Observations of effective teacher-student interactions in secondary school
classrooms: predicting student achievement with the classroom assessment scoring
system – Secondary

Joseph Allen
University of Virginia

Anne Gregory
Rutgers University

Amori Mikami
University of British Columbia

Janetta Lun
University of Maryland

Bridget Hamre and Robert Pianta
University of Virginia

2013

RESEARCH INTO PRACTICE

Observations of Effective Teacher–Student Interactions in Secondary School Classrooms: Predicting Student Achievement With the Classroom Assessment Scoring System—Secondary

Joseph Allen
University of Virginia

Anne Gregory
Rutgers University

Amori Mikami
University of British Columbia

Janetta Lun
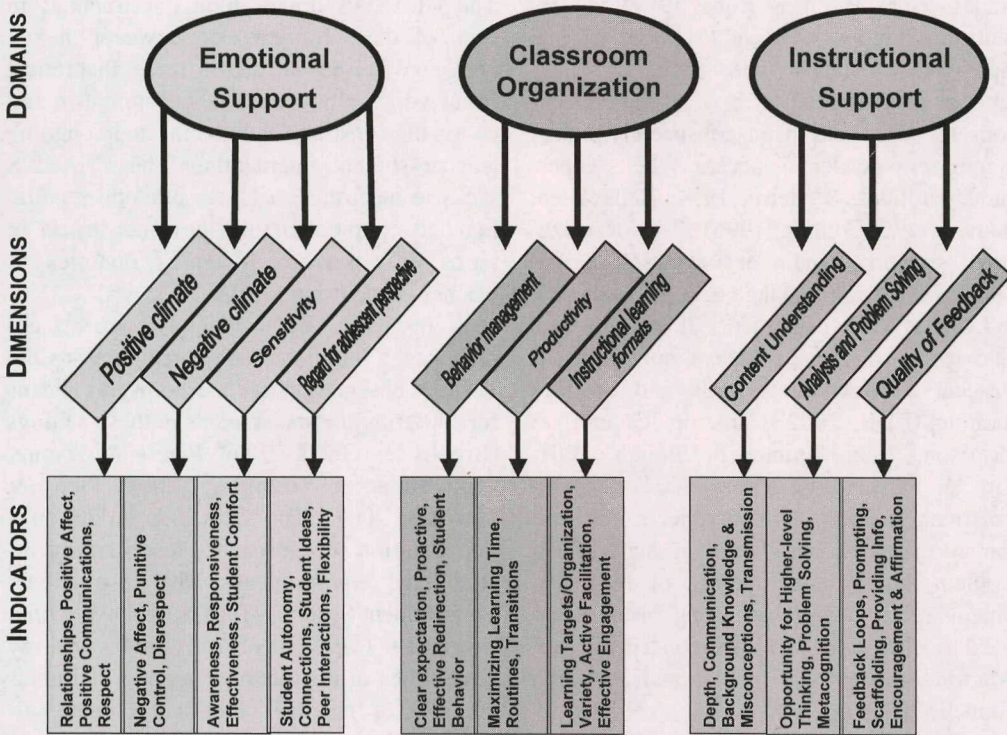*University of Maryland*

Bridget Hamre and Robert Pianta
University of Virginia

Abstract. Multilevel modeling techniques were used with a sample of 643 students enrolled in 37 secondary school classrooms to predict future student achievement (controlling for baseline achievement) from observed teacher interactions with students in the classroom, coded using the Classroom Assessment Scoring System—Secondary. After accounting for prior year test performance, qualities of teacher interactions with students predicted student performance on end-of-year standardized achievement tests. Classrooms characterized by a positive emotional climate, with sensitivity to adolescent needs and perspectives, use of diverse and engaging instructional learning formats, and a focus on analysis and problem solving were associated with higher levels of student achievement. Effects of higher quality teacher–student interactions were greatest in classrooms with fewer students. Implications for teacher performance assessment and teacher effects on achievement are discussed.

Correspondence regarding this article should be addressed to Joseph Allen, University of Virginia, Department of Psychology, Box 400400, Charlottesville, VA 22904-4400; e-mail: allen@virginia.edu

DOMAINS

Emotional Support

Classroom Organization

Instructional Support

DIMENSIONS

Positive climate

Negative climate

Sensitivity

Regard for adolescent perspective

Behavior management

Productivity

Instructional learning formats

Content Understanding

Analysis and Problem Solving

Quality of Feedback

INDICATORS

| Relationships, Positive Affect, Positive Communications, Respect | Negative Affect, Punitive Control, Disrespect | Awareness, Responsiveness, Effectiveness, Student Comfort | Student Autonomy, Connections, Student Ideas, Peer Interactions, Flexibility | Clear expectation, Proactive, Effective Redirection, Student Behavior | Maximizing Learning Time, Routines, Transitions | Learning Targets/Organization, Variety, Active Facilitation, Effective Engagement | Depth, Communication, Background Knowledge & Misconceptions, Transmission | Opportunity for Higher-level Thinking, Problem Solving, Metacognition | Feedback Loops, Prompting, Scaffolding, Providing Info, Encouragement & Affirmation |

Figure 1. Classroom Learning Assessment Scoring System—Secondary framework for assessing effective teacher–student Interactions

Improving the quality of teacher–student interactions within the classroom depends upon a solid understanding of the nature of effective teaching for adolescents. A number of descriptions of classroom environments or quality teaching have been put forth in the educational and developmental literatures listing factors likely to be related to student learning (e.g., Brophy, 1999; Eccles & Roeser, 1999; Pressley et al., 2003; Soar & Soar, 1979). Hamre and Pianta (Hamre & Pianta, 2010; Hamre, Pianta, Burchinal, & Downer, 2010) developed an assessment approach that organizes features of teacher–student interactions into three major domains: emotional supports, classroom organization, and instructional supports. This approach to assessing classroom interaction qualities has been tested and validated for the grades of prekindergarten to five, with evidence supporting this latent structure of dimensions and domains across

grades and across content areas (Hamre et al., 2010).

The Classroom Learning Assessment Scoring System—Secondary (CLASS-S; Pianta, Hamre, Hayes, Mintz, & LaParo, 2008) was developed for secondary schools as an upward extension of previous work. Within each domain considered are specific dimensions of classroom interactions that past research suggests are likely to be important to student learning (see Figure 1). The importance of qualities of emotional and relational support in the classroom is suggested by both attachment and self-determination theories (Bowlby, 1969/1982; Connell & Wellborn, 1991; Pianta, 1999), and is captured via assessments of the dimensions of positive classroom climate, teacher sensitivity, and regard for adolescent perspectives. Although the need for emotional support of students is perhaps more self-evidently important in the lower

grades (Ladd, Birch, & Buhs, 1999), adolescents are highly sensitive to the emotional rapport they establish with adults in school settings, and the experience of strong connections to adults has been consistently linked to long-term academic success (Allen, Kuperminc, Philliber, & Herre, 1994; Bell, Allen, Hauser, & O'Connor, 1996). The organizational support domain encompasses dimensions of classroom management, productivity, and use of varied instructional learning formats, which facilitate the development of adolescent self-regulation skills and enhance learning (Blair, 2002; Cameron, Connor, & Morrison, 2005; Emmer & Stough, 2001; Paris & Paris, 2001; Raver, 2004). The instructional support domain reflects teachers' content understanding, focus on analysis and problem solving, and quality of feedback, which are areas that have long been recognized to allow students to learn on a deep level (Marton & Saljo, 1976; National Research Council, 2005).

The domain approach used by the CLASS-S aligns well with constructs from several existing theoretical and practical approaches. For example, Brophy (1999) described 12 principles of effective teaching encompassing many of the same dimensions as the CLASS-S, including supportive classroom climate, opportunities to learn, curricular alignment, thoughtful discourse, scaffolding engagement, and achievement expectations. Similarly, Pressley and colleagues (2003) draw from their studies of effective teachers (e.g., Bogner, Raphael, & Pressley, 2002; Wharton-McDonald, Pressley, & Hampston, 1998) to suggest that effective teaching strategies can be organized into decisions regarding motivational atmosphere, classroom management, and curriculum and instruction. Eccles and Roeser (1999) suggest that schooling is optimally characterized by organizational, social, and instructional processes that help regulate children and adolescents' development across cognitive, social-emotional, and behavioral domains. A similar approach has been taken by McCaslin, Burross, and Good (2005) in their examination of classroom setting effects on student motivation.

