




The Relationship Between School Administrator and Teacher Ratings of Classroom Practices and Student Achievement in High-Poverty Schools

Assessment for Effective Intervention
2021, Vol. 46(2) 87–98
© Hammill Institute on Disabilities 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1534508419862863
aei.sagepub.com
SAGE

Linda A. Reddy, PhD¹ , Anh N. Hua, PhD¹, Christopher M. Dudek, MEd¹ ,
Ryan Kettler, PhD¹ , Ilona Arnold-Berkovits, PhD¹, Adam Lekwa, PhD¹,
Kevin Crouse, MA¹, Alexander Kurz, PhD¹, and Jiefang Hu, PhD¹

Abstract

This study examined the relationship of school administrator and teacher self-ratings of instructional and behavioral management practices to student growth on statewide achievement tests (Partnership for Assessment of Readiness for College and Career [PARCC]). The study included 78 teachers and 1,594 students from fourth through eighth grades in nine high-poverty charter schools. Observation scores completed by school administrator and teacher self-ratings were collected on the Classroom Strategies Assessment System (CSAS), an observational assessment that reports outcomes as discrepancy scores: differences between recommended frequency and observed frequency of specific instructional and behavior management strategies for teachers. Correlations revealed negative relations between both informants' discrepancy scores and PARCC growth scores, demonstrating that teachers with lower discrepancy scores tended to have students with greater PARCC growth scores. Hierarchical multiple regression analyses revealed school administrator and teacher CSAS total discrepancy scores were related to student performance on PARCC mathematics, but not English Language Arts (ELA), and teachers' CSAS Total discrepancy scores explained an additional 4.8% of variance in PARCC mathematics. Implications of findings for professional development and research are offered.

Keywords

progress monitoring, achievement assessment, instruction

Educators' ability to provide instruction and manage the classroom environment is one of the most influential factors for improving student achievement. How we support teachers' acquisition and use of research-based classroom practices is essential for enhancing the quality of instruction, student learning, and school improvement (Reddy, Dudek, & Shernoff, 2016). This is particularly important for teachers who serve students in high-poverty schools, as teacher attrition, burnout, and gaps in student achievement remain high in comparison with other school contexts (e.g., Borman & Dowling, 2008; Boyd et al., 2012). High-poverty schools typically have more difficulties recruiting and retaining high-quality teachers (Darling-Hammond & McLaughlin, 1995, 2003). As a result, students in these settings are more likely to be exposed to less effective instructional practices and experience more negative interactions with students (e.g., Krei, 1998; Langford, Loeb, & Wyckoff, 2002; Nye, Konstantopoulos, & Hedges, 2004). Thus, it is not surprising that students who

live in high poverty are at significant risk of academic, behavior, and social failure (Belfiore, Auld, & Lee, 2005; Espinosa, 2005; Stormont, 2007).

Teachers working in high-poverty schools can benefit from valid assessment-based feedback on their use of effective instructional and behavioral management practices as well as their relations to student achievement (Reddy, Dudek, & Lekwa, 2017). Indeed, reliable assessment scores that permit valid inferences about effective classroom practices are an important first step in addressing the needs of teachers and their students, especially those working with the most vulnerable student populations.

¹Rutgers University, Piscataway, NJ, USA

Corresponding Author:

Linda A. Reddy, Graduate School of Applied and Professional Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA.
Email: LReddy@Rutgers.edu

The need for valid assessments of classroom practices becomes even more paramount as the benefits and consequences of the recent Elementary and Secondary Education Act (ESSA, 2015) passage take effect in states across the nation. Prior to ESSA, the federal government's involvement in teacher accountability systems was heavily prescriptive, especially in its requirement that there be a "highly qualified" teacher in core classrooms, and that states seeking flexibility waivers under the No Child Left Behind (NCLB, 2001) Act implement educator accountability systems based in significant part on students' test scores. This requirement was reflected in federally funded programs and legislation (e.g., Race to the Top Fund [U.S. Department of Education, 2009]; Teacher Incentive Fund, 2012); and to meet these demands, the majority of states (i.e., 90%; National Council on Teacher Quality, 2015) adopted formal legislation on teacher evaluation policies that included student achievement scores (McGuinn, 2012; Sawchuk, 2016). These policies resulted in widespread implementation of multimethod teacher evaluation approaches that combined various measures of teacher practice with student academic achievement measures. Now, however, the passage of ESSA (2015) has removed federal prescriptions from states' procedures for licensing and evaluation of teachers (Klein, 2017a; Sawchuk, 2016), and permitted states to remove student test scores from teacher evaluation systems all together. According to the National Council on Teacher Quality (2015), six states so far have removed student scores from their teacher evaluation systems, with the potential for more states to follow suit as the political climate changes (Klein, 2017b). This turnaround in federal policy offers new pathways for defining educator effectiveness and addressing challenges inherent to design and measurement in teacher evaluation systems.

Although the multimethod evaluation policies enacted prior to ESSA (2015) were improvements from the reliance on one-time student proficiency (status) scores under the NCLB Act of 2001 era, the contemporary approaches of evaluation still in place post-ESSA (2015) fail to systematically capture *and* quantify teacher input of the instructional environment. For more than two decades, research has shown that teacher self-report assessment can improve instructional delivery and teachers' willingness to engage in self-reflection about their practice (e.g., Koziol & Burns, 1986; Reddy, Dudek, & Shernoff, 2016). Desimone and colleagues' (Desimone, Smith, & Frisvold, 2010) meta-analytic study indicated teachers' self-reports on their teaching quality are strongly related with classroom observations and teacher instructional planning records. Likewise, combining teacher self-assessments with classroom observation assessments can enhance professional development (PD) conversations that lead to targeted supports and changes in

teaching effectiveness related to student achievement and social behavior (Reddy, Fabiano, Dudek, & Hsu, 2013a).

There are few valid multirater classroom assessment approaches that assess teachers' use of evidence-based classroom instructional *and* behavioral management practices. There are even fewer tools that quantify *both* observer and teacher input on the same instructional metrics (Reddy, Dudek, Fabiano, & Peters, 2015). Likewise, it remains unknown how the combined use of school administrators' and teachers' self-assessment of best practices relate to student academic growth and can inform PD supports.

