Hammill Institute on Disabilities

Assessment for Effective Intervention
2019, Vol. 44(2) 81–94

DOI: 10.1177/1534508418772919
aei.sagepub.com
SAGE

Improving Universal Classroom Practices Through Teacher Formative Assessment and Coaching

Christopher M. Dudek, MEd[1], Linda A. Reddy, PhD[1], Adam Lekwa, PhD[1], Anh N. Hua, PhD[1], and Gregory A. Fabiano, PhD[2]

Abstract

This article presents the Classroom Strategies Coaching (CSC) Model, a data-driven coaching approach that uses teacher formative assessment data to drive improvements in universal practices. The classroom strategies assessment system (CSAS), a formative assessment of evidence-based instructional and behavioral management practices was used to facilitate the coaching process. Results from 32 elementary school teachers who received brief coaching after participating as waitlist controls in a randomized controlled trial are presented. Teachers' practices remained stable across baseline periods. Following coaching, teachers displayed improvements toward their behavioral management goals (e.g., *d*s = .50–.83). Results also showed meaningful reductions in the overall need for change in instruction (*d* = .88) and in behavior management practices (*d* = .68) at postintervention. Findings illustrate the benefits of integrating teacher formative assessment in coaching to improve teaching practices. Implications for practice and future directions are outlined.

Keywords

classroom formative assessment, instructional coaching, instruction and behavior management practices

Teachers' professional development (PD) is a key factor for education quality in schools (Darling-Hammond, 1993; Desimone, 2009) and is a critical mediator in the effectiveness of policies related to teaching practice and student-related outcomes (Desimone, Smith, & Frisvold, 2007). Yet, one-shot traditional workshops, the most common form of PD, have been found to be largely ineffective, often leading to limited skill transfer and subsequently poor student outcomes (Bush, 1984; Darling-Hammond, Chung Wei, Andree, Richardson, & Orphanos, 2009; Gulamhussein, 2013). The failure of workshop-style models stems from their inability to help teachers with two critical aspects of learning new skills: engaging in implementation and receiving support during implementation (Carlisle & Berebitsky, 2011; Gulamhussein, 2013). Workshop models assume the difficulty in learning new skills is a lack of knowledge (Gulamhussein, 2013). However, for teachers the struggle is not learning a new skill, but transferring that knowledge and implementing the skill in their classroom (Fullan, 2001; Gulamhussein, 2013; Joyce & Showers, 1982). Teachers' initial efforts when implementing a new teaching skill are often met with failure, which can lead to frustration, abandonment of the new skill, and a return to business as usual (Ermeling, 2009; Fullan, 2001). These initial struggles are further complicated by the fact that teachers only change their beliefs and subsequent actions after success with students is evident (Guskey, 1984, 2002).

French (1997) concluded that teachers may require up to 50 hours to master and implement a new skill successfully. Other researchers have similarly corroborated that teachers need between 50 and 80 hours (Banilower, 2002; Yoon, Dunca, Wen-Yu Lee, Scarloss, & Shapley, 2007) to develop mastery. Skill mastery thus requires significant time and practice despite initial challenges and failed attempts. Similarly, research has shown that the provision of ongoing supports during initial implementation stages, such as observation, practice, and performance feedback, can facilitate mastery of new teacher skills (Riley-Tillman & Ecker, 2001; Scheeler, Rhul, & McAfee, 2004; Solomon, Klein, & Politylo, 2012). Teachers provided with these kinds of supports during implementation have a 95% chance of skills transfer to the classroom (Truesdale, 2003). In sum, for PD

[1]Rutgers, The State University of New Jersey, Piscataway, USA
[2]University at Buffalo, NY, USA

Corresponding Author:
Christopher M. Dudek, School System Improvement Project, Rutgers, The State University of New Jersey, 41 Gordon Road, Room Suite C, Piscataway, NJ 08854, USA.
Email: cdudek@gsapp.rutgers.edu

to be effective, it must occur over time, be ongoing, and offer opportunities for support during the process (Gulamhussein, 2013).