The CLASS-S draws upon constructs from each of these frameworks; however, a key challenge raised by all of these theoretical frameworks is how best to operationalize and assess the constructs they delineate in ongoing teacher–student interactions. The CLASS-S seeks to meet this challenge by conceptualizing and operationalizing these constructs in terms of observable, ongoing qualities of teacher–student interactions.

In terms of actual assessment approaches, a long line of research suggests the value of observation in classrooms as a means for capturing the social assets in those settings (Brophy & Good, 1986; Pianta & Hamre, 2009; Shinn & Yoshikawa, 2008; Tseng & Seidman, 2007). The CLASS-S builds on a rich tradition of approaches to observing instructional environments. The Instructional Environment System—II (Ysseldyke & Christenson, 1987, 1993) approach outlines an array of qualities of instructional support and notes the role of positive emotional climate, although it has not been linked to achievement outcomes in typical classroom environments. Similarly, the Ecobehavioral Assessment Systems Software (Greenwood, Carta, & Dawson, 2000; Greenwood, Carta, Kamps, & Terry, 1994), initially developed for use in alternative school classrooms, addresses many similar constructs to the CLASS-S (Wallace, Anderson, Batholomay, & Hupp, 2002; Watson, Gable, & Greenwood, 2011). One key distinction, however, is that the CLASS-S focuses on broad patterns of interaction assessed at a molar level, as opposed to the time-sampling and counting of discrete behaviors as in the Ecobehavioral Assessment Systems Software. Thus, the CLASS-S is in keeping with principles of developmental psychology that suggest the importance of a focus on the broader organization of molar patterns of behavior as a means to get at subtle processes not easily captured via counts of discrete behaviors (Sroufe, Egeland, Carlson, & Collins, 2005), such as emerging self-regulation skills and resilience. At the younger grades, there is emerging evidence that a broad range of classroom interaction qualities can be observed and linked to student learning gains using this

more global yet standardized approach to assessment (Landry et al., 2006; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). Similarly, an intervention approach based on the CLASS-S framework has demonstrated evidence of efficacy along with evidence that program effects may be at least partially mediated via a single global measure of teacher–student interaction qualities from this system (Allen, Pianta, Gregory, Mikami, & Lun, 2011).

Unfortunately, outcome-based research on classroom observation as a predictor of actual student achievement is still relatively rare in secondary education. A modest amount of research beyond the elementary grades has successfully observed teaching in specific content areas or for specific lessons (Johnson, Kahle, & Fargo, 2007; Roth et al., 2006; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Seidel and Shavelson (2007) conducted a broader meta-analysis of 125 studies using a variety of methodologies to link teacher behavior to student achievement. They did not break out the tiny amount of observational research from nonobservational research at the secondary level, but did note that overall, assessments of teaching relied heavily upon distal and proxy variables (e.g., survey data) of questionable reliability and validity. They concluded that effects on secondary student academic performance from current approaches to measuring teaching quality were too small to be of practical significance (e.g., effect sizes ≤ 0.04). Thus despite the strong theoretical interest in identifying qualities of teacher–student interactions linked to student achievement, scientific evidence is quite sparse regarding our capacity to identify and observe the critical features of teacher–student interactions that actually predict student learning within the secondary school classroom. Virtually no evidence exists regarding the effectiveness of assessment systems designed to capture broad interactional patterns and apply across diverse content areas at the secondary level.

One significant potential confound in efforts to identify qualities of effective teacher–student interactions linked to student achievement is the likelihood that high-quality inter-

actions may come more easily among students who are already academically motivated and successful. Given the likelihood that students are to some degree tracked into higher and lower achieving groups in secondary schools (either explicitly or implicitly), different teachers are likely to face students with very different characteristics at the start of an academic year. End-of-year student test scores are highly dependent on preexisting student levels of academic proficiency and are typically highly correlated with prior year test scores. Failure to account for prior year test scores would thus misattribute variance in student achievement that would be more directly accounted for by background student proficiencies. In keeping with the growing recognition of the importance of "value-added" approaches to assessing student learning (Hanushek, Rivkin, Figlio, & Jacob, 2010; Rothstein, 2010), we assess end-of-year test scores after first accounting for prior year test scores, which we consider to be an indicator of student academic proficiency independent of the current classroom environment.

Purpose

Some qualities of teacher–student interactions may primarily reflect student characteristics as they enter class at the start of the year, particularly to the extent that students are implicitly or explicitly grouped into higher and lower achieving classes. Without awareness of this possibility, it would be all too easy to misattribute the qualities of classroom interactions to teacher skill levels, rather than recognizing that they may primarily reflect the academic characteristics of the students they are teaching. Thus, to provide the most helpful and balanced information to school personnel, this study also focused on identifying qualities of teacher–student interaction that are linked to preexisting student characteristics at the beginning of a school year, which are also qualities that might otherwise be misattributed to teacher skill. By identifying such student-driven qualities, this study seeks to provide appropriate contextual balance to our emerging picture of the role of teacher–student in-

teractions: showing not only where these interactions predict future achievement, but also where they may also simply reflect preexisting student characteristics, rather than simple teacher skill.

This study addressed three overarching goals:

1. Examine whether a broad, molar, observational measure of the quality of teacher–student interactions could predict achievement in classrooms at the secondary level across diverse content areas.

2. To the extent the first goal was met, examine the 10 individual dimensions of the CLASS-S to try to identify specific qualities of teacher–student interactions linked to future student achievement and to do so at a level of granularity sufficient to guide future research on specific teaching practices.

3. Provide a balanced picture to school personnel by assessing not only qualities of teacher–student interactions that predict student performance (after accounting for prior performance levels), but also to identify the extent to which certain qualities of teacher–student interactions appear largely determined by student academic skills upon *entry* into the teacher's classroom.

In pursuing these goals, we account for key contextual and background factors as both covariates and potential moderators (e.g., class size, and student gender, grade level, and family poverty status). Analyses also considered whether the predictive efficacy of the CLASS-S might be moderated by classroom content domain (e.g., math/science vs. English/social studies).

Method

Participants

Participants were 643 students enrolled in 37 classrooms (across 11 schools in six districts), which served as the control, teaching-as-usual condition classrooms in a larger study of an intervention to improve classroom

interaction qualities. The larger intervention study included 78 classrooms altogether and 1267 students in Year 1 and is further described in Allen, Pianta, Gregory, Mikami, & Lun (2011). Students were eligible for participation if they (a) were in a classroom participating in the control condition of the intervention study, (b) had parental consent, and (c) provided their own informed assent to participate. Teacher informed consent to participate in the study was also obtained. Each teacher was asked to identify one class that he or she considered "challenging" to teach, and only one classroom per teacher was assessed in the study. Classes averaged 23 students ($SD = 6.1$) in size and an average of 76.4% of students participated from each class. Classrooms were approximately equally divided between math or science courses ($N = 17$), and history, social studies, or English courses ($N = 20$). Table 1 presents descriptive information regarding both students and teachers; the table indicates that both groups consisted of racially and ethnically diverse samples, and that the participating teachers reflected a broad range of experience levels.