Classroom Strategies Assessment System (CSAS)

One classroom assessment designed to fill this void in available school-based assessments for both school administrators and teachers is the CSAS (Reddy & Dudek, 2014). The CSAS is a multirater, multisource assessment that formatively evaluates and supports teachers' instructional and behavioral management practices. The instructional strategies of the CSAS are guided by the direct (explicit), constructivist, and differentiated learning models of instruction (e.g., Brophy & Good, 1986; Hattie, 2009; Walberg, 1986), and also examine opportunities to respond, techniques for metacognitive and critical thinking, and performance feedback delivery to students. The behavior management strategies come from prevention and antecedent approaches as well as the positive behavioral interventions and supports literature, and include such strategies as behavioral reinforcement, classroom routines, and rules (Gable, Hester, Rock, & Hughes, 2009; Kerns & Clemens, 2007).

Studies have examined the relationship between CSAS–Observer Form (CSAS-O) scores and student performance on statewide achievement. For example, Reddy, Fabiano, Dudek, and Hsu (2013b) found the CSAS-O Instructional Strategy Rating Scale scores predicted New York statewide mathematics and English Language Arts (ELA) proficiency scores in a sample of 662 third- through fifth-grade students from 32 classrooms. CSAS scores reflecting teachers' limited use of evidence-based practices were associated with lower student proficiency scores (approximately 27% and 25% lower odds of proficiency in ELA and mathematics). In a sample of 829 fourth- through fifth-grade students from six urban high-poverty schools, Dudek, Reddy, and Lekwa (2018) found the Instructional Strategy and Behavior Management Total Rating Scale scores predicted New Jersey statewide mathematics and ELA proficiency scores, similarly yielding approximately 38% and 32% reductions in odds of students performing at proficiency for teachers with limited use of evidence-based practices. In sum, these two investigations offer emerging evidence of the CSAS' use for assessing qualities of teachers' classroom practices as they are related to student proficiency scores in schools.

However, research has yet to examine how ratings from both school administrators and teachers relate to student achievement growth in the context of high-poverty schools.

The goal of the current study is to address how the use of school administrators' and teachers' ratings of best practices as measured by CSAS-O and CSAS-Teacher Forms (CSAS-T) relate to student academic growth in mathematics and ELA on the Partnership for Assessment of Readiness for College and Career (PARCC, 2018). This is the first study of the CSAS-O and CSAS-T discrepancy scores in relation to achievement growth for youth in high-poverty schools. Discrepancy scores on the CSAS represent differences between recommended frequencies and observed frequencies of specific instructional and behavior management strategies for teachers. Thus, lower discrepancy scores represent higher teaching quality. We hypothesize the following:

Hypothesis 1: CSAS Observer and Teacher Strategy Rating Scale discrepancy scores would be negatively correlated with student performance on PARCC.

Our research questions included the following:

Research Question 1: Do school administrators' ratings of classroom practices, as measured by CSAS-O, predict student growth on PARCC mathematics and ELA scores?

Research Question 2: Do teacher self-report ratings of classroom practices, as measured by the CSAS-T, predict student growth on PARCC mathematics and ELA beyond what is explained by observer ratings?

Method

Fourth- through eighth-grade teachers ($n = 78$) from nine high-poverty charter schools were observed by 17 school administrators as part of their routine educator evaluation in New Jersey during the 2014–2015 school year. Participating teachers were predominantly White and female (approximately 67%) with a mean age of 33.6 years ($SD = 10.4$ years) and a mean teaching experience of 5.2 years ($SD = 4.8$ years). Among the total 78 teachers, 10 taught in fourth through fifth grades and 23 in sixth through eighth grades. The remaining 45 teachers served in multiple grade levels between fourth and eighth grades. The average number of students per classroom was 21. Teachers' median student growth percentile (mSGP) scores are based on rostered student growth percentiles (SGPs) during the school year. Most teachers ($n = 66$) received mSGPs for mathematics, and all teachers received mSGPs for ELA. mSGPs are used by the state of New Jersey (Evaluation scoring, n.d.) and many other state departments of education as a metric for teacher evaluation.

School principals were predominantly White females (76%), with a mean age of 42.6 years ($SD = 12.5$ years) and an average of 6.5 years ($SD = 6.8$ years) of school administrative experience. Approximately 65% of principals held master's degrees, and 23% held doctoral degrees. The mean years of teaching experience possessed by principals was 10.65 ($SD = 6.0$ years).

The study sample included a total of 1,594 students; 1,317 students had SGPs for mathematics and 1,553 students had SGPs for ELA. About half of the student sample ($n = 817$) was female. The student sample included representation from African American (42%), Latino American (35%), and White (11%) students. The large majority of students (77%) were eligible for free or reduced-priced lunch.

Measures

CSAS. The CSAS-O and CSAS-T assess teachers' use of evidence-based instructional and behavioral management practices. The CSAS-O is a direct observation measure completed by observers (i.e., school administrator) and consists of discrete teacher behavior counts (Strategy Counts), Strategy Ratings Scales, and a Classroom Checklist. The CSAS-T consists of the same Strategy Ratings Scales and Classroom Checklist as the CSAS-O, and is completed through teacher self-report (see Tables 1 and 2 for definitions of the strategies on both CSAS forms). In the current study, the Classroom Checklist was not utilized. When used together, the CSAS-O and CSAS-T enhance PD dialogue between teachers and their school administrators about the effective instructional practices that occurred in an observed lesson by focusing both informants' understanding on a shared understanding of effective teaching practices. Individually, both measures have demonstrated adequate factor structure, interobserver agreement, internal consistency, and test-retest reliability (Reddy et al., 2015; Reddy, Fabiano, Dudek, & Hsu, 2013a) as well as evidence of predictive validity for mathematics and ELA proficiency status on statewide testing (Dudek et al., 2018; Reddy et al., 2013b; Reddy et al., 2019).

To complete the CSAS-O, school administrators must fill out information both during and after an observed lesson. During the observation period, school administrators complete the Strategy Counts section by tallying eight instructional or behavior management strategies used during the observation (lesson) period (see Table 1). In addition, during the observation period administrators take notes related to the Strategy Rating scales dimensions as well as lesson content, activities, student interactions, and student learning. Following the observation period, school administrators complete the Strategy Rating Scales (54 items total), which consist of Instructional Strategies (IS; 28 items) and Behavior Management Strategies (BMS; 26 items) scales (see Table 2 for scale definitions).

Table 1. Descriptions of the CSAS–Observer Form—Strategy Counts.

