



University of Pennsylvania
ScholarlyCommons

CPRE Journal Articles

Consortium for Policy Research in Education
(CPRE)

9-2005

Teacher Quality and Educational Equality: Do Teachers With Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?

Geoffrey D. Borman

Steven M. Kimball

Follow this and additional works at: http://repository.upenn.edu/cpre_articles

 Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Borman, Geoffrey D. and Kimball, Steven M., "Teacher Quality and Educational Equality: Do Teachers With Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?" (2005). *CPRE Journal Articles*. 2.
http://repository.upenn.edu/cpre_articles/2

[View on the CPRE website.](#)

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/cpre_articles/2
For more information, please contact libraryrepository@pobox.upenn.edu.

Teacher Quality and Educational Equality: Do Teachers With Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?

Abstract

Using standards-based evaluation ratings for nearly 400 teachers, and achievement results for over 7,000 students from grades 4–6, this study investigated the distribution and achievement effects of teacher quality in Washoe County, a mid-sized school district serving Reno and Sparks, Nevada. Classrooms with higher concentrations of minority, poor, and low-achieving students were more likely to be taught by teachers with lower evaluation scores. Two-level multilevel models, nesting students within classrooms, tended to show higher mean achievement in classrooms taught by teachers of higher than lower quality, with differences of approximately one-tenth of 1 standard deviation. Findings relating teacher quality to closing within-classroom achievement gaps, though, were mixed. Implications are discussed related to teacher evaluation, teacher quality, and educational inequality.

Disciplines

Educational Assessment, Evaluation, and Research

Comments

[View on the CPRE website.](#)

Teacher Quality and Educational Equality: Do Teachers with Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?

Geoffrey D. Borman

Steven M. Kimball

University of Wisconsin—Madison

Abstract

Using standards-based evaluation ratings for nearly 400 teachers, and achievement results for over 7,000 students from grades 4–6, this study investigated the distribution and achievement effects of teacher quality in Washoe County, a mid-sized school district serving Reno and Sparks, Nevada. Classrooms with higher concentrations of minority, poor, and low-achieving students were more likely to be taught by teachers with lower evaluation scores. Two-level multilevel models, nesting students within classrooms, tended to show higher mean achievement in classrooms taught by teachers of higher than lower quality, with differences of approximately one-tenth of 1 standard deviation. Findings relating teacher quality to closing within-classroom achievement gaps, though, were mixed. Implications are discussed related to teacher evaluation, teacher quality, and educational inequality.

According to the work of Hanushek (1992), the difference between being taught by a good and a bad teacher can translate into a full grade level of achievement in a single school year. Beyond these potential short-term benefits, the research of Sanders and Rivers (1996) has indicated that teacher effects can be enduring and cumulative, whether they advance student achievement or leave children behind. As Sanders and Rivers demonstrated, after 2 years, the performance of fifth-grade students was still affected by the quality of their third-grade teacher. Further, students whose initial achievement was comparable can have vastly different academic outcomes as a result of the sequence of teachers to whom they are assigned. Indeed, evidence of the strong effects of teachers on student

achievement can be traced to the classic Coleman Report, which concluded that teacher characteristics tended to explain more variance in student achievement than any other school resource (Coleman et al., 1966).

Some research also shows that low-income and minority students are disproportionately taught by underqualified teachers, including teachers who are out-of-field, inexperienced, or fail to meet their state's teacher licensing and certification standards. Examining Texas data, Kain and Singleton (1996) found that African American and Latino children were far more likely to be taught by teachers who scored poorly on the Texas Examination of Current Administrators and Teachers (TECAT). Indeed, as the percentage of nonwhite children in the school increased, the average teacher score declined. Finding the same pattern, Ferguson (1998) wrote that "in Texas, as certainly in other places, attracting and retaining talented people with strong skills to teach in the districts where black students are heavily represented is part of the unfinished business of equalizing educational opportunity" (p. 354).

In contrast, analyses of national data have shown relatively equal distributions of teacher qualifications, including years of experience, highest degree earned, and certifications, across schools serving students of higher and lower poverty (Borman & Rachuba, 1999). Rowan, Correnti, and Miller (2002) also found, based on the national Prospects data, that differences in classroom effects were not systematically disadvantaging students from different demographic groups. Deflections from expected gains were not correlated with deflections in the next year, indicating that achievement inequality was due to chance assignment of students to teachers who varied in effectiveness.

Nonetheless, responding to data on both the importance of good teaching and the potential inequalities in the distribution of teacher talent, Congress set out, when it

passed the No Child Left Behind (NCLB) Act, to improve the teaching force in general and, most important, the teaching force in schools receiving Title I funds. The teacher-quality provisions of NCLB require states to adopt minimum standards for who can be considered a "highly qualified" teacher, measure the extent to which the state provides such teachers to all students, and adopt goals and plans to ensure that all students are taught by qualified teachers.

Despite these calls for greater teacher quality for disadvantaged students in Title I schools, there is surprisingly little evidence that highly qualified teachers successfully close the achievement gaps within the classrooms that they teach. The analysis by Rowan et al. (2002) of Prospects data looked at the effects of teachers on students from different social backgrounds, allowing random effects across classrooms for student minority status, gender, and socioeconomic status (SES). They found that relations between these background variables and annual achievement gains generally exhibited reliable variation across classrooms. In other words, the size of the achievement gaps separating children of different backgrounds, for instance, minority and non-minority students, varied depending on which classroom the students attended from within the same school. Although Rowan et al. were less successful in identifying teacher characteristics and instructional practices that explained these classroom-to-classroom differences, the results did suggest that some teachers may be more effective than others in closing achievement gaps.

In a paper examining the validity of the National Board certification process, Goldhaber and Anthony (2004) explored differential effects on students taught by current and future National Board Certified Teachers (NBCTs). Teacher quality, as determined by meeting the National Board for Professional Teaching Standards, was associated with the strongest effects for students from lower-SES backgrounds. Among classes taught by current NBCTs, the size of the ef-

fect was larger for students receiving free or reduced-price lunch (.13 *SDs* in reading and .11 *SDs* for math) than for those who did not receive subsidized lunch (.02 *SDs* for reading and .06 *SDs* for math).