Measures

The CLASS-S was the primary source of standardized observational data on teacher-student interactions in the classroom. The pre-K and elementary versions of the CLASS are among the most current and widely used standardized assessments of social and instructional interactions in classrooms (Burchinal et al., 2008; Howes et al., 2008; McCaslin et al., 2005; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). The CLASS-S version was modified to capture precisely those aspects of classroom interactions that we hypothesize to be critical resources for educational achievement in adolescence. The CLASS-S consists of a set of global 7-point rating scales (one rating scale for each dimension below) with behaviorally anchored scale points providing detailed descriptions of a specific dimension of classroom process and its scaling from low to high. The CLASS-S scales are organized into three overarching domains,

Table 1
Means and Standard Deviations of Primary Measures and
Demographic Variables

		Mean	SD
Student Grade in School		8.3	1.5
Teacher Age		41.2	10.6
Teacher Education Level		5.4	1.2
Number of Years Teaching		10.8	10.7
Class Size		22.8	6.1
Prior Year Achievement Test Score		461.6	75.85
End-of-year Achievement Test Score		456.9	74.15
			N
Student Gender	Male		348
	Female		295
Student Race/Ethnicity	Asian		7
	African-American		154
	Hispanic		26
	White		456
Students' Family Poverty Status	Low-income		200
	Non-low-income		420
Teacher Gender	Male		17
	Female		20
Teacher Race/Ethnicity	African-American		1
	White		32
	Pacific Islander		1
	Multi-racial		3
Teacher Education Level	Bachelor's Degree		11
	Post Bachelor's Training		26

similar to those reported in factor analyses of the elementary version:

Emotional Support. The Emotional Support domain is composited from subscales for Positive Climate, reflecting warmth and sense of connectedness in classroom; Negative Climate, reflecting expressed negativity in classroom; Teacher Sensitivity, reflecting responsiveness to student academic/emotional needs; and Regard for Adolescent Perspectives, reflecting the teacher's ability to recognize and capitalize on student needs for autonomy, active roles, and peer interaction in the classroom.

Classroom Organization. The Classroom Organization domain is composited from subscales for Behavior Management, re-flecting teacher ability to use effective methods to encourage desirable behavior and prevent/redirect misbehavior; Productivity, reflecting teacher ability to manage the classroom so as to maximize instructional time; and Instructional Learning Formats, reflecting teacher use of varied and interesting materials and teaching techniques in an organized fashion.

Instructional Support. The third domain is Instructional Support, which is composited from subscales for Content Understanding, reflecting teacher presentation of content within a broader intellectual framework; Analysis and Problem Solving, reflecting emphasis upon engaging students in higher order thinking skills; and Quality of Feedback,

reflecting provision of contingent feedback designed to challenge students and expand their understanding of a concept.

Student achievement. Student academic achievement was assessed using the Standards of Learning (SOL; Commonwealth of Virginia, 2005). The SOL is the state-mandated accountability measure for the Commonwealth of Virginia that is administered to meet the requirements of No Child Left Behind. It was first used in 1998, with a 7-year period for schools to align their curricula and adjust to the testing requirement before the school accreditation process began (Commonwealth of Virginia, 2005). The SOL program has now been in place for almost a decade, making it one of the oldest such programs in the nation. Students take SOL tests (which consist of between 45 and 63 multiple-choice questions, depending upon the specific test used) at the end of the course in core subjects taught by their teacher, and each is standardized on a 200–600 point scale, with 400 defined as a passing score with real-world implications both for student graduation and school accreditation. In terms of validity, the subject tests have strong unidimensionality and correlate between .50 and .80 with the Stanford 9 achievement tests (Hambleton et al., 2000). In terms of reliability, high school assessments had KR-20 coefficients of .87 and .91 (Hambleton et al., 2000).

External reviewers have also reviewed the tests and found that the "reliability evidence for the SOL assessments is solid and typical of high quality assessments" (Hambleton et al., 2000, p. 8). Although tests use the same scales across subject matter, student scores for a given test were adjusted using statewide normative data for each test to assure that scores could be fully equated across different tests. We obtained each students' baseline SOL score, which was taken from a similar course the year before and each students' end-of-year SOL score, which was directly linked to the instructional content of the classrooms under examination (i.e., math courses were paired with prior year math courses, social studies courses were paired

with prior year social studies courses, and so on).

Student, teacher, and classroom characteristics. School records were used to identify students' gender, race/ethnicity, and grade level. Records also indicated whether students came from low-income families (coded based on student eligibility for free and reduced-price lunch, which extends to families with incomes up to 185% of federal poverty line). Teachers reported on their gender, race/ethnicity, years of experience teaching, and education level. Class size was obtained from teachers' enrollment rosters.

Procedures

Student SOL data were collected via standardized end of course assessments both for the study year and for the year prior to the study. Classroom observation data were collected within a limited window between the 4th and 8th weeks of the school year. Teachers were asked to record a typical classroom session and coding was performed based on video-recording of this classroom interaction for a teacher, with the camera positioned to capture both the teacher as well as a significant number of students. Teachers were told to record a class session in which they were actually teaching (as opposed to giving a test, watching a video, and so on). The window of observation at the beginning of the school year was deliberately selected to minimize the degree to which qualities of student–teacher interactions might be influenced by ongoing student engagement and achievement (as opposed to predicting such achievement).

A team of advanced undergraduate and graduate student coders were trained in a two-day workshop on the CLASS-S system. Coders learned to rate each of the ten specific CLASS-S dimensions along a 1–7 scale, with a 1 or 2 indicating low quality; 3, 4, or 5 indicating midrange quality; and 6 or 7 indicating high quality. At the end of the workshop, each coder passed a reliability test, in which they scored within one point of the master-coded tapes on 80% of scores, across five video segments. The master coders had

extensive knowledge of the CLASS-S instrument. In addition, team members met regularly during the year to jointly code master tapes in order to prevent drift and increase coding agreement. The recorded classroom session was divided into two 20-min segments and each segment was coded independently by two trained coders, with a different pair of coders used for each of the two segments.

Scores were averaged across raters. Reliabilities between the two raters for domains, assessed via intraclass correlations, using Cicchetti's (1981) norms for interpreting intraclass correlation coefficients, ranged from good (.73 for Instructional Support) to excellent (.77 and .82, for Emotional Support and Classroom Organization, respectively). For the specific dimensions, with the exception of Negative Climate (intraclass correlation coefficient [ICC] = .50—in the "fair" range), all other dimensions also had ICC's in the good to excellent range (ranging from .64 to .78).

Analytic Strategy

This study used a nested design that included multiple students within each classroom. Initial analyses revealed a significant ICC at the classroom level and related design effect (.52 and 9.43, respectively). Classrooms were also nested within schools, although numbers of schools and classrooms within them were too small to permit adequate modeling of school level effects. Hierarchical linear modeling (Raudenbush & Bryk, 2002) was thus used as the conceptual and analytic framework for specifying two-level models that examined the association between measures of classroom quality and individual-level child outcomes (end-of-year achievement test scores), after accounting for student grade level, gender, family poverty status, and classroom size. PROC Mixed in SAS, using restricted maximum likelihood estimation, was used to specify the models derived from the following equations (Singer, 1998). In the first level of the two-level model (Equation 1), an end-of-year test score (Y) for a student (i) who is in classroom (j) is a function of the mean post-test score for students in this class (β_{0j})

after adjusting for pretest scores (β_{01}) and demographic characteristics of students (β_{02-4}), and the error term associated with this estimated mean (r_{ij}).

$$Y_{ij} = \beta_{0j} + \beta_{01} \text{ (pretest)}$$

$$+ \beta_{02} \text{ (student gender)}$$

$$+ \beta_{03} \text{ (student grade level)}$$

$$+ \beta_{04} \text{ (student poverty status)} + r_{ij} \quad (1)$$

Equation 2 specifies in the second-level model that the adjusted mean post-test score for students in each classroom (β_{0j}) is a function of the grand mean post-test score (γ_{00}), classroom interaction quality (γ_{01}), and class size (γ_{02}) and the error term associated with this estimated mean (u_{0j}).

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{ (quality)}$$

$$+ \gamma_{02} \text{ (class size)} + u_{0j} \quad (2)$$

Because all Level 1 coefficients for the controls were fixed, the only substitution is for β_{0j}, resulting in Equation 3:

$$Y_{ij} = \gamma_{0j} + \gamma_{01} \text{ (quality)} + \gamma_{02} \text{ (class size)}$$

$$+ \beta_{01} \text{ (pretest)} + \beta_{02} \text{ (student gender)}$$

$$+ \beta_{03} \text{ (student grade level)}$$

$$+ \beta_{04} \text{ (student poverty status)}$$

$$+ r_{ij} + u_{0j} \quad (3)$$

All Level 1 coefficients aside from β_{0j} were fixed, meaning that they were not allowed to vary across classes. For ease of interpretation, the outcome variable was standardized at the grand mean for all analyses.

We conducted three sets of primary analyses. The first examined whether each global domain of teacher–student interaction quality predicted student achievement, after first taking into account student-level and classroom-level demographic characteristics.