Instructional Strategies	
Concept summaries	Teacher summarizes or highlights key concepts or facts taught during the lesson. Summarization statements are typically brief and clear. This teaching strategy helps students organize and recall material taught.
Academic response opportunities	Teacher creates opportunities for students to share their understanding of the lesson content with the teacher or class. These opportunities can be verbal or nonverbal response (e.g., explain answers, repeat key points, brainstorm ideas, and show answers on the board).
Academic praise	Teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate academic performance.
Academic corrective feedback	Teacher gives verbal or nonverbal statement or gesture to provide feedback for incorrect academic performance.
Behavior management strategies	
Clear directives	Teacher gives a verbal instruction that specifically directs a behavior to occur immediately. These directives are clear and they provide specific instructions to students to perform a behavior. They are declarative statements (not questions), describe the desired behavior, and include no more than two steps.
Vague directives	Teacher gives a verbal instruction that is unclear when directing a behavior to occur immediately. These directives are vague, may be issued as questions, and often include unnecessary verbalizations of more than two steps.
Behavior praise	Teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate behavior.
Behavior correct feedback	Teacher gives verbal or nonverbal statement or gesture to provide feedback for inappropriate behavior.

Note. CSAS = Classroom Strategies Assessment System.

The IS scale includes 28 items that collectively yield a total scale, two composite scales, and five subscales. The Instructional Methods Composite Scale includes 17 items subdivided into the Adaptive Instruction subscale (four items), Student Directed Instruction subscale (five items), and the Direct Instruction subscale (eight items). The Monitoring and Feedback Composite Scale includes 11 items subdivided into the Promotes Students' Thinking subscale (five items) and Academic Performance Feedback subscale (six items). The BMS scale includes 26 items that create a total scale, two composite scales, and four subscales. The Proactive Methods Composite contains 14 items subdivided into the Proactive Methods subscale (eight items) and the Directives subscale (six items). The Behavior Feedback composite is composed of 12 items subdivided into the Praise subscale (five items) and Behavioral Corrective Feedback subscale (seven items).

For the IS and BMS Rating scales, school administrators rate how often (*observed frequency rating*) teachers used specific instructional and behavior management strategies on a 7-point Likert-type scale (1 = *never used*, 3 = *sometimes used*, 7 = *always used*) and then rate how often the teachers *should have used* each strategy (*recommended frequency*) on a 7-point Likert-type scale (1 = *never used*, 3 = *sometimes used*, 7 = *always used*). The Strategy Rating scales then produce a third score, the *discrepancy* score, which is calculated by subtracting the *observed frequency* from the *recommended frequency* and taking the absolute value of the difference (i.e., |recommended frequency – frequency ratings|). Absolute value discrepancy scores indicate whether any change

(regardless of direction) was needed. Larger discrepancy score values indicate greater amount of change in the practices is needed. Discrepancy scores are first calculated at the item level, and then summed to create their corresponding subscale, composite, and total discrepancy scores.

To complete the CSAS-T, teachers self-rate their performance for the observed lesson on the same 28 IS and 26 BMS items of the Strategy Rating scales, and the same Classroom Checklist items as the CSAS-O.

PARCC. The PARCC assessments are end-of-year summative, computer-based tests for students that have been used in several states since 2015. Aligned with the Common Core State Standards (National Governors Association 2010), PARCC provides annual evaluation of students' academic achievement in Grades 3 through 11 for mathematics and ELA. Cronbach's alpha reliability estimates ranged from .85 to .94 for the mathematics tests and from .89 to .93 for the ELA tests (Educational Testing Services [ETS], Pearson, & Measured Progress). In addition to providing information about student academic achievement levels near the end of a school year, the PARCC assessments have also been used by states to provide estimates of growth in achievement for students in Grades 4 through 8 through SGP scores (Betebenner, 2011). As per New Jersey Department of Education (NJ DOE) procedures, an SGP is calculated based on a student's performance from 2013–2014 to 2014–2015 on the PARCC assessments, and it indicates the percentage of academic peers (i.e., students with similar academic achievement in the prior year), above whom the

Table 2. Descriptions of the CSAS Strategy Rating Scales: IS and BMS Rating Scales for Observer and Teacher Forms.

Name of Scale	Definitions
Total scale	The Total IS and BMS Rating scales reflects the overall use of IS and BMS.
IS Scales	
IS Total scale	The Total IS scale reflects the overall use of Instructional Methods and Academic Monitoring/ Feedback.
Student Directed Instruction subscale	Strategies for engaging students in the lesson, creating active learners, and encouraging self-initiative in the learning process. These practices encompass direct experience, hands-on instructional techniques, linking lesson content to personal experiences, and cooperative learning strategies.
Direct Instruction subscale	Methods for conveying information to students and strategies employed while teaching lesson content/concepts. These practices include modeling, advanced organizers, summarizing, and other instructional methodology.
Promotes Student Thinking subscale	Practices for stimulating students' metacognitive and higher order thinking abilities. They encourage students to critically think about the lesson material (why/how analysis), generate new ideas, and examine their own thought processes.
Academic Performance Feedback subscale	How teachers provide feedback to students on their understanding of the material. These practices assess teacher efforts to explain what is correct or incorrect with student academic performance.
BMS Scales	
BMS Total scale	The Total BMS scale reflects the overall use of Proactive Methods and Behavior Feedback.
Praise subscale	Verbal and nonverbal strategies teachers use to praise students for specific appropriate behaviors in the classroom.
Corrective Feedback subscale	Verbal and nonverbal strategies teacher use to redirect or correct students' inappropriate behavior in the classroom.
Proactive Methods subscale	Proactive verbal and nonverbal strategies teachers use to promote positive classroom functioning and establish effective learning environments. These practices include taking actions to prevent problem behaviors from occurring, establishing clear and consistent expectations, and creating a positive atmosphere in the classroom.
Directives subscale	Strategies teachers use to communicate their behavioral requests to students and manage the movement and behavior of students during class transitions.

Note. CSAS = Classroom Strategies Assessment System; IS = Instructional Strategies; BMS = Behavior Management Strategies.

student scored. SGPs are calculated within grade levels. Based on the PARCC 2017 technical report, PARCC scores and standard errors across Grades 4 through 8 for ELA and mathematics are similar (PARCC, 2018).