Other classroom-based interventions have shown equity effects as well. For instance, the Tennessee class-size experiment provided evidence that small classes of approximately 15 children had particularly strong effects on African American students (Krueger & Whitmore, 2001). Similarly, studies of the Success for All comprehensive reading program have demonstrated positive effects for all participating students, with the largest effects for students starting the intervention among the lowest 25% on the pretest (Slavin & Madden, 2001). However, there remains much to be explained about the effect of teacher characteristics and behaviors on student achievement. As Ferguson (1998) asserted, "social scientists are unable to identify and measure most of the characteristics that make one teacher more effective than another. No one characteristic is a reliable predictor of a teacher's performance. Nor are most teachers uniformly good or bad in every subject or with all types of students" (p. 351).

In the current study, we used teacher ratings from a standards-based teacher evaluation system from Washoe County, Nevada, to measure teacher quality. First, we ascertained the extent to which teacher quality, as measured by the evaluation system, was equally distributed across classrooms varying in poverty, minority concentration, and achievement. Second, net of student background and teachers' experience, we assessed the relation between this measure of teacher quality and both overall classroom mean achievement and within-classroom effects on social equality. Most fundamentally, we asked whether high-quality teaching was related to better outcomes for all children and more equal outcomes among children from poor and nonpoor, minority and white, and higher and lower achievement backgrounds.

How Teachers Achieve Educational Equality

There are few strong theoretical or empirical perspectives on how teacher quality might affect disadvantaged students' achievement. For instance, does a high-quality teacher simply have a stronger effect on the educational outcomes of minority, poor, and low-achieving students? In other words, do these students exhibit a differential sensitivity to high-quality instruction that enables them to benefit more than their more advantaged peers? Does the broader construct of high-quality teaching include a belief in equity, and instructional skills, that enable teachers to help all low-achieving students catch up with their peers? In this case, simply assigning a high-quality teacher to teach minority, low-SES, and low-achieving students would help close achievement gaps.

However, teachers may need to possess more highly specialized skills, beliefs, or other characteristics that enable them to achieve educational equality in their classrooms. Being a "highly qualified" teacher may not be enough. Instead, teachers who achieve equality in their classrooms may have to, for instance, be adept at applying principles of multicultural education, be highly committed to the ideal of equity, or be of the same race and social class of the children they are teaching. Characteristics such as these may not be captured and adequately measured by traditional constructs of teacher quality.

Standards-Based Teacher Evaluation

Standards-based teacher evaluation systems are based on a common conception of teaching, developed from empirical and theoretical literature on effective teaching behaviors, and assessed using multiple, authentic sources of teaching evidence (Danielson, 1996; Danielson & McGreal, 2000). These systems are designed to assess teaching practice using a comprehensive set of standards and rubrics with the intention of enhancing instruction and strengthening educational accountability. Evaluation sys-

tems adapted from the Framework for Teaching (Danielson, 1996) include four broad domains of practice covering instructional planning, classroom management, instructional interactions, and professional responsibilities. Each domain includes a set of rubrics delineating specific teaching behaviors. Considered in total, the domains, rubrics, and sources of evidence can be used to construct an overall measure of teaching quality for use in formative and summative evaluation decisions. If evaluation scores on such standards-based systems are shown to be valid measures of teaching practice and have a positive relation to student achievement, they could provide a useful source of information on the distribution and effects of teacher quality (Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004; Odden, Borman, & Fermanich, 2004).

Standards-based evaluation systems may also help identify teaching behaviors and strategies that improve achievement of traditionally underserved students, including those with a history of low achievement or from low-SES or minority backgrounds. The rubrics provide evaluators and teachers with a guide to monitor and evaluate teaching performance intended to benefit all students. Included are elements related to tailoring instruction to the needs of lower-achieving and minority students. For example, in the Framework for Teaching, Danielson (1996) asserted that excellent instructional design not only reflects a deep understanding of content and pedagogy but also "includes sound assessment methods, and is appropriate to the range of students in the class" (p. 30). Teachers who rate highly on the evaluation system may also be better able to reduce achievement gaps between students from different social backgrounds and with differential achievement (i.e., low-performing students).

In this article we explore whether the Washoe County, Nevada, teacher evaluation system results, which are based on evaluation standards representing a common conception of teaching, explain differ-

ences in achievement for students who vary in prior achievement and social background. In addition, we examine the distribution of teacher quality among classrooms. Teacher quality is represented by scores on the teacher performance evaluation system. The study thus addresses two primary questions: (1) Is teacher quality distributed equally across classrooms of varying compositions; and, (2) Is teacher quality associated with both excellence and equality in terms of student achievement?

Method

Geographically, the Washoe County School District encompasses urban, suburban, and rural schools and is the second largest in Nevada. The district serves the communities of Reno and Sparks. There is rapid population growth in this region, socioeconomic diversity, and a sizable minority population, including a large number of Latino families. Given the growth, the district has a continual demand for teachers, with about 400 hired annually. In addition, the district has been operating its standards-based teacher evaluation system since the fall of 2000 and therefore has several consecutive years of data on teacher performance. Finally, the state and district administer a variety of measures to assess student performance relative to state academic standards.

Measures

Students. Student demographic data were made available by the district and used to construct dummy variables for minority status and student eligibility for free and reduced-price lunch. We obtained student achievement results from district and state norm-referenced and criterion-referenced tests (CRTs) for mathematics and reading in the third through sixth grades. Each of the three assessments we used for the study was designed to measure proficiency on the Nevada State Content and Performance Standards.

The first pair of assessments we used,

the grades 3 and 5 state CRTs, were developed by Harcourt Assessment, Incorporated in collaboration with representatives from the Nevada Department of Education, Nevada state educators, and personnel from the WestEd Regional Education Laboratory (Harcourt Assessment, Inc., 2004). Test items and results from this assessment were reviewed by the Nevada Department of Education and Harcourt and were found to be reliable, valid, and free of bias. The second pair of assessments, the grades 4 and 6 district CRTs, were also developed by Harcourt Assessment, Incorporated in collaboration with district content experts. Finally, we also used the norm-referenced Comprehensive Test of Basic Skills Fifth Edition (CTBS/5) TerraNova exam, which was administered in the fourth grade, in the analyses. The CTBS/5 Terra Nova, published by CTB Macmillan/McGraw-Hill, is a widely used and highly regarded achievement test. Prior analyses of national norming sample data indicated high item and scale reliabilities, absence of ceiling and floor effects, absence of obvious cultural biases, and a low error of measurement.