Second, we conducted a fine-grained analysis of each dimension within each global domain of teacher–student quality as earlier, entering student and classroom demographic factors first in the model. This identified which specific dimensions were significant predictors of achievement. Finally, we ran similar analyses to assess the extent to which certain observed qualities of student–teacher interactions were actually predictable from students' prior levels of achievement.

For each set of these primary analyses, we also tested to see whether the findings obtained might be moderated by classroom or student characteristics. Equations 4 and 5 provide examples of the models used for testing moderation of the relation of CLASS-S scores to achievement via variables at the classroom level (Equation 4) and at the student level (Equation 5). Equation 4 specifies adding an interaction term to the second level model so that the adjusted mean post-test score for students in each classroom (β_{0j}) is a function of the grand mean post-test score (γ_{00}), classroom interaction quality (γ_{01}), class size (γ_{02}), the interaction between both Level 2 main effects (γ_{03}), and the error term associated with this estimated mean (u_{0j}).

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{ (quality)} + \gamma_{02} \text{ (class size)}$$

$$+ \gamma_{03} \text{ (quality} \times \text{class size)} + u_{0j} \quad (4)$$

Equation 5 instead specifies a cross-level interaction between student poverty status (as one example) and teacher–student interaction quality by combining Equation 1 and Level 2 equations, where $\beta_{0j} = \gamma_{00} + \gamma_{01}$ (quality) $+ u_{0j}$ and $\beta_{1j} = \gamma_{10} + \gamma_{11}$ (quality) $+ u_{1j}$ to yield the following:

$$Y_{ij} = \gamma_{00} + \gamma_{10} \text{ (pretest)}$$

$$+ \gamma_{20} \text{ (student gender)}$$

$$+ \gamma_{30} \text{ (student grade level)}$$

$$+ \gamma_{40} \text{ (student poverty status)} + \gamma_{01} \text{ (quality)}$$

$$+ \gamma_{11} \text{ (student poverty status} \times \text{quality)}$$

$$+ (u_{0j} + u_{0j} \text{ (poverty status)} + r_{ij}) \quad (5)$$

Results

Preliminary Analyses

Means and standard deviations for all primary variables of interest in the study are reported in Table 1. Table 2 presents a matrix of the bivariate correlations between the three overarching domains of Emotional Support, Classroom Organization, and Instructional Support along with the specific dimensional scales for measures of teacher–student interaction quality. Confirmatory factor analysis, using maximum likelihood analysis, supported the loading of scales onto the domains as presented earlier, with all factors loading above 0.4 on the relevant factor ($\chi^2 = 39.60$, $p = .06$, confirmatory factor analysis $= 0.97$, root mean square error of approximation $= 0.07$; exact factor loadings are available from the authors upon request). As expected, prior year achievement test scores were substantially correlated with current year achievement test scores ($r = .69$, $p < .001$). All analyses also accounted for a range of potentially confounding variables: contextual factors, including student grade level, gender, and family poverty status, and classroom size. Potential moderating effects of these contextual factors were also examined (via interaction terms created from centered versions of these variables) and yielded no significant interactions, with the exception of a moderating effect of classroom size, discussed further later.

Qualities of Teacher–Student Classroom Interactions to Predictors Student Achievement

Table 3 presents results of three hierarchical linear modeling models, one for each of the three overarching domains assessing teacher–student interaction quality. In these models, end-of-year student achievement test scores are predicted from baseline scores, from student grade in school, gender, and fam-

Table 2
Descriptive Statistics and Correlations Among Teacher-Student Interaction Quality Rating Scales

	Mean	SD	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
Overarching Domains:														
1. Emotional Support	4.7	.56	52*	79*	83*	−58*	85*	85*	33*	41*	59*	45*	72*	72*
2. Classroom Organization	5.0	.61	—	71*	39*	−59*	47*	34*	86*	91*	77*	53*	41*	56
3. Instructional Support	3.8	.63		—	61*	−47*	72*	67*	43*	60*	78*	70*	81*	86*
Dimensional Scales:														
4. Positive Climate	4.2	.74			—	−41*	54*	61*	22	25	53*	30	53*	55*
5. Negative Climate	1.2	.35				—	−51*	−29	−57*	−45*	−47*	−28	−34*	−40*
6. Teacher Sensitivity	4.7	.81					—	61*	34*	39*	49*	43*	60*	73*
7. Regard for Adolescent Perspectives	3.4	.88						—	13	32	43*	40*	72*	55*
8. Behavior Management	5.4	.76							—	77*	42*	40*	18	40*
9. Productivity	5.3	.67								—	57*	50*	37*	51*
10. Instructional Learning Formats	4.4	.71									—	47*	51*	53*
11. Content Understanding	3.8	.73										—	32	50*
12. Analysis & Problem Solving	3.2	.92											—	65*
13. Quality of Feedback	3.9	.82												—

Note. * $p \leq .05$. All correlations are multiplied by 100; correlations between overarching domains and individual teacher-student interaction quality scales on a given domain on which it loads are underlined. $N = 37$ classrooms.

Table 3
Predicting End-of-year Achievement from Teacher-Student Interaction Quality (Covarying Prior Year Achievement and Demographic Factors)

| | Predicting End-of-year Achievement Scores From: | | | | | |
| | Emotional Support | | Classroom Organization | | Instructional Support | |
	B	SE	B	SE	B	SE
Intercept	31.30***	10.63	359.8***	59.42	372.6***	46.64
Baseline Achievement Score	.61***	.029	.61***	.029	.61***	.029
Student Characteristics						
Student Grade-Level	5.40	3.60	4.50	3.73	6.33	3.77
Student Gender	1.90	3.39	2.05	3.39	1.95	3.39
Student Family Poverty Status	−3.31	4.32	−3.23	4.33	−3.00	4.33
Classroom Size	−0.06	1.03	−.047	1.07	−0.56	1.04
Interaction Quality Variable	31.30**	10.63	21.65*	10.35	25.58**	10.13

*** $p < .001$. ** $p \leq .01$. * $p < .05$.

ily poverty status (at the child level), and from classroom size (at the classroom level), followed by the classroom interaction quality variable of interest. The domains of teacher–student interactions overlapped sufficiently that when analyses examined all three simultaneously as predictors, it was not possible to identify significant unique predictive variance from any single domain, after accounting for the others. Thus analyses were conducted separately by domain.

Results indicate that each of the three domains of teacher–student interaction were predictive of higher student achievement test scores at the end of the year, even after covarying prior-year scores and relevant student and classroom characteristics. Using the approach suggested by Peugh (2010), we examined the proportional reduction in classroom level variance obtained by using this composite, relative to the classroom variance remaining after considering the effects of student grade, gender, and family poverty status, and classroom size. Results indicated a proportional reduction in classroom-level variance in student outcomes of 12.8%, 5.3%, and 8.9%, respectively, from the single assessments of emotional, organizational, and instructional support. For ease of general interpretation, we examined effects not simply in terms of raw score equivalents, but in percentile terms (examining the change in percentile terms in test scores associated with a one standard deviation increment in an observed domain score). The magnitude of the strongest prediction, from the Emotional Support domain, indicates that, after accounting for other measured variables, a student entering with average prior test scores (i.e., 50th percentile) in a class that was one standard deviation below the mean in Emotional Support would on average place in the 41st percentile in end-of-year tests; whereas an average student with the same background characteristics in a class that was one standard deviation above the mean in Emotional Support would on average place in the 59th percentile in end-of-year tests.