Study Procedures

Observer training and reliability. All administrators received a 3-day training on the CSAS-O that consisted of a four-step process. First, administrators were oriented to the scientific literature guiding the CSAS theory, evidence, and the effective instruction and behavior management literature guiding the CSAS construct definitions (e.g., Brophy & Good, 1986; Hattie, 2009; Walberg, 1986). Second, administrators were trained on how to score, and practiced the scoring mechanics of the observed and recommended frequency ratings. They were trained to score the items according to (a) the effective instruction literature, (b) CSAS scoring principles, and (c) the administrators' targeted notes aligned to the CSAS constructs and items during classroom

observations. Third, they completed and passed knowledge tests to ensure acquisition of training content. Fourth, administrators practiced coding classroom videos using the CSAS, and practice feedback was provided by a CSAS trainer/master coder. Finally, administrators were required to pass a video coding criterion test on the CSAS, which required them to rate five classroom videos using the CSAS and reach the minimum reliability criterion level. All administrators in the current study passed the criterion test and were provided with additional co-observation practice in live classrooms with a certified CSAS trainer/master coder. In this study, pairs of observers completed CSAS-O observations concurrently for approximately 30 min. Average rates of agreement for discrepancy scores for IS Total (86%) and BMS Total (83%) were obtained.

The CSAS-T training consists of a two-step process to enhance teachers' self-assessment of teaching practices. First, the teachers received a 3-hr didactic training session from a CSAS trainer/master coder, which included discussion of the CSAS theory, evidence, construct definitions,

and criteria for scoring. Similar to school administrators, teachers were oriented to the scientific literature guiding the development of the CSAS and the recommended frequencies of strategies, to ensure teachers operate from a common knowledge base for judging the recommended frequency of the CSAS Strategy Rating scales. Second, training on the recommended frequency of strategies was informed by a group discussion of the effective instruction literature in relation to CSAS constructs and items, as well as review and discussion of brief classroom video clips with a trained CSAS master coder.

CSAS procedures. In the current study, three observations were conducted by school administrators for each teacher as part of yearly evaluation processes. Observations occurred throughout the academic school year during the fall, winter, and spring of 2014 and 2015, and were conducted in accordance with the teacher evaluation system procedures. Teachers completed the CSAS-T immediately following each lesson observed by their administrator. Each teacher had three sets of CSAS-O scores and three sets of corresponding CSAS-T scores.

For the Strategy Rating Scale scores, item-level discrepancy scores were first calculated for each observation separately, and then summed to create their corresponding subscale, composite, and total discrepancy scores for each observation. The subscale, composite, and total discrepancy scores for the three CSAS-O observations were then averaged together, respectively, to create overall subscale, composite, and total scores for the year. This process was repeated for the CSAS-T discrepancy scores.

PARCC procedures. The PARCC was administered to students in New Jersey in third grade through eighth grade during the months of April and May 2015. Depending on grade level, testing occurred for approximately four to five mornings. The test scores of students were then matched to teachers through an online student–teacher course rostering sponsored by the NJ DOE. The research team computed mSGPs for each participating teacher in the study. The NJ DOE computes teachers' mSGPs using 20 or more students who are rostered to them in the most recent school year or 20 students who were rostered to them in three consecutive years. In this study, teachers' mSGPs were computed based on 2013–2014 and 2014–2015 school years. We calculated mSGPs for teachers who had five or more students rostered during 2013–2014 and 2014–2015 school years to (a) maximize our samples of teachers (charter schools tend to have smaller numbers of students per class than traditional school districts) and (b) increase statistical power.

Data Analysis

Descriptive statistics including Pearson correlations were computed on all CSAS scales and mathematics and

ELA mSGP scores. Bivariate correlations with magnitudes in the .00s were considered nonexistent, .10s and .20s small, .30s and .40s medium, .50s and .60s large, .70s and .80s very large, and .90s nearly perfect (Cohen, 1992). Also, hierarchical multiple regression analyses were conducted at the teacher level to determine whether teacher self-report ratings of classroom practices predict teachers' mSGPs beyond what was explained by observer ratings. In Models 1 and 2, the overall scores for CSAS-O and CSAS-T are entered as predictors, for mathematics and ELA, respectively. In Models 3 and 4, the subdomain scores, IS and BMS, are entered as predictors, for mathematics and ELA, respectively. In each model, a predictor is entered into the regression equation one step at a time to determine the change in R^2 associated with that variable (Keith, 2014). Thus, ΔR^2 , and its corresponding change in F (ΔF) and one-tailed p values are the statistics of interest in the analyses, and they inform whether adding a predictor significantly improves the model's ability to predict PARCC mSGPs.

Missing data. Twelve teachers, out of the total sample of 78, were missing in the mathematics data set because their rostered students did not receive PARCC growth scores from 2014 to 2015 school years. To investigate whether there were systematic patterns in the missing data, two Little's Missing Completely at Random (MCAR) tests were conducted for Model 1 and Model 3, where the dependent variables were teachers' mSGPs in mathematics. For both models, Little's MCAR tests indicated that the data were missing completely at random, $\chi^2(2) = 0.88, p = .64$, and $\chi^2(4) = 5.17, p = .27$, respectively. As a result, listwise deletion was used to address missing data (Enders, 2010).

Multicollinearity diagnostics. To confirm multicollinearity was not an issue in our analyses, tolerance and variance inflation factor statistics were obtained for each of the four models. For Model 1 and Model 2, tolerance values ranged from .949 to .979 and variance inflation factor values ranged from 1 to 1.054. For Model 3 and Model 4, tolerance values ranged from .25 to .48 and variance inflation factor ranged from 2.07 to 4.00. The sizes of these statistics indicated that the predictor variables were sufficiently independent from each other and that the magnitudes of the inflated variance factors due to multicollinearity were minimal.

Results

Table 3 presents the descriptive statistics of CSAS-O, CSAS-T, and PARCC mSGP scores. Teachers who received PARCC mathematic growth scores yielded comparable means and standard deviations with those who received ELA growth scores. In particular, across both subjects, the distributions for CSAS-O Total and CSAS-T

Table 3. Descriptive Statistics for CSAS-O and CSAS-T and PARCC Mathematics and ELA SGPs.