The grade-specific analyses of pretest-to-posttest outcomes for grades 4, 5, and 6 relied on the following assessments:

Grade 4 outcomes:

- Pretest: Grade 3 state CRT from spring 2002
- Posttest: Grade 4 district CRT from spring 2003

Grade 5 outcomes:

- Pretest: Grade 4 CTBS TerraNova from spring 2002
- Posttest: Grade 5 state CRT from spring 2003

Grade 6 outcomes:

- Pretest: Grade 5 state CRT from spring 2002
- Posttest: Grade 6 district CRT from spring 2003

Due to scaling differences, we standardized all test scores to have a common mean of 0 and standard deviation of 1.

Teachers. The evaluation system implemented in Washoe County School District

was modeled closely after the Framework for Teaching (Danielson, 1996). The district made relatively minor changes to the evaluation rubrics and designed the system to apply many of the suggested evidence sources to evaluation decisions. Teachers are evaluated by their principal or assistant principal. The sources of evidence that evaluators use can include teacher self-assessments, lesson and unit plans, classroom and nonclassroom observations with pre- and postobservation conferences, instructional materials (e.g., assignments and student work), reflection sheets, and logs of professional development and parental contact activities.

All teachers undergo an annual evaluation, but teachers are evaluated in different domains and on different elements depending on their stage in the evaluation cycle. There are three stages of evaluation under the system: probationary, postprobationary major, and postprobationary minor. Probationary (nontenured) teachers receive a comprehensive evaluation across all four performance domains: (1) planning and preparation, (2) classroom environment, (3) instruction, and (4) professional responsibilities. Probationary teachers are observed by the evaluator at least nine times over three periods of the year.

Once achieving postprobationary (tenured) status, teachers undergo a 3-year major-minor evaluation cycle. During the first year, they participate in a major evaluation in two performance domains and are formally observed by the evaluator three times during the school year. Over the next 2 years, the minor evaluations focus on one domain each year and involve at least one formal observation. Over this 3-year cycle, evaluators assess the postprobationary teachers in each of the four performance domains. Because of concerns by school board members that postprobationary teachers should be evaluated on their instruction each year, however, teachers who are not formally evaluated in the instruction domain are required to be evaluated on a

subset of key instruction-related standards. From this subset of standards, one can calculate composite scores that represent psychometrically sound summary measures of teachers' instructional performance.

Because teachers are evaluated in different domains depending on their stage in the 3-year major-minor evaluation cycle, maximizing the number of teachers who could be compared on the same teaching performance standards was a concern. To obtain the largest representation of teachers, the analyses used the composite measures of teacher instructional performance. The composite scores represent key elements from the planning and preparation domain and the instruction domain. Teachers are rated on each of the four composite scores as unsatisfactory (0), target for growth (1), proficient (2), and area of strength (3). We took the simple average of the four composite scores to obtain an overall measure of teacher quality. The composite measure is made up of the following standards:

- The teaching displays solid content knowledge and uses a repertoire of current pedagogical practices for the discipline being taught. (This standard includes 10 items from two performance domains.)
- The teaching is designed coherently, using a logical sequence, matching materials and resources appropriately, and using a well-defined structure for connecting the individual activities to the entire unit. Instruction links student assessment data to instructional planning and implementation. (This includes nine items from two performance domains.)
- The teaching provides for adjustments in planned lessons to match the students' needs more specifically. The teacher is persistent in using alternative approaches and strategies for students who are not initially successful. (This includes three elements from one domain.)
- The teaching engages students cognitively in activities and assignments, groups are productive, and strategies are congruent to instructional objec-

tives. (This includes three elements from one domain.)

The item intercorrelations for these composite scores ranged from .69 to .75, and the coefficient alpha reliability was .91. Teacher evaluation results, as measured by the overall composite, averaged about 2.63 on the 0–3-point scale.

Composite scores could be calculated for all probationary teachers. All postprobationary teachers also had composite scores, with the exception of one group. Specifically, postprobationary teachers who were evaluated in the instruction domain but who were not evaluated in the planning and preparation domain did not have composite scores. Because the teachers received an evaluation in the instruction domain, they were not required to be evaluated using the composite. However, without evaluation scores for the planning and preparation domain, on which the composite score also depends (along with scores from the instruction domain), composite scores could not be determined. This group represented approximately 19% of the teachers evaluated in Washoe County. Independent samples *t* tests, which compared this group of teachers with missing composite data to teachers who had complete data on the composite, revealed no statistically significant differences between the groups with respect to experience, the minority and free-lunch composition of their classes, and the baseline reading and math achievement of their students.

We measured teacher experience using each teacher's step position on the district teacher salary schedule. Because the district credits some qualified teaching experience in other states or districts, this measure takes into account relevant prior teaching experience. It should be noted, however, that with a maximum of 20 steps on the salary schedule, teachers can reach the top of the schedule and continue teaching for a number of years. Average experience based on the salary schedule ranged from 10.12

for fourth-grade teachers to 10.80 in the fifth grade.

Sample

Table 1 includes descriptive statistics for students and teachers in the sample. The table shows that over 2,500 students were included in the analyses for grades 4 and 6, and 2,176 students were included in grade 5. The student demographics were similar for each grade, with from 31% to 34% minority (Latino, African American, and Native American), and 23% to 25% eligible for free and reduced-price lunch. Latino students were the largest minority group with about 27% representation across the three grades.

As is often the case in education research, this sample of teachers and students had a nested structure. For instance, in the fourth-grade sample, there were 2,527 students nested within 131 classrooms (and teachers), for an average of 19.3 students per teacher. These 131 teachers and their classrooms were also nested within 55 schools, with an average of 2.4 teachers per school. For the fifth-grade sample, there were 135 teachers and 2,176 students within 43 schools, for an average of about 16 students per teacher and three teachers per school. The sixth-grade sample included 131 teachers with 2,632 students in 52 schools, with an average of between 2.5

teachers per school and 20.1 students per teacher.