A growing body of literature supports instructional coaching as an effective form of PD for enhancing teacher practices and student outcomes (e.g., Briere, Simonsen, Sugai, & Myers, 2015; Denton & Hasbrouck, 2009; Dufrene, Lestremau, & Zoder-Martell, 2014; Knight, 2007). Instructional coaching often aims to enhance existing skills or develop new skills that lead to improved practices. Instructional coaching actions such as observing, modeling, practicing, and providing ongoing performance feedback have been found effective for promoting teachers' skill development and implementation (e.g., Becker, Darney, Domitrovich, Keperling, & Ialongo, 2013; Joyce & Showers, 2002; Kretlow & Bartholomew, 2010; Marzano & Simms, 2013). In contrast to workshop-model forms of PD, instructional coaching reflects principles of active learning approaches (e.g., modeling, practice, and feedback) to facilitate teacher skill development (Desimone, 2009). Instructional coaching is a job-embedded support specific to teachers' classroom contexts, allowing for focus on immediate teacher, student, and classroom challenges. This enhances the relevance of instructional coaching, uptake of teacher learning, and promotes transfer of skills into the classroom (e.g., Carlisle & Berebitsky, 2011; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Riley-Tillman & Ecker, 2001; Salas, Tannenbaum, Kraiger, & Smith-Jentsch, 2012).

However, current instructional coaching models vary in foci and efficacy. Many instructional coaching models focus on individual students instead of classroom-wide ecology (Sheridan, Welch, & Orme, 1996), which makes it difficult to generalize to other students or situations (Coffee & Kratochwill, 2013; Riley-Tillman & Ecker, 2001). For example, the functional assessment of academic behavior (Ysseldyke & Christenson, 2002) focuses on individual students with learning problems whereas consultation and coaching approaches making use of schoolwide positive behavior supports programs address systems-level change (Horner, Sugai, & Anderson, 2010; Sugai, Horner, & Gresham, 2002). In contrast, the Classroom Check Up developed by Reinke, Lewis-Palmer, and Merrell (2008) addressed classroom-level system changes. Most instructional coaching models focus on single targets for change and neglect to address improvements in both instruction and behavior management (Dufrene et al., 2014; Reddy & Dudek, 2014).

The Classroom Strategies Coaching (CSC) Model

To this end, we describe the CSC Model, to demonstrate the utility of a brief data-driven coaching approach that centers on using teacher formative assessment to enhance evidence-based classroom practices. The instructional coaching process, theoretical background, and core components of the CSC Model are described. As a demonstration of the CSC's effectiveness for changing teachers' classroom practices, we present findings from a follow-up study on 32 teachers that received CSC after participating as waitlist controls in a previous randomized controlled trial (RCT) examining the efficacy of the CSC Model (see Fabiano, Reddy, & Dudek, in press; Reddy, Dudek, & Lekwa, 2017).

The CSC Model reflects principles of behavioral consultation frameworks (e.g., Bergan & Kratochwill, 1990; Sheridan & Kratochwill, 2008) and incorporates active learning techniques from systems and social learning theories (e.g., Bandura, 1977; Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001). In this model, teachers are viewed as active collaborators and decision makers throughout the instructional coaching process focused on teachers' needs. Coaches move beyond indirect service provider roles and are active agents who frequently maintain a presence in the classroom by conducting classroom observations. Coaches model how to use specific instructional and classroom behavioral management strategies (BMSs), provide practice opportunities for teachers in sessions, and provide feedback from the ongoing classroom observations' data.

The CSC Model focuses on helping teachers develop both evidence-based instructional and BMSs, commonly considered as universal teaching practices from a multitiered system of support perspective (Reinke, Herman, & Stormont, 2013). To identify and shape effective teaching strategies, the CSC Model uses a classroom observation assessment, the Classroom Strategies Assessment System (CSAS; Reddy & Dudek, 2014), a multirater classroom assessment system that measures teachers' use of evidence-based instructional and classroom BMSs (e.g., Gable, Hester, Rock, & Hughes, 2009; Harris, 1998; Hattie, 2009; Kerns & Clemens, 2007; Marzano, Pickering, & Pollock, 2001). Although recently developed, the CSAS and CSC Model target teacher classroom practices that are supported by decades of research (e.g., Gable et al., 2009; Harris, 1998; Hattie, 2009; Kerns & Clemens, 2007; Marzano et al., 2001). Targeted instructional strategies (ISs) are grounded in the effective teaching model literature such as direct instruction, differentiated instruction, and constructivist models (Harris, 1998; Tomlinson & McTighe, 2006). The BMSs represent well-established prevention and antecedent approaches for managing students' classroom behavior, behavioral reinforcement, classroom routines, and rules (Acker & O'Leary, 1987; Alberto & Troutman, 2008; Gable et al., 2009; Kounin, 1970; O'Leary, Kaufman, Kass, & Drabman, 1970). School personnel can use the CSAS to (a) measure how often teachers use empirically supported instructional and BMSs, (b) identify strengths

and areas for improvement, (c) formulate practice goals, and (c) monitor progress toward goals.