Samples & Measures	<i>M</i>	<i>SD</i>	Range	Skew	Kurtosis
Mathematics sample (<i>n</i> = 66)					
CSAS-O Total discrepancy	21.65	15.04	1.00–62.00	0.74	–0.11
IS total discrepancy	11.28	7.75	0.00–31.33	0.57	–0.45
BMS total discrepancy	10.37	8.43	0.00–35.33	1.00	0.51
CSAS-T Total discrepancy	29.16	20.13	2.00–110.33	1.54	3.31
IS total discrepancy	14.91	9.90	1.33–57.67	1.46	3.62
BMS total discrepancy	14.26	11.05	0.67–56.50	1.75	3.68
PARCC mSGP	56.71	16.70	17.5–86.00	–0.40	–0.66
ELA sample (<i>n</i> = 78)					
CSAS-O Total discrepancy	21.06	14.56	1.00–62.00	0.76	–0.01
IS total discrepancy	10.72	7.45	0.00–31.33	0.68	–0.20
BMS total discrepancy	10.34	8.28	0.00–35.33	0.93	0.35
CSAS-T Total discrepancy	29.44	20.89	2.00–110.33	1.46	2.61
IS total discrepancy	15.18	10.54	1.33–57.67	1.43	2.85
BMS total discrepancy	14.26	11.10	0.67–56.50	1.64	3.07
PARCC mSGP	55.77	14.59	21.00–96.00	–0.13	0.15

Note. CSAS = Classroom Strategies Assessment System; CSAS-O = CSAS–Observer Form; CSAS-T = CSAS–Teacher Forms; PARCC = Partnership for Assessment of Readiness for College and Career; ELA = English Language Arts; SGP = Student Growth Percentile; IS = Instructional Strategies; BMS = Behavior Management Strategies; mSGP = median Student Growth Percentile.

Total scores were similar: CSAS-O Total: $M = 21.6$, $SD = 15.0$ for mathematics and $M = 21.1$, $SD = 14.6$ for ELA; CSAS-T Total: $M = 29.2$, $SD = 20.1$ for mathematics and $M = 29.4$, $SD = 20.9$ for ELA. The distributions for the IS and BMS Total scales across the CSAS-O and CSAS-T measures were also similar for mathematics and ELA. The distributions of mSGPs were similar between mathematics ($M = 56.7$, $SD = 16.7$) and ELA ($M = 55.8$, $SD = 14.6$). Finally, due to the small sample sizes, distributions for mathematics and ELA were similarly skewed (skewness values ranged from -0.40 to 1.75 for mathematics, and -0.13 to 1.64 for ELA) and similarly kurtotic (kurtosis values ranged from -0.66 to 3.68 for mathematics, and -0.20 to 3.07 for ELA). Because ordinary least squares regression makes no assumptions about the shape of the independent or dependent variables but rather the shape of the residuals (Keith, 2014), the kurtotic and skewed raw data would not likely affect the results. Visual inspections of the Q–Q plots showed that there were no obvious violations of normality of the residuals across models. In addition, plots of unstandardized residuals against the independent variables indicated that the assumption of linearity was met. They also showed that the data were fairly homoscedastic across different levels of the predicted mSGPs.

Based on mSGPs, the current sample performed slightly higher than the population of test takers. Compared with a normative mean of growth at the 50th percentile, students in the current sample were near the 56th percentile for mathematics and near the 57th percentile for ELA. The variance of the current sample demonstrates that this performance is a little more

than $1/3$ SD above the normative mean. These estimates are based on the average of the median student performance by classroom, which may be different than the grand mean or median of the student sample; calculation this way is appropriate in this study because teachers are the unit of analysis.

The correlations between the CSAS-O and CSAS-T ranged from $r = .15$ to $r = .25$ for mathematics and $r = .07$ to $r = .17$ for ELA, as shown in Table 4. These correlations suggest that, besides measurement error, the two measures capture somewhat different attributes of teaching, and provide preliminary support for why it is important to examine the incremental validity of CSAS-T beyond CSAS-O in predicting teachers' mSGPs.

Next, correlations between administrator versus teacher self-ratings and PARCC mSGPs were also computed separately for mathematics and ELA scores. As hypothesized, negative correlations were found between the CSAS-O or CSAS-T discrepancy scores and PARCC mSGPs, indicating that a larger need for change in teachers' classroom practices was associated with smaller mSGPs in mathematics and ELA (Cohen, 1992). For the CSAS-O, the correlations were significant for mathematics (Total $r = -.40$, IS $r = -.26$, BMS $r = -.48$; $ps < .05$) and were in the small and medium ranges. For ELA, the correlation between the PARCC and the BMS was significant ($r = -.29$, $p < .05$) and in the small range, whereas the correlations between the PARCC and the total score ($r = -.16$, $p > .05$) and the IS ($r = -.01$, $p > .05$) were nonsignificant. For the CSAS-T, the correlations were significant for mathematics (Total $r = -.31$, IS $r = -.29$, BMS $r = -.30$; $ps < .05$) and were in the small and medium ranges (Cohen, 1992). For ELA, the

Table 4. Correlations Between CSAS-O and CSAS-T and PARCC Mathematics and ELA SGPs.

Samples & Measures	CSAS-O Total	CSAS-O IS	CSAS-O BMS	CSAS-T Total	CSAS-T IS	CSAS-T BMS
PARCC mathematics sample						
CSAS-O Total						
CSAS-O IS	.92***					
CSAS-O BMS	.94***	.73***				
CSAS-T Total	.23*	.17	.25*			
CSAS-T IS	.21*	.15	.24*	.96***		
CSAS-T BMS	.22*	.18	.23*	.97***	.85***	
mSGPs	-.40***	-.26*	-.48***	-.31**	-.29**	-.30**
PARCC ELA sample						
CSAS-O Total						
CSAS-O IS	.92***					
CSAS-O BMS	.93***	.71***				
CSAS-T Total	.14	.10	.16			
CSAS-T IS	.11	.07	.14	.96***		
CSAS-T BMS	.16	.13	.17	.97***	.86***	
mSGPs	-.16	-.01	-.29**	-.18	-.17	-.18

Note. CSAS = Classroom Strategies Assessment System; CSAS-O = CSAS–Observer Form; CSAS-T = CSAS–Teacher Form; PARCC = Partnership for Assessment of Readiness for College and Career; ELA = English Language Arts; SGP = Student Growth Percentile; IS = Instructional Strategies; BMS = Behavior Management Strategies; mSGP = median Student Growth Percentile.

*Significant (one-tailed) at the .05 level; ** .01 level; and ***.001 level.

Table 5. Hierarchical Multiple Regression: CSAS-O and CSAS-T Discrepancy Scores Predicting PARCC SGPs.