Class size interacted with Emotional Support ($B = -4.81$, $SE = 2.00$, $p = .02$) and with Instructional Support ($B = -3.54$, $SE = 1.78$, $p = .046$). These interactions are depicted in Figure 2 using standardized scores for all variables and plotting lines for classrooms that were one standard deviation above and below the mean in class size (reflecting class sizes of approximately 29 and 17, respectively). Figure 2 shows that measured Emotional and Instructional Support in the

Interaction of Class Size and Emotional Support in Classroom Predicting Relative Achievement Gains Over Time
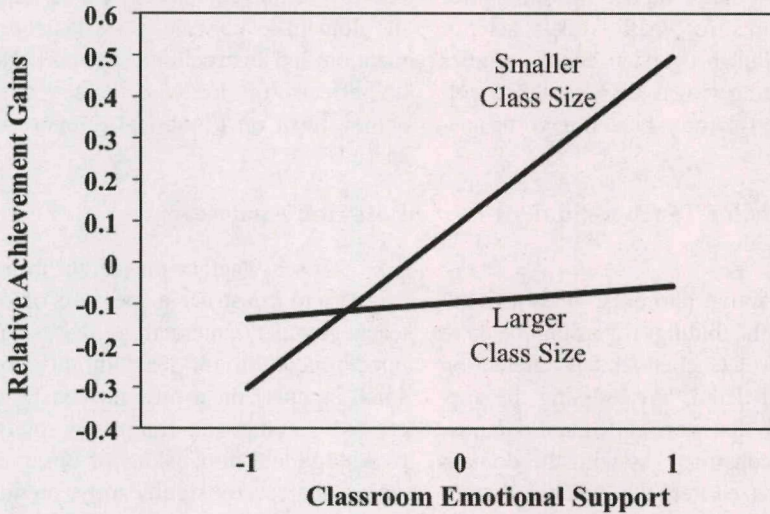
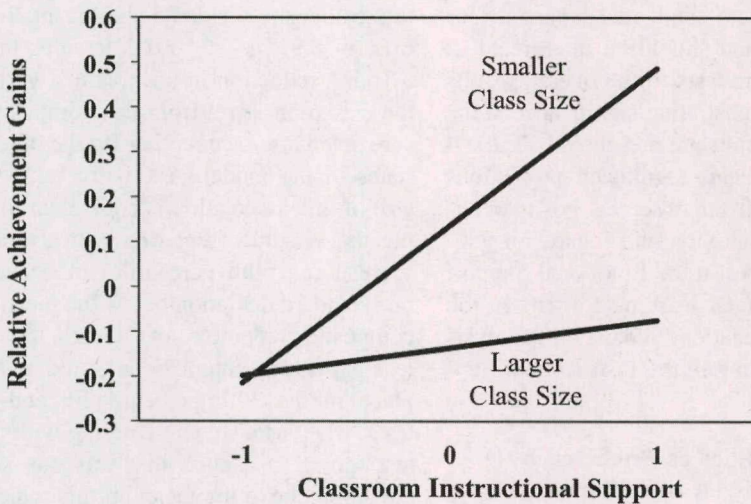## Interaction of Class Size and Instructional Support in Classroom Predicting Relative Achievement Gains Over Time

Figure 2. Interaction of class size and emotional support in classroom predicting future student achievement (controlling for baseline achievement)

classroom was of greatest predictive value for student academic achievement in smaller as compared to larger classrooms.

No moderator effects of course content area were found, which indicates that relations between past and current test scores were not

significantly different for math/science versus English/social studies courses and that results did not significantly differ depending on the type of course being taught. (There were also no differences in the predictive power of prior years' test scores for math versus science courses or English versus social studies courses, indicating it was appropriate to collapse into these two broad categories for analytic purposes).

Specific Effects for Teacher–Student Interactions

For descriptive purposes, analyses next followed up on the findings regarding the three global domains of teacher–student interaction in finer grained detail, by assessing the specific dimensions that were coded and composited into each construct. As with the domain scores discussed earlier, the specific dimensions overlapped sufficiently that when analyses examined them simultaneously as predictors, it was not possible to identify significant unique predictive variance from any single dimension, after accounting for the others. These analyses were conducted separately for each dimension and should be interpreted as post hoc follow-up tests to the overall results noted earlier demonstrating significance at the domain level. Results are presented in Table 4 These results indicate significant predictions of achievement from observed positive climate, teacher sensitivity, and regard for adolescent perspectives in the Emotional Support domain, instructional learning formats in the Classroom Organization domain, and analysis and problem solving in the Instructional Support domain.

Classroom Qualities as Predicted by Prior Student Levels of Achievement

Finally, analyses sought to determine the extent to which measures of observed teacher–student interactions could be predicted from levels of student achievement prior to entering the class. Such predictions can identify which aspects of teacher–student interaction may be influenced heavily by student entry characteristics. These results are presented in Table 5. Higher quality behavior management, instructional learning formats, content understanding, and quality of feedback were significantly predicted by students' baseline achievement test scores, along with the domain-level scales for Classroom Organization and Instructional Support. However, no effects of baseline levels of student achievement on Emotional Support were observed.

Post Hoc Analyses

To see whether one might utilize these findings to construct a measure of observed teacher–student interactions that would optimize prediction (at least in this somewhat small sample) on a post hoc basis, we next created a composite comprised solely of the five individual dimensions of observed interaction that were significantly predictive of end-of-year test scores: positive climate, teacher sensitivity, regard for adolescent perspectives, instructional learning formats, and analysis and problem solving. When considered in models identical to those in Table 3, this composite yielded a significant B of 30.2 ($SE = 8.91$, $p < .001$). Results indicated a 16.3% reduction in unexplained variance at the classroom level from this composite measure, even after accounting for the other covariates in the model. Alternatively, the effect size of this association is equivalent to a finding that a student entering with average test scores (i.e., 50th percentile) in a classroom one standard deviation below the mean on this composite reflective of overall quality of teacher–student interactions would on average place in the 37th percentile in end-of-year tests; whereas a student entering with average test scores in a class that was one standard deviation above the mean on this scale would on average place in the 63rd percentile on end-of-year tests.

Discussion

This study identified specific features of teacher–student interactions in the classroom, observed using standardized measurements, that were directly linked to student achieve-

Table 4
Relation of Dimensional Teacher-Student Quality Interaction Scales to End-of-year Test Scores (Covarying Prior Year Achievement and Demographic Factors)

Predicting End-of-year Achievement Scores From:

	Emotional Support								Classroom Organization						Instructional Support					
	Positive Climate		Negative Climate		Teacher Sensitivity		Regard for Adolescent Perspective		Behavior Management		Productivity		Instructional Learning Formats		Content Understanding		Analysis & Problem Solving		Quality of Feedback	
	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)	B	(SE)
Intercept	363.7***	(44.37)	480.8***	(32.24)	389.8***	(47.39)	400.6***	(36.29)	438.1***	(55.70)	372.8***	(59.10)	367.9***	(42.73)	423.3***	(41.19)	417.0***	(35.54)	431.7***	(45.38)
Baseline Achievement Score	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)	.61***	(.03)
Student Grade-Level	4.78	(3.58)	4.13	(3.89)	5.35	(3.78)	4.67	(3.6)	3.93	(3.87)	3.88	(3.76)	6.17	(3.65)	3.98	(3.79)	5.97	(3.82)	5.30	(4.03)
Student Gender (1 = M; 2 = F)	2.02	(3.39)	1.98	(3.39)	1.88	(3.39)	1.83	(3.39)	1.98	(3.39)	2.01	(3.39)	2.14	(3.39)	1.94	(3.39)	1.92	(3.39)	1.93	(3.39)
Student Family Poverty Status	−2.97	(4.33)	−3.37	(4.34)	−3.45	(4.33)	−3.33	(4.32)	−3.27	(4.34)	−3.15	(4.33)	−3.52	(4.32)	−3.01	(4.34)	−3.12	(4.33)	−3.17	(4.34)
Classroom size	−0.09	(1.03)	−.48	(1.15)	−0.37	(1.15)	−0.19	(1.04)	−.51	(1.13)	−0.34	(1.09)	−.85	(1.02)	−0.84	(1.11)	−0.29	(1.07)	−0.45	(1.12)
Interaction Quality Variable	23.1**	(7.89)	−9.25	(18.75)	16.36***	(7.95)	18.03**	(6.73)	5.89	(8.53)	17.68	(9.47)	24.94**	(8.41)	14.11	(9.27)	15.19*	(6.89)	9.32	(8.55)

Note. *** $p < .001$. ** $p \leq .01$. * $p < .05$. 'Interaction Quality Variable' results for each column refer to the specific classroom interaction quality noted in the column heading.