CSC Model Process

The CSC Model is carried out in phases that enable the identification of practice needs and goals, development of implementation plans matched to instructional and behavioral management practice needs, monitoring plan implementation, and finally evaluation of progress toward goals. *Phase 1* focuses on establishing rapport and the identification of instructional and behavior management strategies to target throughout coaching. During *Phase 1*, coaches review baseline data on the CSAS with teachers and collaboratively identify which of the eight strategy counts should be selected for intervention. *Phase 2* then encompasses development of implementation plans based on the targeted strategies. Coaches define and model the first set of targeted strategies (either instructional or BMSs). Teachers practice during the coaching session and begin implementing in their classrooms. Meanwhile, coaches begin conducting ongoing classroom observations between meetings with the CSAS to provide data-informed feedback to teachers. *Phase 3* primarily focuses on defining and modeling the second set of targeted strategies (for example, if ISs were discussed earlier, then behavior management would be the focus). *Phase 4* focuses on methods to sustain strategy improvements through discussions on preplanning lessons to make use of the targeted strategies and use of recall strategies and techniques (e.g., penny in the pocket; visual cues and reminders).

Current Study

The current study presents findings from 32 elementary school teachers who received instructional coaching after participating as waitlist controls in the first RCT that examined the efficacy of the CSC 4 Session Model (see Fabiano et al., in press; Reddy et al., 2017). Compared to the first RCT that examined between-group differences, we examined within-group effects for independently observed changes in instruction and behavior management practices following the CSC intervention in the waitlist controlled teachers. The current study serves as an illustrative example of an assessment-driven coaching model that targets universal classroom practices. In addition, this study further tests the efficacy of the CSC Model with general education teachers in elementary school and findings from this study can provide insights for iterative refinement of the CSC Model dosage, components, and processes (Shernoff, Lekwa, Reddy, & Coccaro, 2017). Iterative development design research prior to and following quasi-experimental and experimental design investigation helps spotlight active ingredients and processes of promising interventions (e.g., Hoagwood, Burns, & Weisz, 2002; Shernoff, Lekwa, Reddy & Coccaro, 2017).

In the previous RCT, the CSC 4 Session Model focused on enhancing teachers' use of eight teaching strategies related to instruction and behavioral management. Compared to the waitlist controls, teachers receiving immediate coaching evidenced significant improvements in their usage of behavior management strategies (i.e., improved rates), along with a reduced need to make changes in behavior management practices (i.e., lower rating scale scores) as measured by the CSAS. In addition, high ratings of satisfaction were reported by teachers receiving immediate coaching. The current study extends this work by investigating the outcomes of the same CSC 4 Session Model provided to the 32 teachers serving as waitlist controls. We examined within-group effects for independently observed changes in the same eight teaching strategies related to instruction and behavior management practices following the CSC intervention in the waitlist controlled teachers. The current study addressed the following research questions:

Research Question 1: Do waitlist control teachers receiving delayed CSC evidence improvements (i.e., improved rates) in eight teaching strategies related to instruction behavior management as measured by CSAS?
Research Question 2: Do waitlist control teachers receiving delayed CSC evidence reduced need for changes (i.e., lower rating scale scores) in instruction and behavior management strategies as measured by the CSAS?
Research Question 3: Do waitlist control teachers rate the CSC as highly satisfactory as a PD support?

Based on the outcomes from the previous RCT, we hypothesized that waitlist control teachers would evidence improvements (i.e., improved rates) in using the targeted behavior management strategies of the CSAS. We also hypothesized that waitlist control teachers would exhibit a reduced need for change (i.e., lower rating scale scores) in both instructional and BMSs on the CSAS. Finally, we hypothesized that waitlist control teachers would report high satisfaction with the CSC intervention.