Step and Predictors Entered	Model 1: Mathematics		Model 2: ELA		Model 3: Mathematics		Model 4: ELA	
	R ²	ΔR ²	R ²	ΔR ²	R ²	ΔR ²	R ²	ΔR ²
1. CSAS-O Total	.162***		.025					
2. CSAS-T Total	.210***	.048*	.050	.025				
1. CSAS-O IS total					.066*		.000	
2. CSAS-O BMS total					.249***	.183***	.173***	.173***
3. CSAS-T IS total					.279***	.030	.186**	.013
4. CSAS-T BMS total					.287***	.007	.190**	.004

Note. CSAS = Classroom Strategies Assessment System; CSAS-O = CSAS–Observer Form; CSAS-T = CSAS–Teacher Form; PARCC = Partnership for Assessment of Readiness for College and Career; SGP = Student Growth Percentile; ELA = English Language Arts; IS = Instructional Strategies; BMS = Behavior Management Strategies.

*Significant (one-tailed) at the .05 level; ** .01 level; and ***.001 level.

correlations with PARCC were nonsignificant (Total $r = -.18$, IS $r = -.17$, BMS $r = -.18$; $p > .05$).

School Administrator Ratings Predict Student Growth

Model 1 results show CSAS-O Total discrepancy scores significantly predicted mSGPs in mathematics, $R^2 = .162$, $F(1, 64) = 12.36$, $p < .001$, contributing 16.2% of variance, although the CSAS-O Total did not significantly predict mSGPs in ELA, $R^2 = .025$, $F(1, 76) = 1.98$, $p > .05$; Model 2. For Model 3, CSAS-O IS was entered at the first step and BMS was entered at the second step. For mathematics, CSAS-O IS significantly predicted mSGPs, $R^2 = .066$,

$F(1, 64) = 4.54$, $p < .05$, contributing approximately 7% of the variance, and adding BMS explained an additional 18% of the variance, $\Delta R^2 = 18.3\%$, $\Delta F(1, 63) = 15.35$, $p < .001$. However, for ELA, CSAS-O IS did not predict teachers' mSGPs, $R^2 = .00$, $F(1, 76) < 1$, whereas BMS did, yielding an additional 17% of variance explained, $\Delta R^2 = 17.3$, $\Delta F(1, 75) = 15.70$, $p < .001$.

Teacher Ratings Predict Student Growth Beyond School Administrator Ratings

CSAS-T Total, when entered at the second step in Model 1, as shown in Table 5, explained a significant, additional amount of 4.8% of the variance in mathematics beyond that

of CSAS-O Total, $\Delta R^2 = .048$, $\Delta F(1, 63) = 3.868$, $p < .05$. However, in Model 2, CSAS-T did not explain a significant additional amount of variance in ELA, $\Delta R^2 = .025$, $\Delta F(1, 75) = 1.98$, $p > .05$. Similarly in Model 3 and Model 4, CSAS-T IS and BMS did not predict a significant amount of variability in mathematics beyond what was predicted by the CSAS-O IS and BMS; Model 3—IS: $\Delta R^2 = .03$, $\Delta F(1, 62) = 2.58$, $p > .05$; BMS: $\Delta R^2 = .007$, $\Delta F < 1$; and Model 4—IS: $\Delta R^2 = .013$, $\Delta F(1, 74) = 1.20$, $p > .05$; BMS: $\Delta R^2 = .004$, $\Delta F(1, 73) < 1$.

Discussion

Although ESSA (2015) granted states more flexibility in their educator evaluation system design, many states have yet to undo Obama-era legislation that embraced the measurement of qualities of instructional delivery (i.e., processes) and students' gains in achievement (i.e., outputs). As with any teaching enterprise, it is imperative that teacher process components demonstrate relationships with student achievement. This relationship becomes even more imperative as state education agencies move forward under the freedom of ESSA to redesign their evaluation systems and consider abandoning student achievement metrics. This study examined the use of school administrator and teacher self-report assessments of classroom practices in mathematics and ELA SGPs in high-poverty schools. Findings offer initial evidence of the relations of CSAS scores and student achievement on a mandated achievement test.

Findings revealed negative correlations between school administrator and teacher CSAS Rating Scale discrepancy scores and growth scores in mathematics and ELA. This means that teachers' greater need for change in using evidence-based practices was associated with lower achievement growth within a school year. These results were consistent with previous studies examining the relation of CSAS-O and achievement scores (Dudek et al., 2018; Reddy, Fabiano, Dudek, & Hsu, 2013b).

The current study found that the association between school administrators' and teachers' self-reported discrepancy scores and achievement growth was generally more robust for mathematics than for ELA. Although this study found that the observational measures accounted for a modest portion of the variation in teachers' mSGPs in mathematics, that is, 21% in Model 1 and 29% in Model 3, these findings are slightly higher in magnitude to those reported in similar classroom observational research (e.g., McLean, Sparapani, Tostec, & Connor, 2016; Rogosa, 2002).

It was interesting to see that when the CSAS-O ratings were broken down into IS and BMS Total scales, the BMS contributed unique variance (i.e., 18%, 17%) to students' growth in mathematics and ELA, above and beyond the variance explained by the IS scale. One possible explanation for BMS' robust contributions is that challenging

classroom behaviors may negatively affect instructional time, content, and quality of learning opportunities, and subsequently reduce student gains over time (Dolan et al., 1993; Rowan, Camburn, & Correnti, 2004). Research has pointed to the link between high poverty, levels of classroom disruptive behaviors, and lower rates of student academic engagement and learning (Greenwood, 1991).

Findings in this study offer support for the use of teacher self-report assessment in identifying instructional effectiveness and improvements related to student outcomes. The value of self-report as a powerful technique for self-improvement has been noted in the literature—irrespective of its incremental validity (McDonald & Boud, 2003; Ross, Hogaboam-Gray, & Rolheiser, 2002). For example, Ross and Bruce (2007) reported that the use of a teacher self-assessment tool contributed to teacher effectiveness when bundled with other professional growth strategies (i.e., coaching and independent observation). We found that the inclusion of teacher self-assessment of their classroom practices, as measured by the CSAS-T Total, explained significant, additional variance in student growth in mathematics.

Overall, findings from this study suggest that the combined use of school administrator and teacher assessments could offer valid and helpful data for the evaluation of teacher effectiveness and PD supports related to student achievement in high-poverty contexts. Coaching school personnel working in high-stress, high-poverty settings need reliable and valid assessment-driven feedback and support to enhance the instructional environment for students at risk of poor school performance and dropout (Reddy et al., 2017).

Limitations

Findings should be interpreted considering limitations. First, this study aimed to predict teacher practice measurement to the PARCC, which uses SGPs. Although growth scores are meaningful for estimating progress in learning over time, research has indicated SGPs may have large amounts of random error (Monroe & Cai, 2015), and this error may be increased when a small subset of the mSGPs were calculated based on fewer than 20 students. In addition, errors in teacher SGPs may be correlated with average student achievement (Castellano & McCaffrey, 2017), warranting further investigation. Second, our best model only captured 29% of the variance in students' academic growth, which indicates that other factors, such as student self-regulated learning, instructional match, task difficulties, or alignment of curriculum with tested standards, need to be investigated. Third, with the passage of ESSA (2015), several states have moved to abandon student-achievement metrics as part of their evaluation systems and other states have abandoned Common Core Standards and PARCC

testing all together. Therefore, the widespread relevance of the PARCC assessment as a metric for student achievement may be limited in the current climate of change to the states that continue to implement this assessment. Subsequently, the generalizability of the current study's results may be limited to PARCC states only. Fourth, teachers were asked to self-assess their classroom practices with the CSAS-T only for specific lessons observed by their supervisor. It is unclear whether teacher self-assessment on the CSAS-T influenced their use of classroom practices during subsequent observations conducted by school administrators. Finally, participant characteristics may limit generalizability of findings to other states, populations, and contexts. For example, school administrators and teachers were predominantly female White from high-poverty charter schools in New Jersey. Despite this limitation, characteristics of teachers were comparable with the school personnel characteristics reported by the state of New Jersey.

Future Directions for Research and Practice

The present investigation is the first to examine the relation of the school administrator and teacher CSAS discrepancy scores on student growth on state-mandated mathematics and ELA testing in high-poverty schools. Specifically, this study examined the quality of instructional and behavioral management practices on student achievement growth on a mandated state assessment (PARCC). Thus, further investigations on instructional conditions (opportunities to learn, task difficulty), classroom- and/or student-related factors (classroom management skills, level of disruption, and student academic engagement) in relation to student growth in achievement are warranted.

For example, we found with this sample that teachers' IS and BMS scores differentially predicted student gains in mathematics and ELA. This finding is in keeping with conclusions reached by researchers in earlier studies in which the relationship between teacher practices and student learning was mediated by student behavior (McLean et al., 2016; Ponitz, Rimm-Kaufman, Grimm, & Kurby, 2009); yet the ways in which teacher practices, as measured by the CSAS, relate to students' classroom behavior appear to be complex, as qualities of instructional strategies have also been associated with students' engagement in learning activities (e.g., Lekwa, Reddy, & Shernoff, 2019). The results of this and similar studies on assessment of teacher practices underscore the value of using distinct measures of behavior management and instructional strategies for the purposes of guiding coaching programs. Based on these results, it appears feasible that there are circumstances in which teacher coaching that is intended to boost academic achievement should target behavior management first, but future research needs to determine what those circumstances are.

Moreover, questions remain about the degree to which curricular content and teacher practices contribute separately to student learning. Results showed that student gains in mathematics appeared to relate more strongly to teachers' strategy use than students' gains in ELA. Although the IS and BMS scales within the CSAS represent distinct groups of strategies used by teachers, they were designed to be agnostic of course content or student grade level; there is no a priori reason to expect that either or both scales might relate differently to separate academic subjects. This result therefore warrants further investigation.

Several additional implications of these results for future research and practice are also of note. First, future validity research on the CSAS-O and CSAS-T is warranted to include more culturally diverse populations of school administrators, teachers, and students. Second, because the CSAS-O and CSAS-T were designed to be used in conjunction with each other for promoting PD conversations, more investigations are warranted to examine the relationships of both school administrator and teacher input (discrepancy scores) on student achievement. Third, research that examines how school administrator and teacher ratings of classroom practices inform instructional coaching actions and decisions (i.e., identifying practice needs, setting goals, implementation plans of action and monitoring progress toward goals) and influence educational outcomes (student engagement, achievement and social behavior) would be beneficial. Finally, studies that examine the convergent and discriminant validity of the CSAS-O and CSAS-T with other known teacher observation instruments (i.e., Framework for Teaching; Danielson, 2013) may inform new multi-method teacher evaluation approaches.

Conclusion

This study examined the relation of school administrator and teacher assessments of instruction *and* classroom management practices on student achievement in high poverty settings. Overall, findings offer some evidence of score inferences to achievement, highlighting the utility of school administrator and teacher ratings as a complimentary assessment for evaluation and PD decision making. Findings indicate stronger utility of the CSAS school administrator and teacher discrepancy scores for students' mathematics performances, compared with ELA performance in high poverty settings. Overall, findings underscore teacher self-assessment of their own classroom practices may provide additional understanding of the variations in student achievement.

Authors' Note

The positions and opinions expressed in this article are solely those of the authors.

Declaration of Conflicting Interests

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


Funding

The authors received the following financial support for the research, authorship, and/or publication of this article: The current study was implemented as part of the School System Improvement (SSI) Project, a collaboration between multiple universities and charter schools funded by the U.S. Department of Education's Office of Innovation and Improvement as part of the Teacher Incentive Fund program (awarded to Rutgers, The State University of New Jersey; #S374A120060).