Table 5

Relation of Dimensional Teacher-Student Interaction Quality Scales to Prior Year Test Scores

Relation to *Prior* Year Achievement Scores:

	Emotional Support					Classroom Organization				Instructional Support			
	Overall Domain Score	Positive Climate	Negative Climate	Teacher Sensitivity	Regard for Adolescent Perspective	Overall Domain Score	Behavior Management	Productivity	Instructional Learning Formats	Overall Domain Score	Content Understanding	Analysis & Problem Solving	Quality of Feedback
	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
Intercept	-87.97 (46.90)	-94.59 (53.45)	23.49 (3.43)	-95.45 (55.52)	-49.75 (38.34)	-162.9* (77.50)	-141.5* (66.23)	-77.96 (76.85)	-128.5* (57.66)	-167.6** (54.47)	-163.55*** (44.77)	-48.87 (35.26)	-124.88** (44.73)
Student Grade-Level	4.01 (5.19)	3.56 (5.18)	3.43 (5.29)	3.82 (5.21)	3.81 (5.26)	3.51 (5.11)	2.92 (5.09)	2.74 (5.26)	4.27 (5.13)	5.91 (4.97)	1.34 (4.77)	4.27 (5.43)	6.53 (5.12)
Interaction Quality Variable	27.68 (16.62)	19.93 (12.61)	-28.64 (27.38)	17.99 (11.71)	11.13 (10.81)	30.07* (15.27)	23.93* (12.06)	12.56 (14.41)	26.66* (12.96)	40.89** (14.05)	39.98*** (11.54)	11.83 (10.74)	28.88** (11.15)

Note. *** $p < .001$. ** $p \leq .01$. * $p < .05$. 'Interaction Quality Variable' results for each column refer to the specific classroom interaction quality noted in the column heading.

ment over the course of an academic year, even after accounting for prior levels of student achievement, demographic characteristics, and classroom size. Moreover, these findings applied regardless of student grade levels and content area of instruction; in short, "good teaching" was good regardless of content or grade level. These results have implications not only for theories regarding the role and value of interactions and proximal processes in social settings serving youth (Tseng & Seidman, 2007), but also for contemporary policies and practices related to the assessment and improvement of teacher performance (Weisberg, Sexton, Mulhern, & Keeling, 2009).

These findings are consistent with theoretically driven expectations regarding the mechanisms through which secondary school instruction affects adolescents' learning, as operationalized via the CLASS-S observations. These results suggest that the "value-added" of classroom settings may in part be attributable to qualities of teacher–student interactions. These results also extend similar findings obtained in prekindergarten and elementary school settings, in which observational approaches using the elementary version of the CLASS have been used to predict measures of student learning (e.g., Hamre & Pianta, 2005; Rimm-Kaufman et al., 2009). Collectively, these studies provide empirical support for the argument that critical teacher behaviors can be measured in a standardized observational assessment.

Although studies of student achievement have been important in laying a foundation for inquiry into classroom effects (Ladd, 2008; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005), they had not yet succeeded in identifying specific processes that may lead to student learning and positive social adjustment across an array of content areas. As a result, the field has thus far been left to rely largely upon an atheoretical, post hoc approach as reflected in Hanushek's (2002) definition of teacher quality: "Good teachers are ones who get large gains in student achievement for their classes; bad teachers are just the opposite" (p. 3). This definition, although apt as far as it goes,

provides only limited guidance to efforts to produce effective teaching (Cochran-Smith & Zeichner, 2005). The present study extends our understanding of the connection between specific observed teacher–student interactions and student achievement into secondary school classrooms and supports the theoretical proposition that properties of these interactions have value for student learning and development regardless of the content being taught.

In the present study, aspects of teacher–student interactions reflecting instructional support and classroom organization were predicted by the baseline level of achievement of students in a given class (although these still went on to predict achievement after accounting for these baseline levels). In contrast, however, features of teachers' emotional support appeared independent of students' baseline levels of achievement. These findings have implications for the observation of instruction (i.e., teacher oversight) in contexts in which value-added achievement test analyses cannot be readily conducted. These findings suggest that observed levels of instructional support and classroom organization, although important in predicting future student achievement, are likely to reflect both teacher skill and student background. Thus, to the extent students are implicitly or explicitly grouped or tracked into higher achieving and lower achieving classes, it would be important not to assume that observed qualities of instructional support and classroom organization were solely a result of teacher skills or efforts. Observed emotional support, in contrast, also predicts future student achievement after accounting for baseline achievement but appears relatively independent of student background characteristics, and thus may be more likely to be determined by individual teacher qualities. Notably, there is evidence now from experimental studies that teachers' behavior in all three CLASS domains can be improved with teacher training that targets these interactive behaviors via training either through college courses (Hamre et al., 2010) or through ongoing, job-embedded consultation (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Brown, Jones,

LaRusso, & Aber, 2010; Pianta, Belsky et al., 2008; Raver et al., 2008).

With regard to emotional support, attachment theorists posit that when adults provide emotional support and respond contingently, children develop self-reliance and are better able to explore (Ainsworth, Blehar, Waters, & Wall, 1978; Bowlby, 1969/1982), a premise that has been repeatedly validated in school environments (Birch & Ladd, 1998; Hamre et al., 2010; Howes, Hamilton, & Matheson, 1994). Self-determination theory (Connell & Wellborn, 1991; Skinner & Belmont, 1993) posits that motivation to learn is in part related to adults' support for competence, relationships, and autonomy (Roeser, Eccles, & Sameroff, 1998). Instructional interactions have been put in the spotlight in recent years as more emphasis has been placed on the translation of cognitive science, learning, and developmental research to educational environments (Bransford, Brown, & Cocking, 1999; Carver & Klahr, 2001; National Research Council, 2005). Teachers who use strategies that focus students on higher order thinking skills, give consistent, timely, and process-oriented feedback, and work to extend students language skills, tend to have students who achieve more academically (Hamre & Pianta, 2005; Justice, Meier, & Walpole, 2005; Meehan, Hughes, & Cavell, 2003; Taylor, Pearson, Peterson, & Rodriguez, 2003).

Emotional support may be particularly important in instructional settings serving adolescents. Unlike in elementary education, in which most students can be expected to have a disposition toward seeking to please teachers and comply with authority, engaging adolescent students emotionally may be critical to maximizing their academic motivation in the classroom. Further, as adolescents seek greater autonomy with respect to parents, having other settings in which they can receive emotional support from adults may provide a powerful motivation to engage in those settings (National Research Council, 2004).

That adolescents achieve at higher levels across a range of content areas in the context of more positive classroom interactions, although not yet well documented in the

educational literature, is actually quite consistent with developmental theory. In contrast to assessments targeted at specific content areas (e.g., seventh-grade math instruction), this assessment broadly targeted instruction across the full range of core secondary instructional content areas. Moderator analyses provided no evidence that predictions differed in magnitude across different content areas, suggesting that the generality of the approach to assessing instruction across widely different content areas was successful in this instance.

Follow-up descriptive analyses shed light on specific dimensions of teacher–student interaction that were most predictive of student outcomes. These analyses were explicitly descriptive and exploratory in nature; however, of necessity, the dimensions examined contained a degree of redundancy, both with the larger domains examined, but also with the other dimensions observed. Although we present them in terms of conventional significance levels, we recommend they be interpreted with caution, requiring further replication. In the Emotional Support domain, teachers' ability to establish a positive emotional climate (Positive Climate), their sensitivity to student needs (Teacher Sensitivity), and their structuring of their classroom and lessons in ways that recognize adolescents' needs for a sense of autonomy and control, for an active role in their learning, and for opportunities for peer interaction (Regard for Adolescent Perspectives) were all associated with higher relative student achievement, after covarying baseline levels of such achievement. Similarly, use of instructional learning formats that encouraged active participation by students and that provided variety in classroom approaches (Instructional Learning Formats) was also predictive of student achievement, as were lessons that required high levels of analysis and problem solving by students (Analysis and Problem Solving).

Overall, the particular constellation of interactions that was most linked to future achievement seemed to focus upon tailoring a classroom experience to be maximally emotionally and intellectually engaging to the adolescent. In post hoc analyses, this constella-

tion could account for a difference in student achievement test performance spanning the 37th–63rd percentiles, a magnitude that has considerable ramifications for teachers, classrooms, and schools in terms of accountability. Stated differently, on the particular tests used, the swing in scores reflected in moving from an average classroom to one that was one standard deviation above the mean score in overall teacher–student interaction quality would be sufficient in the abstract to reduce failure rates on these high-stakes tests from 17% to 11%.