Method

Participants

The sample included 32 elementary school teachers across kindergarten through Grade 5 from New Jersey and New York. Teacher participants were predominantly female (96.8%) and Caucasian (96.7 %) ranging in age from 23 to 62 years (M = 37.94, SD = 11.78). The average number of years of teaching experience was 10.74 (SD = 7.18).

Table 1. Descriptions of the CSAS Strategy Counts.

Name	Definition
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.
Behavioral 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. Only the strategy counts were selected as targets for change by teachers in coaching. CSAS = Classroom Strategies Assessment System.

A total of 18 CSC and CSAS trained and supervised graduate students served as coaches. Coaches were predominantly female (83%) and Caucasian (100%) and had a mean age of 24.7 years (*SD* = 6.65). A total of 13 CSAS trained and supervised independent observers (blind to teachers' assignment within the RCT) performed repeated baseline and postintervention observations in the study. Independent observers were undergraduate and graduate students in psychology and education programs, predominantly female (62%) with a mean age of 25.25 (*SD* = 6.93).

Measures

CSAS. The CSAS-observer form (CSAS-O) was used as formative assessment to measure changes in teachers' usage of evidence-based instructional and behavioral management practices (Reddy & Dudek, 2014; Reddy, Fabiano, Dudek, & Hsu, 2013a, 2013c). The CSAS-O contains three sections, which observers complete as part of the classroom observation process. The *Strategy Counts* tally the frequency of eight discrete teaching behaviors (see Table 1). The *Strategy Rating Scales* assess how the teacher used specific strategies during the observed lesson using a 7-point Likert-type scale. The rating scales are composed of five subscales related to ISs and four subscales related to BMS; see Table 2). Finally, the *Classroom Checklist* measures the presence of key classroom structural elements and procedures (e.g., classroom rules are posted). Results on the Classroom Checklist were not the focus of the instructional coaching and are not included in the present study.

The Strategy Counts are completed during the observation period. During the observation, observers also take notes related to the Strategy Rating Scales' nine dimensions, the lesson content, activities, and students. After the observation, observers use the strategy counts and their notes to facilitate completion of the Strategy Rating Scales. To complete the Strategy Rating Scales, observers first rate how often (*observed frequency rating;* 1 = *never used*, 3 = *sometimes used*, 7 = *always used*) the teacher used specific strategies in each subscale, and then, observers rate how often the teacher *should have* used each strategy (*recommended frequency*) on the same 7-point Likert-type scale.

The *observed frequency* and *recommended frequency* ratings produce a third score, the discrepancy score, which is calculated by subtracting the observed frequency rating from the recommended frequency rating of each item. Large scores in either direction (positive or negative) suggest a greater need for improvement in teaching practices (i.e., increase or decrease in use) and smaller scores indicate minimal or no need for change in teaching practices. In the current study, absolute value discrepancy scores were used: |recommended frequency—observed frequency|. Discrepancy scale scores were created by summing the respective absolute value item scores together. Thus, larger scale scores indicated greater need for change independent of direction.

Table 2. Descriptions of the *CSAS Strategy Rating Scales*.

Scale name	Definition
ISs total	The total IS scale reflects the overall use of instructional methods and academic monitoring/feedback.
Adaptive instruction	Strategies teachers use to respond to their students' learning needs while teaching. These practices reflect teacher flexibility and responsiveness to students' needs, as well as methods of differentiated instruction.
Student-directed instruction	Strategies teachers use to actively engage students in the learning process. These practices encompass constructivist and hands-on instructional techniques, linking lesson content to prior learning, personal experiences, and cooperative learning.
Direct instruction	Strategies teachers use to deliver academic content or convey information to students. These practices include direct instruction techniques, modeling, identifying, and summarizing.
Promotes students' thinking	Strategies teachers use to activate students' thinking about the lesson material. These practices assess teachers' efforts to get their students to think about their thinking process (i.e., open-ended, what, how, why).
Academic performance feedback	Strategies teachers use to provide specific feedback to their students on their understanding of the material. These practices assess teachers' efforts to explain what is correct or incorrect with student academic performance. These practices also measure teachers' efforts to reinforce (i.e., praise) students learning.
BMSs total	The total BMS scale reflects the overall use of prevention methods and behavior feedback.
PMs	Verbal and nonverbal strategies teachers use to prevent student disengagement, and problem behaviors from occurring in classroom. These practices assess how teachers create a positive classroom environment.
Directives	Strategies teachers use for issuing directions or instructions to students and behavioral expectations in the classroom.
Praise	Verbal and nonverbal strategies teachers use to positively reinforce specific appropriate behaviors in the classroom. These practices assess how teachers respond to positive behavior in the classroom.
Corrective feedback	Verbal and nonverbal strategies teachers use to correct students' inappropriate behavior. These practices asses how teachers respond to negative behavior in the classroom.