Our analyses represented this nested structure through a multilevel modeling approach, which nested students within teachers' classrooms. Our two-level hierarchical linear models captured both student- and teacher-level sources of random variation. Another possible hierarchical model might have included schools as an additional source of variation. However, with so few teachers nested within schools, and with our focus on teacher-level outcomes, our exploratory analyses suggested that this strategy would have produced considerably less reliable and less informative results.

Results

Distribution of Teacher Quality

We began by examining how the teacher evaluation scores were distributed across classrooms of varying poverty and minority concentrations and varying reading and mathematics pretest scores. In all cases, we compared classrooms that were above the sample mean on poverty and minority concentration and pretest level to classrooms that were below the mean. A *t* test was used to evaluate the statistical significance of the differences between the two groups of classrooms on the dependent measure, the

TABLE 1. Student and Teacher Descriptive Statistics

Variables	Fourth Grade		Fifth Grade		Sixth Grade	
	Mean	SD	Mean	SD	Mean	SD
Students:	<i>N</i> = 2,527		<i>N</i> = 2,176		<i>N</i> = 2,632	
Math:						
Pretest	.03	.98	.03	.97	.04	.99
Posttest	.10	.94	.14	.93	.09	.99
Reading:						
Pretest	.02	1.00	.03	.99	.05	.98
Posttest	.09	.95	.16	.89	.08	.96
Minority	.31	.46	.34	.47	.31	.46
Free lunch	.23	.42	.24	.43	.25	.43
Teachers:	<i>N</i> = 131		<i>N</i> = 135		<i>N</i> = 131	
Evaluation score	2.63	.43	2.63	.48	2.62	.41
Experience	10.12	6.52	10.80	6.61	10.47	6.82

teacher evaluation score. In Table 2, we present the results from these comparisons.

The first comparison, by classroom poverty level, revealed that grade 4–6 classrooms with above average, or high, concentrations of poor children were taught by teachers with lower evaluation scores than classrooms with below average, or low, concentrations of poverty, $t(376) = 5.07$, $p < .001$ (two-tailed). This difference was equivalent to approximately half of a standard deviation on the teacher evaluation composite score.

The results for grade 4–6 classrooms with high and low concentrations of minority students are also presented. Classrooms with high concentrations of minority students were taught by teachers with lower evaluation scores than classrooms with low concentrations of minority children, $t(370) = 6.13$, $p < .001$ (two-tailed). The magnitude of this difference between classrooms with high and low minority concentrations was essentially the same as that found for classrooms of varying poverty levels: it was equivalent to about half of a standard deviation on the teacher evaluation composite score.

Finally, Table 2 shows results for classrooms with high and low reading and mathematics pretest scores. Classrooms composed of lower-achieving children on the pretest were taught by teachers with lower evaluation scores than classrooms with higher-achieving students, $t(367) = -5.78$, $p < .001$ (two-tailed) and $t(367) = -5.46$, $p < .001$ (two-tailed). Again, the magnitude of this difference between classrooms composed of students with higher and lower pretest math and reading scores was consistent. In both cases, it was equivalent to about half of a standard deviation on the teacher evaluation composite score.

Hierarchical Linear Model Analyses of Teacher Effects on Achievement

We used a two-level HLM to estimate teacher effects on classroom mean achievement and the within-classroom distributions of achievement. These two-level models were designed to examine the variation in student-specific achievement outcomes within classrooms and permitted analyses of achievement differences among students of different backgrounds as a consequence

TABLE 2. Average Teacher Evaluation Scores for Grade 4–6 Classrooms, by Poverty Level, Minority Concentration, and Reading and Math Pretest Scores

	N	Evaluation Score		t	df
		M	SD		
Poverty level: ^a					
Low	214	2.71	.39	5.07***	370
High	158	2.48	.48		
Minority concentration: ^b					
Low	206	2.74	.37	6.13***	370
High	166	2.47	.49		
Reading pretest scores: ^c					
Low	178	2.48	.48	-5.78***	367
High	191	2.74	.37		
Math pretest scores:					
Low	178	2.49	.49	-5.46***	367
High	191	2.74	.37		

^aLow-poverty classrooms were below the sample average of 24% free lunch, and high-poverty classrooms were at or above the sample average.

^bClassrooms with low minority concentrations were below the sample average of 34% minority, and high-minority classrooms were at or above the sample average.

^cClassrooms with low reading and math pretest scores were below the sample mean, and classrooms with high pretest scores were above the sample mean on the pretest measures.

*** $p < .001$.

of the classroom attributes, most importantly, teacher evaluation scores.

The one-way ANOVA with random effects served as a preliminary HLM model from which we computed estimates of the proportion of variation in the outcome that exists within and between classrooms. The level 1 model was specified as

$$Y_{ij} = \beta_{0j} + r_{ij},$$

where Y_{ij} is the achievement outcome for student i in classroom j , β_{0j} represents each classroom j 's mean achievement outcome, and r_{ij} is a student-specific random error term. At level 2, the one-way ANOVA with random effects model is

$$\beta_{0j} = \gamma_{00} + \mu_{0j},$$

where each classroom's mean achievement, β_{0j} , is represented by the function of the grand classroom mean, γ_{00} , plus random error, μ_{0j} .

The subsequent model introduced student-level predictors, including pretest score, a free-lunch status dummy code (1 = free-lunch recipient, 0 = non-free-lunch recipient), and a minority status dummy code (1 = African American, Hispanic, and Native American; 0 = Asian and white) to explain variation on the posttest. This level 1 model is written as

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{PRETEST})_{ij} + \beta_{2j}(\text{FREE LUNCH})_{ij} + \beta_{3j}(\text{MINORITY})_{ij} + r_{ij},$$

which represents the achievement for student i in classroom j regressed on the pretest score, free-lunch status, and minority status. The term r_{ij} is the level 1 residual variance that remains unexplained after accounting for the pretest, free-lunch status, and minority status.

The two-level model also assessed the relations between teachers' evaluation scores and experience and overall classroom mean achievement and the extent to which the within-school relations between

each of the three student-level predictors—pretest, free-lunch status, and minority status—and achievement varied across classrooms. For those within-classroom slopes that exhibited reliable variation across classrooms at level 2, we then began to examine the extent to which the teacher characteristics moderated these achievement differences. Specifically, the two-level models assessed the overall effects of teacher quality on classroom mean achievement and the potential equalizing effects of teacher quality in terms of the degree to which higher teacher evaluation scores closed the within-school gaps between minority and nonminority students, poor and more advantaged children, and students who scored higher and lower on the pretest. For example, such a level 2 model, with a classic aptitude-by-treatment interaction (ATI) effect of teacher quality, may be specified as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{EVALSCOR})_j + \gamma_{02}(\text{YRSEXP})_j + \gamma_{03}(\text{MEANPRET})_j + \mu_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{EVALSCOR})_j + \gamma_{12}(\text{YRSEXP})_j + \mu_{1j},$$

where the mean achievement intercept for classroom j , β_{0j} , is regressed on the classroom-level teacher quality rating, the teacher experience covariate, and the classroom-level pretest mean covariate, plus a residual, μ_{0j} . The β_{1j} represents the pretest slope, or the average relation in classroom j between students' pretest and posttest outcomes, which is regressed on the teacher quality rating and teacher experience variable. Respectively, these two level 2 formulas permitted us to examine the relation between teacher quality, as represented by the evaluation ratings, and overall classroom mean achievement and the degree to which teacher quality attenuated the relation between pretest and posttest scores. In other words, these analyses generated estimates of both the overall effects of teacher quality on classroom-level achievement and

the potential within-classroom equalizing effects of teacher quality.

Grade 4 mathematics. The first multi-level models, shown in Table 3, assessed student- and classroom-level effects on the mathematics posttest. The first set of columns in Table 3 presents the results for the fully specified model predicting grade 4 math achievement. In this model, the coefficient for the pretest suggested that each one-unit increase was associated with a 0.59-point increase on the posttest. The average within-classroom poverty gap, the difference between free-lunch and non-free-lunch students, on the posttest was -0.03 points. The average within-classroom minority gap on the posttest was -0.12 points. Three classroom-level effects—classroom mean achievement, the pretest slope, and the poverty gap—showed statistically significant random variation across classrooms, but the minority gap did not show statistically significant random variation. The pretest slope and the poverty gap were group-mean centered and modeled as random effects across classrooms, and the minority gap was treated as fixed at level 2 and grand-mean centered.

For the classroom mean achievement outcome, the evaluation composite and teachers' experience were not statistically significant predictors of between-classroom variation. The classroom-level mean pretest score, though, was a statistically significant predictor, indicating that classrooms with higher mean pretest scores also tended to achieve higher posttest scores. Finally, the teacher quality measure and teacher experience variable showed no statistically significant classroom-level relations to overall achievement, the pretest slope, or the poverty gap. In other words, the measure of teacher quality had no reliable relation to overall achievement and exhibited no within-classroom equalizing effects. The classroom-level model explained over 62% of the between-classroom variability on mean achievement but no variability in terms of the pretest slope or poverty gap.

Grade 5 mathematics. In the second group of columns in Table 3, we present the HLM results for the grade 5 mathematics sample. In this model, two classroom-level effects, classroom mean achievement and the pretest slope, exhibited statistically significant between-classroom random variation. The minority gap and poverty gap, though, did not vary across classrooms. The pretest slope was group-mean centered and modeled as a random classroom-level effect, and the minority and poverty gaps were treated as fixed at level 2 and grand-mean centered.

For the mean achievement outcome, the evaluation composite was a statistically significant predictor. After controlling for teacher experience and the classroom mean pretest score, and after adjusting the classroom mean achievement intercept for differences across classrooms in poverty and minority composition, each one-unit increase in the evaluation composite was associated with a .20-point increase in classroom mean achievement. Similar to the estimates derived from the previous model for grade 4, the intercept for the pretest slope suggested that each one-unit increase in the pretest was associated with a 0.69-point increase on the posttest. Neither the evaluation composite nor the teacher experience predictor accounted for classroom-to-classroom variability in closing the gap between higher and lower achievers. Therefore, though higher scores on the teacher evaluation composite were associated with higher classroom mean achievement outcomes, the measure of teacher quality was not associated with classroom-level differences in the equality of achievement outcomes.

Grade 6 mathematics. The final set of columns at the far right of Table 3 shows the outcomes for the grade 6 math sample. The level 2 model for the grade 6 sample revealed statistically significant between-classroom variation for the mean achievement and pretest slope outcomes, but the minority and poverty gaps did not exhibit

TABLE 3. Hierarchical Linear Models Predicting Mathematics Achievement, by Grade Level

	Grade 4			Grade 5			Grade 6		
	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t
Mean achievement:									
Intercept	.08	.03	2.79**	.10	.03	3.77***	.06	.04	1.81
Teacher experience	.00	.00	.81	-.00	.00	-.08	.00	.01	.00
Evaluation composite	.10	.06	1.54	.20	.06	3.41**	.02	.10	.19
Class mean pretest	.67	.06	11.11***	.65	.05	12.25***	.95	.08	11.86***
Pretest slope:									
Intercept	.59	.02	29.50***	.69	.02	37.99***	.57	.02	26.42***
Teacher experience	-.00	.00	-.70	-.00	.00	-1.07	-.00	.00	-.03
Evaluation composite	-.03	.04	-.81	.01	.04	.32	-.02	.06	-.33
Poverty gap:									
Intercept	-.03	.04	-.62	-.01	.03	-.45	-.04	.03	-1.24
Teacher experience	-.01	.01	-1.16						
Evaluation composite	.11	.08	1.30						
Minority gap:									
Intercept	-.12	.03	-3.57**	-.13	.03	-4.15***	-.13	.03	-4.99***
Teacher experience									
Evaluation composite									
Variation between classes:									
Mean achievement	.08	103	553.37***	.07	130	594.12***	.14	127	1209.81***
Pretest slope	.02	104	215.53***	.01	131	189.69**	.03	128	349.35***
Poverty gap	.04	104	142.08**						
Variance within classes	.38			.30			.35		
Variance explained:									
Mean achievement (%)	62.62			67.76			60.65		
Pretest slope (%)	0			0			0		
Poverty gap (%)	0								

* $p < .05$.