Classroom observations have been used as measurement tools in educational research for more than three decades (Gage & Needels, 1989). In most studies using observational methods, approaches typically focused on specific teacher pedagogical behaviors. The present study results and other related findings (Pianta, Belsky et al., 2008) demonstrate the predictive utility of a focus on interactions, given the fact that they can be assessed in secondary classrooms via 40 min of video, and that global ratings of interactions, as they demonstrate patterns of behavior and response over the segment, can be coded reliably and account in some part for the value of that classroom setting for student learning. Notably, two of the domains of teacher–student interaction quality that were assessed, emotional and instructional support, were more strongly related to achievement in smaller as opposed to larger classrooms. One explanation for these findings is that qualities such as sensitivity to student needs or provision of high-quality feedback to students might have the greatest effect when they are concentrated among relatively fewer students. Conversely, the effect of these factors might be relatively diluted in very large classrooms.

The particular methods for observation used in this study also warrant discussion as they have implications for the use of standardized observations as measures of teacher performance in states and in districts. Results were obtained based on observations of just 40 min of a single classroom session early in the school year using global, judgment-based standardized ratings of teacher–student inter-

action. Under these conditions, the CLASS-S is likely capturing only a modest portion of the true variance in the quality of teacher–student interactions that exists across an entire year (Mashburn, Downer, Rivers, Brackett, & Martinez, 2013). Thus, the estimates of the effects of classroom interaction quality in this study, although substantial, should be taken not only as a likely minimum estimate of the effects of the specific qualities observed, but also as an indication that stable and reliable between-teacher differences in interaction can be observed in a fairly efficient manner.

At the same time, it should also be acknowledged that this was not an experimental study, and that even longitudinal analyses of end-of-year achievement controlling for baseline levels are not logically sufficient to demonstrate causal relations. It could be, for example, that other unmeasured factors were promoting both student achievement and high-quality instruction in the classroom. Even if this were the case, however, the results of this study would still indicate that the domains assessed were sensitive proxies of important classroom processes related to student achievement. Also, as achievement data could not be obtained on 24% of students in classrooms (because of a lack of parental consent), it remains possible that such additional data might have altered study findings, although it is not clear whether they would have biased findings in any particular direction. Teachers were also asked to select one of their more challenging classes to teach for this study, and thus results should properly be generalized to such classrooms. It is also possible that bias existed in which classroom session teachers selected for video recording, thus introducing unmeasured bias into the assessments obtained. Finally, it was not possible to assess effects of classroom nesting at the school level, which introduces an additional potential source of unmeasured bias into the results.

In sum, it is striking that given the importance of classroom settings as vehicles for the transmission of knowledge and skill in our system of education, that little-to-no population level data exist pertaining to adolescents' exposure to specific practices in the classroom

that are (a) known to relate to academic success or failure, (b) desired on the basis of certain policies or values, or (c) even hypothetically expected to relate to outcomes. In secondary classrooms there is no current work that provides national-level, observational data on these environments. The Measures of Effective Teaching Study (Bill and Melinda Gates Foundation, 2010) will likely produce a number of important findings in this regard in coming years. The MET study includes observations of over 3,000 teachers in Grades 4–9 and is using several observational protocols. Several, like CLASS-S, focused on global instructional process, and some focused specifically on content. We argue that designing standardized observation protocols into current value-added state-standards tests and large-scale student assessments could leverage not only our understanding of classroom effects (Early et al., 2005; National Institute of Child Health and Human Development Early Child Care Research Network, 2005; Pianta et al., 2005; Pianta, La Paro, & Hamre, 2008) but also the capacity to systematically produce effective teaching.

Potential Implications for Practitioners

In the midst of concern about standardized tests, block versus regular class scheduling, school size and budgets, this study emphasizes the fundamental importance of the emotional quality of the classroom as a key to adolescent learning. Although we can think of them as "learners," adolescents are first and foremost highly social and emotional beings. The role of school psychologists and administrators in both supporting this recognition and in helping teachers understand and enact it is potentially paramount. The elements of a positive classroom climate—teacher sensitivity to adolescents' needs, and recognition of their desire for peer interaction and for a sense of autonomy regarding classroom activities—are not simply the niceties in a functional classroom, they are key predictors of adolescent learning, even regarding otherwise "dry" subjects such as algebra or geometry. An advantage of the CLASS-S is that it provides

highly specific markers of these qualities, which can be used not only by those evaluating teacher–student interactions, but also by teachers seeking to enhance them. For example, teachers may enhance the positive climate of a class by laughing with students, engaging them as they enter the room (vs. sitting behind a desk at the front of the room), and asking about events outside of class at the beginning and end of a lesson period. They may give students choices (within a prespecified range) of ways of approaching learning assignments. They may use varied instructional learning formats—moving beyond lecturing to a largely silent class, for example, to asking students to work briefly in small teams, present ideas to the class, and so on—and focus on student analysis and problem solving by asking "why" questions not simply factual questions. All of these elements were part of the ratings of interactions linked to greater student learning in this study.

Equally important, for those charged with evaluating teachers, this study identified qualities of teacher–student interactions that may appear to be linked to quality teaching because they occur more in high-achieving classrooms, but which appear primarily to reflect qualities of students entering the classroom than to predict future achievement gains of those students. The quality of behavior management in a classroom is associated with student achievement, for example; however, the association is primarily to preexisting levels of student achievement rather than to future gains in achievement. This is not to say that behavior management is unimportant, but rather to suggest that great care be used in assessing teachers on this quality unless one also accounts for the qualities of the students being taught.

Perhaps the most important reason to conduct observational assessment of classrooms is for the purposes of professional development. Professional development typically occurs in the absence of a direct link to actual teaching behavior in classrooms, particularly for already-trained and certified teachers (Caspary, 2002). Systematic classroom observation systems provide a standardized ap-

proach to measuring and noting teachers' strengths and weaknesses and evaluating whether professional development activities are actually helping improve the classroom interactions responsible for learning. The current study suggests that professional development focused specifically on teachers emotional, organizational, and instructional interactions with students may enhance teacher effectiveness in ways that have a direct effect on student learning (Mashburn, Downer, Hamre, Justice, & Pianta, 2010), while clearly raising important questions for future research to consider.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation.* Hillsdale, NJ: Erlbaum.

Allen, J. P., Kuperminc, G., Philliber, S., & Herre, K. (1994). Programmatic prevention of adolescent problem behaviors: The role of autonomy, relatedness, and volunteer service in the Teen Outreach Program. *American Journal of Community Psychology, 22,* 617–638.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034–1037.

Bell, K. L., Allen, J. P., Hauser, S. T., & O'Connor, T. G. (1996). Family factors and young adult transitions: Educational attainment and occupational prestige. In J. A. Graber & J. Brooks-Gunn (Eds.), *Transitions through adolescence: Interpersonal domains and context* (pp. 345–366). Mahwah, NJ: Erlbaum.

Bill and Melinda Gates Foundation. (2010). *Working with teachers to develop fair and reliable measures of effective teaching.* Available online at http://www.metproject.org/downloads/met-framing-paper.pdf

Birch, S. H., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology, 34,* 934–946.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57,* 111–127.

Bogner, K., Raphael, K. M., & Pressley, M. (2002). How Grade-1 teachers motivate literate activity by their students. *Scientific Studies of Reading, 6,* 135–165.

Bowlby, J. (1969/1982). *Attachment and loss, Vol. 1.* New York: Basic Books.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brophy, J. (1999). *Teaching* (Vol. Educational Practices Series—1). Geneva, Switzerland: International Academy of Education and International Bureau of Education.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. *Handbook of Research on Teaching, 3,* 328–375.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs Program. *Journal of Educational Psychology, 102,* 153–167.

Burchinal, M., Howes, C., Pianta, R. C., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12,* 140–153.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology, 43,* 61–85.

Carver, S. M., & Klahr, D. (2001). *Cognition and instruction: 25 years of progress.* Mahwah, NJ: Erlbaum.

Caspary, K. (2002). *California's pioneering training and retention initiative for early childhood educators.* Berkeley, CA: PACE.

Cicchetti, D. V. (1981). Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. *Applied Psychological Measurement, 5*(1), 101–104.