Note. IS = instructional strategies; PM = proactive methods; BMS = behavior management strategies; CSAS = Classroom Strategies Assessment System.

The CSAS-O evidences good reliability, content, construct, and predictive validity (e.g., Reddy et al., 2013a, 2013b, 2013c). High levels of internal consistency have been documented across all three sections (Cronbach's $\alpha > .90$) and fair to good test–retest reliability across a 2 to 3 week span ($r > .70$). For inter-rater reliability, acceptable estimates have also been found for strategy counts ($r = .94$), the IS and *BMS Strategy Rating Scales* ($rs = .80$ and .72, respectively), and the classroom checklist ($r = .86$). Differential item functioning analyses have demonstrated that IS and BMS Strategy Rating Scale items are free of item bias for teacher age, educational degree, and years of teaching experience. In addition, the CSAS-O has demonstrated convergent and discriminant validity with other classroom observational assessments, such as the Classroom Assessment Scoring System (Reddy, Fabiano, & Dudek, 2012) and the Danielson Framework for Teaching (Dudek, Kettler, Kurz, & Reddy, under review). The CSAS-O has also demonstrated predictive validity with students' academic outcomes on statewide testing (Dudek, Reddy, & Lekwa, in press; Reddy et al., 2013c) and the Measures of Academic Progress (MAP; Northwest Evaluation Association [NWEA], 2011; see Lekwa, Reddy, Dudek, & Hua, in preparation).

Teacher Coaching Evaluation Scale (TCES). Teachers completed the TCES (Reddy, Fabiano, & Dudek, 2012), a 14 item, 7-point Likert-type scale assessing teachers' satisfaction with the CSC instructional coaching process. The TCES items relate to the coaches' competencies with classroom intervention strategies, applicability, and usefulness of the interventions, teachers' competence using the strategies, the stress demands of the intervention strategies, and teachers' evaluation of progress and outcome (see Table 4). Teachers rate their agreement to each item on a scale of 1 (*strongly disagree*) to 7 (*strongly agree*) for each item and the 14 items scores are summed to produce a total score representing overall satisfaction with the CSC coaching intervention. In the previous RCT, the TCES demonstrated good internal consistency ($\alpha = .97$, $SE \pm 3.58$) and acceptable item-to-total correlations (exceeding .60). In the current study ($N = 32$ teachers), the TCES demonstrated good internal consistency ($\alpha = .91$) and comparable item-to-total correlations.

Procedure

Observer training. Independent observers and coaches were systematically trained as observers on the CSAS-O. Authors

of the CSAS provided training directly to observers and coaches over 3 full days. Training included several didactic sessions with discussion of theory and evidence of the CSAS dimension structure, along with video-based coding practice. The CSAS authors provided ongoing feedback to the observers during video practice sessions about their understanding of the CSAS content, observation skills, and application of the CSAS to video-based practice. All observers were required to engage in a criterion reliability certification testing (i.e., 80% reliability threshold) prior to observing teachers.

At the start of the study, all observers, coaches, teachers, and school leaders signed a ground rules agreement form to prohibit sharing of instructional coaching information within their schools until the study and school year were completed. As per the RCT procedures, the sample of 32 teachers was randomly assigned to a waitlist control group following informed consent procedures. Random assignment was stratified within school to equally distribute each condition within school buildings.

Coach training. In addition to participating in CSAS observer training, coaches received extensive training and weekly supervision on the CSC 4 Session Model from faculty members in school psychology. Coaches were school psychology doctoral students who completed behavior consultation courses prior to the start of coaching. Coaches were trained on using a manualized approach, which included detailed coaching session procedural checklists and forms. Coaching training consisted of a full-day workshop focused on the theory and evidence of the CSC components and delivery of session-specific procedures such as data review/interpretation, modeling of strategies, group practice, effective delivery of performance feedback, evaluation of implementation fidelity and progress toward goals. Adherence to the manualized model was assessed through the independent review of audiotaped sessions for each coach (see the "Results" section).