** $p < .01$.

*** $p < .001$.

statistically significant random variation. As in the grade 5 model, the pretest slope was group-mean centered and modeled as a random effect across classrooms, and the minority and pretest gaps were grand-mean centered and treated as fixed effects at level 2.

In this model, only the classroom mean pretest score accounted for statistically significant between-classroom variability for the mean achievement outcome. Though there was considerable random variation to model for both mean classroom achievement and the pretest slope, the teacher experience and teacher quality measures were not statistically significant predictors of either outcome.

Grade 4 reading. The first multilevel models predicting the reading outcomes are shown in Table 4. In the first set of columns in the table, the results for the fully specified model predicting grade 4 reading achievement appear. For the grade 4 reading outcome, three classroom-level effects—classroom mean achievement, the pretest slope, and the poverty gap—revealed statistically significant between-classroom random variation. The minority gap, though large in magnitude, did not show reliable variation across classrooms. The pretest slope and poverty gap were group-mean centered and modeled as random classroom-level effects, and the minority gap was treated as fixed at the classroom level and was grand-mean centered.

We first examined the classroom mean achievement outcome and found that the evaluation composite was a statistically significant predictor of classroom-to-classroom differences. After controlling for teacher experience and the classroom mean pretest score, and after adjusting the mean achievement intercept for differences across classrooms in their ethnic makeup, we found that the students of teachers with higher composite evaluation scores performed better on the posttest than students taught by teachers with lower evaluation scores. The

outcome for the pretest slope revealed no statistically significant differences associated with the teacher quality measure and classroom-to-classroom differences in the pretest-posttest relation. With regard to the poverty gap outcome, though, the gap between poor and nonpoor students was ameliorated by teachers with higher evaluation scores. Therefore, the measure of teacher quality was associated with higher overall classroom achievement and with reduced achievement gaps between students receiving free lunch and those not receiving it. The classroom-level model explained nearly 82% of the between-classroom variability on mean achievement and over 15% of the variability in the poverty gap.

Grade 5 reading. The second group of columns in Table 4 presents the results for the grade 5 reading sample. We modeled the classroom mean achievement and pretest slope outcomes as random classroom-level effects, but the poverty and minority gaps showed no statistically significant between-classroom random variation and were treated as fixed. The pretest slope was group-mean centered, and the poverty and minority gaps were grand-mean centered.

Only the classroom mean pretest was a statistically significant predictor of classroom mean achievement at posttest. The teacher evaluation composite and the teacher experience measure revealed no statistically significant associations with classroom mean achievement or the pretest slope. Thus, after we controlled for teacher experience and the classroom mean pretest score, and after adjusting the classroom mean achievement intercept for differences across classrooms in poverty and minority composition, the evaluation composite accounted for no classroom-to-classroom variability in mean achievement or in closing the gap between higher and lower achievers.

Grade 6 reading. The final multilevel analysis for grade 6 reading is presented in the far right set of columns in Table 4. Class-

TABLE 4. Hierarchical Linear Models Predicting Reading Achievement, by Grade

	Grade 4			Grade 5			Grade 6		
	Coefficient	SE	t	Coefficient	SE	t	Coefficient	SE	t
Mean achievement:									
Intercept	.07	.02	3.76***	.13	.02	7.64***	.06	.02	3.06**
Teacher experience	.00	.00	.10	.00	.00	.15	-.00	.00	-.38
Evaluation composite	.12	.05	2.27*	.04	.04	1.14	.09	.06	1.66
Class mean pretest	.73	.05	14.79***	.74	.04	20.55***	.91	.04	20.91***
Pretest slope:									
Intercept	.65	.02	34.81***	.66	.02	41.99***	.63	.02	29.37***
Teacher experience	-.00	.00	-0.73	.00	.00	.24	-.00	.00	-1.45
Evaluation composite	-.07	.04	-1.74	.03	.04	.74	.05	.05	.91
Poverty gap:									
Intercept	-.01	.03	-.33	-.08	.03	-3.01**	-.13	.04	-3.55**
Teacher experience	-.01	.01	-1.21						
Evaluation composite	.27	.09	3.06**						
Minority gap:									
Intercept	-.16	.03	-5.42***	-.05	.03	-1.71	-.18	.04	-5.13***
Teacher experience							-.01	.01	-2.49*
Evaluation composite							-.01	.09	-.11
Estimate									
	Estimate	df	χ^2	Estimate	df	χ^2	Estimate	df	χ^2
Variation between classes:									
Mean achievement	.03	103	313.17***	.02	130	314.63***	.03	113	291.84***
Pretest slope	.02	104	181.53***	.01	131	185.67**	.03	114	261.47***
Poverty gap	.02	104	118.27						
Minority gap							.03	114	148.39*
Variation within classes									
Variance explained:	.33			.27			.37		
Mean achievement (%)	81.68			85.49			85.68		
Pretest slope (%)	4.2			0			2.6		
Poverty gap (%)	15.31						16.78		
Minority gap (%)									

* $p < .05$.

** $p < .01$.

*** $p < .001$.

room mean achievement, the pretest slope, and the minority gap showed statistically significant random variation across classrooms. The poverty gap did not vary and was grand-mean centered and treated as fixed. The results showed a statistically significant relation between teacher experience and the minority gap. The outcome, which we did not anticipate, showed that in classrooms taught by teachers with greater experience, the minority gap was exacerbated. With respect to classroom mean achievement, only the classroom mean pretest showed a statistically significant and positive relation to the outcome. The teacher evaluation composite was not a statistically significant predictor of difference in classroom mean achievement, the pretest slopes, or the minority gaps.

Summary of results from the multilevel analyses. The results across the three grades and two subjects are summarized in Table 5. This table shows the outcomes associated with being a higher- versus lower-rated teacher on both overall classroom achievement and on closing the within-classroom gaps. The effect sizes were calculated based on the coefficients reported for the teacher evaluation score found in Tables 3 and 4. Each coefficient was multiplied by two times the standard deviation of the evaluation score composite and divided by the grade- and subject-specific student posttest score

standard deviation. In this way, the effect sizes represented the expected difference associated with a two-standard-deviation difference on the teacher evaluation composite score.