Cochran-Smith, M., & Zeichner, K. (2005). *Studying teacher education: The report of the AERA panel on research and teacher education.* Mahway, NJ: Erlbaum.

Commonwealth of Virginia. (2005). *The Virginia Standards of Learning: Technical report.* Richmond, VA: Author.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. Gunnar & L. A. Sroufe (Eds.), *Self processes in development: Minnesota symposium on child psychology* (Vol. 23, pp. 43–77). Hillsdale, NJ: Erlbaum.

Early, D., Barbarin, O., Bryant, B., Burchinal, M., Chang, F., Clifford, R., Crawford, G., Howes, C., Sharon, R., Kraft-Sayre, M. E., Pianta, R. C., Barnett, S. W., & Weaver, W. (2005). *Pre-kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten and state-wide early educational programs (SWEEP) study.* Washington, DC: National Institute for Early Education Research.

Eccles, J. S., & Roeser, R. W. (1999). School and community influences on human development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (4th ed.) (pp. 503–554). Mahwah, NJ: Erlbaum.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36,* 103–112.

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

Greenwood, C. R., Carta, J. J., & Dawson, H. (2000). Ecobehavioral Assessment Systems Software (EBASS): A system for observation in education settings. In T. Thompson, D. Felce, & F. J. Symons (Eds.), *Behavioral observation: Technology and applications in developmental disabilities* (pp. 229–251). Baltimore, MD: Paul H. Brookes.

Greenwood, C. R., Carta, J. J., Kamps, D., & Terry, B. (1994). Development and validation of standard classroom observation systems for school practitio-

ners: Ecobehavioral Assessment Systems Software (EBASS). *Exceptional Children, 61,* 197–210.

Hambleton, R. K., Crocker, L., Cruse, K., Dodd, B., Plake, B. S., & Poggio, J. (2000). *Review of selected technical characteristics of the Virginia Standard of Learning (SOL) assessments.* Richmond, VA: Commonwealth of Virginia Department of Education.

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

Hamre, B. K., & Pianta, R. C. (2010). Classroom environments and developmental processes: Conceptualization & measurement. In J. Meece & J. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). New York, NY: Routledge.

Hamre, B. K., Pianta, R. C., Burchinal, M., & Downer, J. T. (2010). *A course on supporting early language and literacy development through effective teacher-child interactions: Effects on teacher beliefs, knowledge and practice.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Hanushek, E. A. (2002). *The long run importance of school equality.* Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., Rivkin, S. G., Figlio, D., & Jacob, B. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100,* 267–271.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn: Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*(1), 27–50.

Howes, C., Hamilton, C. E., & Matheson, C. C. (1994). Children's relationships with peers: Differential associations with aspects of the teacher-child relationship. *Child Development, 65,* 253–263.

Johnson, C. C., Kahle, J. B., & Fargo, J. D. (2007). Effective teaching results in increased science achievement for all students. *Science Education, 91,* 371–383.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: Findings from an intervention with at-risk kindergarteners. *Language, Speech, and Hearing Services in Schools, 36,* 17–32.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70,* 1373–1400.

Ladd, H. (2008). *Teacher effects: What do we know.* Paper presented at the Teacher Quality Conference, Northwestern University, Chicago, IL.

Landry, S. H., Swank, P. R., Smith, K. E., Assel, M. A., & Gunnewig, S. B. (2006). Enhancing early literacy skills for preschool children. *Journal of Learning Disabilities, 39,* 306.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology, 46,* 4–11.

Mashburn, A. J., Downer, J. T., Hamre, B. K., Justice, L., & Pianta, R. C. (in press). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Psychology.*

Mashburn, A. J., Downer, J. T., Rivers, S. E., Brackett, M. A., & Martinez, A. (2013). *Improving the power of an experimental study of a social and emotional learning program: Application of generalizability theory to the measurement of classroom-level outcomes.* Prevention Science.

McCaslin, M., Burross, H. L., & Good, T. L. (2005). Change and continuity in student achievement from Grades 3 to 5: A policy dilemma. *Education Policy Analysis Archives, 13*(1).

Meehan, B. T., Hughes, J. N., & Cavell, T. A. (2003). Teacher-student relationships as compensatory resources for aggressive children. *Child Development, 74,* 1145–1157.

National Research Council. (2004). *Engaging schools: Fostering high school students' motivation to learn.* Washington, DC: National Academies Press.

National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom.* Washington, DC: National Academies Press.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). A day in third grade: A large- scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105,* 305–232.

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

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85–112.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9,* 144–159.

Pianta, R. C. (1999). *Enhancing relationships between children and teachers.* Washington, DC: American Psychological Association.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45,* 365–397.

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

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

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS): Manual—Pre-K.* Baltimore, MA: Brookes Publishing.

Pressley, M., Roehrig, A., Raphael, L., Dolezal, S., Bohn, C., Mohan, L., . . . Hogan, K. (2003). Teaching processes in elementary and secondary education. In W. Reynolds & G. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 153–176). Hoboken, NJ: John Wiley & Sons.

Raudenbush, S. W., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75,* 346–353.

Raver, C. C., Jones, A. S., Li-Grining, C. P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*(1), 10–26.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45,* 958–972.

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

Roeser, R. W., Eccles, J. S., & Sameroff, A. J. (1998). Academic and emotional functioning in early adolescence: Longitudinal relations, patterns, and prediction by experience in middle school. *Development & Psychopathology, 10,* 321–352.

Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., . . . Okamoto, Y. (2006). *Teaching science in five countries: Results from the TIMSS 1999 Video Study* (NCES 2006–011). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics, 125*(1), 175–214.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77,* 454–499.

Shinn, M., & Yoshikawa, H. (2008). *Toward positive youth development: Transforming schools and community programs.* New York, NY: Oxford University Press.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational & Behavioral Statistics, 23,* 323–355.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571–581.

Soar, R., & Soar, R. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications* (pp. 97–119). Berkeley, CA: McCutchan.

Sroufe, L. A., Egeland, B., Carlson, E. A., & Collins, W. A. (2005). *The development of the person: The Minnesota Study of Risk and Adaptation from Birth to Adulthood.* New York, NY: Guilford.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*(2), 3–28.

Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology, 39,* 217–228.

Wallace, T., Anderson, A. R., Bartholomay, T., & Hupp, S. (2002). An ecobehavioral examination of high school classrooms that include students with disabilities. *Exceptional Children, 68,* 345–359.

Watson, S. M. R., Gable, R. A., & Greenwood, C. R. (2011). Combining ecobehavioral assessment, functional assessment, and response to intervention to promote more effective classroom instruction. *Remedial and Special Education, 32,* 334–344.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Retrieved June 2, 2009, *from* http://widgeteffect.org/

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States.* Chapel Hill, NC: Horizon Research.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Outstanding literacy instruction in first grade: Teacher practices and student achievement. *Elementary School Journal, 99,* 101–128.

Ysseldyke, J., & Christenson, S. (1993). *The Instructional Environment System—II (TIES-II).* Colorado Springs, CO: Sopris West.

Ysseldyke, J. E., & Christenson, S. L. (1987). Evaluating students' instructional environments. *Remedial and Special Education, 8*(3), 17–24.

Joseph Allen, PhD, is the Hugh P. Kelly Professor of Psychology at the University of Virginia, specializing in the study of adolescent social development.

Anne Gregory, PhD, is an associate professor of psychology in the Graduate School of Applied and Professional Psychology at Rutgers University, specializing in the study of disproportionality in implementation of disciplinary procedures across racial/ethnic groups in schools.

Amori Mikami, PhD, is an assistant professor of psychology at the University of British Columbia, specializing in the study of ways in which school and home environments influence social development.

Janetta Lun, PhD, is currently a research associate in the Department of Psychology at the University of Maryland, specializing in the study of intergroup relations and cultural psychology.

Bridget Hamre, PhD, is a research scientist at the Curry School of Education at the University of Virginia and currently serves as associate director of the Center for Advanced Study of Teaching and Learning at the University of Virginia.

Robert Pianta, PhD, is the Novartis U.S. Foundation Professor of Education and currently serves as dean of the Curry School of Education and director of the Center for Advanced Study of Teaching and Learning at the University of Virginia.