Baseline procedures. The initial baseline assessment consisted of demographics and classroom characteristic forms, as well as initial baseline CSAS administration. Independent observers, who were unaware of the study condition to which the teacher was assigned, observed with the CSAS-O for two different classroom lessons within 7 days of each other. Each observation was 30 min in length and occurred during times of active instruction. Scores were averaged across the two observations to determine the first baseline CSAS-O scores for teachers. The waitlist control teachers in the current sample then engaged in routine teaching for 4 weeks and during this time, participants did not interact with study staff. Approximately 5 weeks after the first baseline assessment, waitlist control teachers received a second round of CSASs, serving as a second baseline assessment. Thus, the 32 control teachers received two baseline CSAS-O administrations (a total of four CSAS observations).

Instructional coaching intervention procedures. The CSC 4 Session Model began for the current sample of teachers following the second baseline administration of the CSAS-O. Instructional coaching consisted of four 30-min sessions once per week over the course of approximately 4 weeks. All instructional coaching sessions were audiotaped to monitor intervention fidelity. Prior to the first meeting, consultants reviewed CSAS-O data gathered by independent observers and CSAS-T self-report data at baseline to facilitate potential targets for teacher change. The instructional coaching intervention then followed the aforementioned CSC 4 Session Model procedures. To counter the potential effects of goal-presentation order, the focus on instructional versus behavioral management goals was counterbalanced for Sessions 2 and 3 across participants. Coaches' observations between meetings occurred during the same lesson content areas observed by the independent observers during the baseline administration. Session 4 included a review of the teachers' progress and development of sustainability plans.

Postintervention procedures. Following completion of the fourth coaching session, the 32 teachers in this study were independently observed using the CSAS-O for the same classroom lessons observed during the baseline periods and coaches' observations. Teacher participants were asked to complete the CSAS-T along with the TCES.

Independent Variables

Research Question 1. The primary outcomes measured were standardized change scores between baseline assessment and postintervention assessment of the strategies (practice goals) selected by the teacher and coach as the focus of the CSC intervention. In the current study, teachers' targeted strategies for improvement (practice goals) were the Strategy Counts of the CSAS-O. During the first coaching session, teachers and coaches collaboratively selected two to four practice goals from the strategy counts. Teachers were asked to select at least one strategy from the IS Counts, and one strategy from the BMS Counts. The Strategy Counts change scores between baseline and postintervention assessment were standardized to account for the distributions and variability of goals selected among teachers.

Research Questions 2 and 3. Secondary outcomes included the Strategy Rating Scales. Specifically, we examined the total discrepancy scores of the IS and BMS rating scales, which represent a need for change in teaching practices respective to instructional or BMSs. Although not the direct focus of the CSC intervention in this study, the IS and BMS

rating scales discrepancy scores contain items that relate to the Strategy Counts, as well as items capturing other non–discretely counted instructional and BMSs that may have been affected by the CSC intervention. Subsequently, these scores were not standardized as they were not the targeted strategies (practice goals) of the CSC intervention.

Data Analysis

Procedural fidelity. Approximately 30% of instructional coaching sessions were audiotaped, reviewed by supervisors, and independently checked for procedural fidelity. Supervisors marked procedures that were completed during each coaching session. The percent of procedures completed within session were computed and averaged across the four coaching sessions for the teacher case. Across the 32 teacher cases (four sessions each), the average procedural fidelity was 98%.

Observer reliability. Interobserver reliability was computed as part of the previous RCT (Fabiano et al., in press). A total of 92 of the baseline- and postintervention CSAS-O observations were utilized to calculate intraclass correlations (ICCs) with fixed measure effects and random observer effects. For the strategy counts, high levels of interobserver agreement for the ICCs (range = .77–.97) were observed and mean differences between observer ratings were also nonsignificant in paired-sample *t* tests ($p > .05$). ICCs for the Strategy Rating Scale discrepancy scores were moderate for IS and BMS (ICC = .55 and .59, respectively; Cicchetti, 1994), although no significant differences in mean scores were found using paired-sample *t* tests ($p > .05$).

Aggregate scoring procedures. In the current study, six CSAS administrations (two per time period) were collected across three time periods (i.e., Baseline 1, