a Just and Caring Education*, 5(3), 269-284.
- Peterson, K. D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Peterson, K. D. (2006). Using multiple data sources in teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (2nd ed., pp. 212-232). Thousand Oaks, CA: Corwin Press.
- Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., & Morrow, L. (1998). *The nature of effective first-grade literacy instruction* (CELA Research Report No. 11007). Albany, NY: Center on English Learning and Achievement.
- Randall, R., Sekulski, J. L., & Silberg, A., (2003). Results of direct instruction reading program evaluation longitudinal results: First through third grade, 2002-2003. Milwaukee, WI: University of Wisconsin-Milwaukee. Retrieved from http://www.uwm.edu/News/PR/04.01/DI_Final_Report_2003.pdf
- Riley, R. W. (1998). Our teachers should be excellent, and they should look like America. *Education and Urban Society*, 31, 18-29.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247-252.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1), 175-214.
- Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on student achievement. *Sociology of Education*, 70, 256-284.
- Rowan, B., Correnti, R., & Miller, R. J. (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(8), 1525-1567.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299-311.
- Sass, T. R. (2008). *The stability of value-added measures of teacher quality and implications for teacher compensation policy*. Washington, DC: Urban Institute, Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.urban.org/UploadedPDF/1001266_stabilityofvalue.pdf
- Schacter, J. (1999). *The impact of education technology on student achievement: What the most current research has to say*. Santa Monica, CA: Milken Exchange on Educational Technology.
- Schalock, H. D., Schalock, M. D., Cowart, B., & Myton, D. (1993). Extending teacher assessment beyond knowledge and skills: An emerging focus on teacher accomplishments. *Journal of Personnel Evaluation in Education*, 7, 105-133.
- Sternberg, R. J. (2003). What is an "expert student"? *Educational Researcher*, 32(8), 5-9.
- Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5, 41-48.
- Stronge, J. H. (2002). *Qualities of effective teachers*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice*. (2nd ed., pp. 1-23). Thousand Oaks, CA: Corwin Press.
- Stronge, J. H. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association of Supervision and Curriculum Development.
- Stronge, J. H., McColsky, W., Ward, T., & Tucker, P. (2005). *Teacher effectiveness, student achievement, and National Board for Professional Teaching Standards*. Greensboro, NC: SERVE, University of North Carolina at Greensboro.
- Stronge, J. H., & Tucker, P. D. (2003, Summer). *Handbook on teacher evaluation*. Larchmont, NY: Eye on Education.
- Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2008). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, 20(3-4), 165-184.
- Taylor, B. M., Pearson, P. D., Clark, K. F., & Walpole, S. (1999). Center for the improvement of early reading achievement: Effective schools/accomplished teachers. *Reading Teacher*, 53(2), 156-159.
- Tschannen-Moran, M. (2000, Spring). The ties that bind: The importance of trust in schools. *Essentially Yours*, 4, 1-5.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17, 783-805.
- Tucker, P. D., & Stronge, J. H. (2006). Use of student achievement in teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice*. (2nd ed., pp. 152-169). Thousand Oaks, CA: Corwin Press.
- Viadero, D. (2008a, May 7). Scrutiny heightens for "value added" research method. *Education Week*, 27(36), 1, 12-13.
- Viadero, D. (2008b, May 7). "Value-added" pioneer says stinging critique of method is off-base. *Education Week*, 27(36), 13.

- Wahlage, G., & Rutter, R. (1986). *Evaluation of model program for at-risk students*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Wang, M., Haertel, G. D., & Walberg, H. (1993). What helps students learn? *Educational Leadership*, 51(4), 74-79.
- Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.
- Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics*. Princeton, NJ: Educational Testing Service.
- Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Milken Family Foundation and Educational Testing Service.
- Wenglinsky, H. (2004). Closing the racial achievement gap: The role of reforming instructional practices. *Education Policy Analysis Archives*, 12(64). Retrieved from <http://epaa.asu.edu/epaa/v12n64/>
- Wentzel, K. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development*, 73, 287-301.
- Wolk, S. (2002). *Being good: Rethinking classroom management and student discipline*. Portsmouth, NH: Heinemann.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67.
- Yesseldyke, J., & Bolt, D. M. (2007). Effect of technology enhanced continuous progress monitoring on math achievement. *School Psychology Review*, 36(3), 453-467.
- Zahorik, J., Halbach, A., Ehrle, K., & Molnar, A. (2003). Teaching practices for smaller classes. *Educational Leadership*, 61(1), 75-77.

About the Authors

James H. Stronge is Heritage Professor in the Educational Policy, Planning, and Leadership Area at the College of William and Mary in Williamsburg, Virginia. His research interests include policy and practice related to teacher quality and teacher evaluation.

Thomas J. Ward is associate dean and professor in the School of Education at the College of William and Mary in Williamsburg, Virginia. His research interests include issues related to school and teacher effectiveness.

Leslie W. Grant is a visiting assistant professor in the School of Education at the College of William and Mary in Williamsburg, Virginia. Her research interests include issues related to student assessment and teacher quality.