ORCID iDs

Linda A. Reddy  <https://orcid.org/0000-0001-8314-2810>

Christopher M. Dudek  <https://orcid.org/0000-0002-8338-4489>

Ryan Kettler  <https://orcid.org/0000-0003-4395-654X>

References

- Belfiore, P. J., Auld, R., & Lee, D. L. (2005). The disconnection in poor urban education: Equal access and a pedagogy of risk-taking. *Psychology in the Schools, 42*, 855–863.
- Betebenner, D. W. (2011). *An overview of student growth percentiles*. Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved from https://www.state.nj.us/education/njsmart/performance/SGP_Detailed_General_Overview.pdf
- Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research, 78*, 367–409.
- Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2012). Recruiting effective math teachers, evidence from New York City. *American Educational Research Journal, 49*, 1008–1047.
- Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research in teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.
- Castellano, K. E., & McCaffrey, D. F. (2017). The accuracy of aggregate student growth percentiles as indicators of educator performance. *Educational Measurement: Issues and Practice, 36*, 14–27.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- National Governors Association. (2010). Common core state standards. *Washington, DC*.
- Danielson, C. (2013). The framework for teaching evaluation instrument, 2013 instructionally focused edition. Retrieved from <http://www.loccsd.ca/~div15/wp-content/uploads/2015/09/2013-framework-for-teaching-evaluation-instrument.pdf>.
- Darling-Hammond, L. (2003). Keeping good teachers: Why it matters, what leaders can do. *Educational Leadership, 60*, 6–13.
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 4*, 597–604.
- Desimone, L., Smith, T., & Frisvold, D. (2010). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24*, 267–329.
- Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., . . . Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14*, 317–345.
- Dudek, C. M., Reddy, L. A., & Lekwa, A. (2018). Measuring teach practices to inform student achievement in high poverty schools: A predictive validity study. *Journal of Contemporary School Psychology, 2018*, 1–14. doi:10.1007/s40688-018-0196-8
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford press.
- Espinosa, L. M. (2005). Curriculum and assessment considerations for young children from culturally linguistically, and economically diverse backgrounds. *Psychology in the Schools, 42*, 837–853.
- Evaluation scoring. (n.d.). Retrieved from <https://www.state.nj.us/education/AchieveNJ/resources/scoring.shtml>
- Every Student Succeeds Act of 2015, 114 U.S.C. § S1177. (2015).
- Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring and reprimands revisited. *Intervention in School and Clinic, 44*, 195–205.
- Greenwood, C. R. (1991). Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children, 57*, 521–535.
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. New York, NY: Routledge.
- Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. New York, NY: Routledge.
- Kerns, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools, 44*, 65–75.
- Klein, A. (2017a, November 14). Are states changing course on teacher evaluation? *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2017/11/15/are-states-changing-course-on-teacher-evaluation.html>
- Klein, A. (2017b, November 28). What's the future of teacher evaluation in the ESSA era? *Education Week*. Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2017/11/essa_evaluation_teacher_test_scores.html
- Koziol, S. M., & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research, 79*, 205–209.
- Krei, M. S. (1998). Intensifying the barriers: The problem of inequitable teacher allocation in low-income urban schools. *Urban Education, 33*, 71–94.
- Langford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*, 37–62.

- Lekwa, A. J., Reddy, L. A., & Shernoff, E. S. (2019). Measuring teacher practices and student academic engagement: A convergent validity study. *School Psychology Quarterly, 34*, 109–118. doi:10.1037/spq0000268
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education, 10*, 209–220.
- McGuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher-evaluation systems*. Center for American Progress. Retrieved from <https://www.americanprogress.org/issues/education-k-12/reports/2012/11/13/44494/the-state-of-teacher-evaluation-reform/>
- McLean, L., Sparapani, N., Tostec, J. R., & Connor, C. M. (2016). Classroom quality as a predictor of first graders' time in non-instructional activities and literacy achievement. *Journal of School Psychology, 56*, 45–58. doi:10.1016/j.jsp.2016.03.004
- Monroe, S., & Cai, L. (2015). Examining the reliability of student growth percentiles using multidimensional IRT. *Educational Measurement: Issues and Practice, 34*(4), 21–30.
- National Council on Teacher Quality. (2015). *State of states: Evaluating teaching, leading, and learning*. Washington, DC: Author. Retrieved from <http://www.nctq.org/>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425. (2002).
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237–257.
- Partnership for Assessment of Readiness for College and Careers. (2018). *Final technical report for 2017 administration*. Retrieved from <https://parcc-assessment.org/wp-content/uploads/2018/03/PARCC-2017-Technical-Report-Final-03162018.pdf>
- Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K., & Curby, T. W. (2009). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review, 38*, 102–120.
- Reddy, L. A., & Dudek, C. M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology, 2*, 71–84.
- Reddy, L. A., Dudek, C. M., Fabiano, G., & Peters, S. (2015). Measuring teacher self-report on classroom practices: Construct validity and reliability of the classroom strategies scale—Teacher form. *School Psychology Quarterly, 30*, 513–533.
- Reddy, L. A., Dudek, C. M., & Lekwa, A. J. (2017). Classroom strategies coaching model: Integration of formative assessment and instructional coaching. *Theory into Practice, 56*, 46–55.
- Reddy, L. A., Dudek, C. M., & Shernoff, E. S. (2016). Teacher formative assessment: The missing link in response to intervention. In *Handbook of response to intervention* (pp. 607–623). New York, NY: Springer.
- Reddy, L. A., Fabiano, G., Dudek, C. M., & Hsu, L. (2013a). Development and construct validation of the classroom strategies scale-observer form. *School Psychology Quarterly, 28*, 317–341.
- Reddy, L. A., Fabiano, G., Dudek, C. M., & Hsu, L. (2013b). Predictive validity of the classroom strategies scale-observer form on statewide testing. *School Psychology Quarterly, 28*, 301–316.
- Reddy, L. A., Hua, A., Dudek, C. D., Kettler, R. J., Lekwa, A., Arnold-Berkovits, I., & Crouse, K. (2019). Use of observational measures to predict student achievement. *Studies in Educator Evaluation, 62*, 197–208.
- Rogosa, D. R. (2002). *Irrelevance of reliability coefficients to accountability systems: Statistical disconnect in Kane-Staiger "volatility in school test scores" CRESST deliverable*.
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education, 23*, 146–159.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5–6 mathematics: Effects on problem solving achievement. *Educational Assessment, 8*, 43–58.
- Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal, 105*, 75–102.
- Sawchuk, S. (2016, January 5). ESSA loosens reins on teacher evaluations, qualifications. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2016/01/06/essa-loosens-reins-on-teacher-evaluations-qualifications.html?cmp=e-ml-enl-eu-news1>
- Stormont, M. (2007). *Fostering resilience in your children at risk for failure: Strategies for grades K–3*. Upper Saddle River, NJ: Pearson.
- Teacher Incentive Fund (2012). United States Department of Education. Retrieved October 10, 2011 at <http://www.ed.gov/programs/teacher-incentive/index.htm>
- U.S. Department of Education. (2009). Overview information; Race to the top fund; Notice inviting applications for new awards for Fiscal Year (FY) 2010. *Federal Register, 74*, 59836–59872.
- Walberg, H. J. (1986). Synthesis of research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 214–229). New York, NY: Macmillan.