In other words, these differences would be what one would expect if one were to compare the classroom achievement outcomes of a “good” teacher scoring at one standard deviation above the average level—the eighty-fourth percentile—on the evaluation composite to a “bad” teacher at the sixteenth percentile of the evaluation score distribution. This comparison showed differences of between approximately one-tenth and one-fifth of one standard deviation on the achievement outcomes. That is, the classroom achievement average for a teacher at the eighty-fourth percentile of the evaluation score distribution was as much as a fifth of a standard deviation higher than the classroom mean for a teacher at the sixteenth percentile in the district. Although equalizing effects were generally close to zero, the results showed that fourth-grade teachers with higher evaluation scores made some progress in closing the achievement gaps separating poor and nonpoor children in reading and, to a lesser extent, in math.

Discussion

Using a standards-based teacher evaluation score as a measure of teacher quality, we

TABLE 5. Estimated Differences between “Good” and “Bad” Teachers on the Outcomes: Classroom Mean Achievement and Within-Classroom Slopes for Pretest, Poverty Status, and Minority Status

Outcome	Grade 4		Grade 5		Grade 6	
	Reading	Math	Reading	Math	Reading	Math
Classroom mean achievement	.11	.09	.04	.21	.08	.02
Pretest slope	-.06	-.03	.03	.01	.04	-.02
Poverty gap	.24	.10
Minority gap	-.01	...

NOTE.—Differences expressed as effect sizes. The effect sizes were calculated using the coefficients for the teacher evaluation score composite reported in Tables 3 and 4 multiplied by two times the standard deviation of the evaluation score composite and divided by the grade- and subject-specific student posttest standard deviation. This metric contrasts the outcomes for teachers at the sixteenth percentile of the evaluation score distribution (“bad” teachers) to teachers at the eighty-fourth percentile of the evaluation score distribution (“good” teachers), or the model-estimated difference between two teachers whose evaluation composite scores were two standard deviations apart.

find that teacher quality is not distributed equitably among classrooms with varying baseline achievement and poverty and minority concentrations. Students from poor, minority, and low-achieving backgrounds have access to teachers of lower quality, as reflected in the teachers' evaluation scores. Better teachers may be assigned, and seek out assignments, to classrooms with more advantaged, nonminority, and higher-achieving students. This inequality could represent an important mechanism that systematically constricts the educational opportunities offered to students from less advantaged backgrounds.

However, the teachers in these different classroom contexts may be of similar quality, but those teaching in less advantaged classrooms may be perceived by evaluators as less effective due to the attributes of the students they are teaching. Researchers such as Delpit (1995), Gordon and Yowell (1994), and Taylor (1991) have noted the academic risks associated with the potential discontinuities between the behavioral patterns and values socialized in the context of low-income and minority families and communities and those expected in mainstream classroom and school contexts. Low achievement and these nonmainstream behaviors of students from less advantaged classrooms could cause some evaluators to rate teachers lower than might be warranted. Similarly, teachers in schools with higher concentrations of more advantaged students may receive inflated ratings. Additional research is being conducted to explore variations in the validity of teacher evaluation ratings among evaluators in the district. Another year of data will also be analyzed to see if the results from this study are replicated.

Attributes of the school context, such as limited school organizational capacity or lack of a strong professional culture, also can constrain the performance of good teachers in high-poverty, high-minority, and low-achieving schools. Conceptual frameworks, including one articulated by Talbert and

McLaughlin (1999), and empirical findings of researchers, including D'Agostino (2000), clearly illustrate environmental effects of schools on teaching and the interactive effects between schools and teachers on student learning. These school effects on teachers and teaching and interactive effects between schools and teachers may represent other mechanisms through which the environment can alter the effectiveness, and evaluation scores, of teachers from varying contexts. More work needs to be done to understand the presence and cause or causes of these differences, because each of these interpretations has important implications for equality of educational opportunity, the evaluation system, and school organizational processes.

Better teaching appears to be related to better learning outcomes. The difference between "bad" and "good" teaching is equivalent to as much as one-fifth of a standard deviation difference in achievement. Generally, though, the differences are closer to one-tenth of a standard deviation. That is, a teacher at one standard deviation below the mean on the evaluation score distribution—the sixteenth percentile—and a teacher with an evaluation score of one standard deviation above the mean, at the eighty-fourth percentile, tend to have classroom achievement scores that are one-tenth of one standard deviation apart. This difference is after taking into account teacher experience and student pretest score, minority status, and free-lunch status. This outcome is a "main effect" that applies to all classrooms included in this study, regardless of the context.

With respect to the cross-level interaction effects, the results of our study are mixed. Teachers rated higher on the teacher evaluation system do not appear to be reducing gaps in achievement between low- and high-achieving students and students from low-income or minority backgrounds. In some instances, this finding is due to the fact that there was limited variability across classrooms to measure. Only two of six models

show reliable between-classroom variability for the poverty gap. In the fourth-grade reading example, higher-scoring teachers are closing the gap for free and reduced-price lunch students. Only one of six models demonstrates reliable between-classroom variability for the minority gap, but the effect of the teacher quality measure is not related to closing the gap. Finally, in every instance there is sufficient between-classroom variability for the pretest slope. In no instance, however, is teacher quality related to closing the gap between low and high achievers.

Implications

These results leave open the question of how one should define teacher quality if a central goal is reducing inequality. This analysis suggests that teacher quality, as defined and applied in the evaluation system of one school district, may not show reliable relations to closing achievement gaps between poor and more advantaged, minority and nonminority, and low- and high-achieving students. The implications for the evaluation system are important, especially if a key component of teacher quality is an ability to close achievement gaps.

The implications for larger national efforts to place more high-quality teachers in high-poverty Title I schools are also significant. These efforts also construe teacher quality as a broadly defined construct. Our results lend support to this idea to some extent, in that they show consistent achievement differences between classrooms that are taught by higher- and lower-quality teachers. However, our results also suggest that approximately 75% of the variability in achievement occurs within classrooms and about 25% occurs between classrooms. If equity and closing the within-classroom gaps are prominent goals, then the definition of a high-quality teacher may require some refinement.

This teacher evaluation system could be adjusted fairly easily to measure classroom effects on educational equality. We identi-

fied 22 rubrics across the four evaluation domains that appear to relate to the goal of reducing achievement gaps between different groups of students. We did not use these measures for the teacher quality variable in the current study because only a small number of teachers had evaluation scores on each measure. If the district were interested in further exploring teachers' efforts to achieve equality in their classrooms, a subset of these rubrics could be applied to all teachers, much like the four measures of the teacher evaluation performance composite are applied to teachers not evaluated in the instruction domain. The scores could then be substituted for the composite score used in this study. The results would not only help address the utility of the evaluation standards in predicting differential effects of teacher quality but also could be applied to district professional development focused on improving educational equality.

Finally, this leads to the question, Are there teacher preparation and professional development strategies that are likely to increase the effectiveness of teachers in achieving educational equality? The small amount of research that exists on the issue generally supports the idea that efforts to prepare teachers to teach in urban and diverse contexts can be beneficial. Field placement in an urban school, training in multicultural awareness, and effective recruitment and screening of teacher candidates are three efforts that have some research support. For instance, Cook and Van Cleaf (2000) and Stallings, Bossung, and Martin (1990) compared student teachers in an urban field placement with control groups of student teachers in nonurban placements and found that the urban placement was more helpful in enabling teachers to feel comfortable in their environments. This finding, along with research on alternative-route programs involving internships in urban schools, lends some support to the importance of field placements for preparing teachers to work in large-city settings. Two other studies suggested that

training in multicultural sensitivity was helpful to teacher candidates planning to teach in culturally diverse schools (Clewell & Villegas, 2001; Stoddart, 1990). Finally, though it is not a preparation or professional development plan, effective recruitment and selection strategies appear to increase the likelihood that teaching candidates who are accepted will be successful (Clewell & Villegas, 2001; Lutz & Hutton, 1989; Stoddart, 1990).

Conclusion

In these ways, our findings have implications for the broader construct of teacher quality and how it should be reconciled with prominent local and national efforts to promote educational equality. In addition, designers of teacher evaluation systems need to consider the goals of these efforts and whether measuring a teacher's ability to promote equality within the classroom is a prominent concern. Finally, the institutions that prepare teachers and the professional development programs that continue teachers' training might also consider featuring innovative mechanisms to enhance teachers' abilities to work in diverse settings and to achieve equality within their classrooms more routinely.

Note

The research reported in this article was supported in part by a grant from the U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Educational Governance, Finance, Policy-Making and Management, to the Consortium for Policy Research in Education (CPRE) and the Wisconsin Center for Education Research, School of Education, University of Wisconsin—Madison (grant no. OERI-R308A60003). The opinions expressed are ours and do not necessarily reflect the views of the organizations listed above. We wish to acknowledge the efforts and input of Tony Milanowski, who compiled the core database for this study and helped initiate this line of inquiry.

References

- Borman, G. D., & Rachuba, L. T. (1999). Qualifications and professional growth opportunities of teachers in high- and low-poverty elementary schools. *Journal of Negro Education, 68*, 366–381.
- Clewell, B. C., & Villegas, A. M. (2001). *Evaluation of the DeWitt Wallace–Reader's Digest Fund's Pathways to Teaching Careers Program*. Washington, DC: Urban Institute.
- 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: U.S. Government Printing Office.
- Cook, D. W., & Van Cleaf, D. (2000). Multicultural perceptions of first-year elementary teachers' urban, suburban, and rural student-teaching placements. *Urban Education, 35*, 165–174.
- D'Agostino, J. V. (2000). Instructional and school effects on students' longitudinal reading and mathematics achievements. *School Effectiveness and School Improvement, 11*, 197–235.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: New Press.
- Ferguson, R. F. (1998). Can schools narrow the black-white test score gap? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 318–374). Washington, DC: Brookings.
- Goldhaber, D., & Anthony, E. (2004, March). *Can teacher quality be effectively assessed?* Paper presented at the annual meeting of the American Educational Finance Association, Salt Lake City, UT.
- Gordon, E. W., & Yowell, C. (1994). Educational reforms for students at risk: Cultural dissonance as a risk factor in the development of students. In R. J. Rossi (Ed.), *Educational reforms and students at risk* (pp. 51–69). New York: Teachers College Press.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*, 84–117.
- Harcourt Assessment, Inc. (2004, March). *Nevada technical report for the CRT and HSPE assessments, 2002–2003 cycle*. San Antonio, TX: Author.

- Kain, J. F., & Singleton, K. (1996, May/June). Equality of educational opportunity revisited. *New England Economic Review*, pp. 87–111.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, *79*(4), 54–78.
- Krueger, A. B., & Whitmore, D. M. (2001). *Would smaller classes help close the black-white achievement gap?* (Working Paper No. 451). Princeton, NJ: Princeton University, Department of Economics—Industrial Relations Sections.
- Lutz, F. W., & Hutton, J. B. (1989). Alternative teacher certification: Its policy implications for classroom and personnel practice. *Educational Evaluation and Policy Analysis*, *11*, 237–254.
- Milanowski, A. (2004). The criterion-related validity of the performance assessment system in Cincinnati. *Peabody Journal of Education*, *79*(4), 33–53.
- Odden, A., Borman, G. D., & Fermanich, M. (2004). Assessing teacher, classroom, and school effects, including fiscal effects. *Peabody Journal of Education*, *79*(4), 4–32.
- Rowan, B. R., 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*, 1525–1567.
- Sanders, W., & Rivers, J. C. (1996, November). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Slavin, R. E., & Madden, N. A. (2001). *One million children: Success for all*. Thousand Oaks, CA: Corwin.
- Stallings, J. A., Bossung, J., & Martin, A. (1990). Houston Teaching Academy: Partnership in developing teachers. *Teaching and Teacher Education*, *6*, 355–365.
- Stoddart, T. (1990). Los Angeles Unified School District Intern Program: Recruiting and preparing teachers for an urban context. *Peabody Journal of Education*, *67*, 84–122.
- Talbert, J. E., & McLaughlin, M. W. (1999). Assessing the school environment: Embedded contexts and bottom-up research strategies. In American Psychological Association (Ed.), *Measuring environment across the life span: Emerging methods and concepts* (pp. 197–227). Washington DC: APA.
- Taylor, A. R. (1991). Social competence and the early school transition: Risk and protective factors for African-American children. *Education and Urban Society*, *24*, 15–26.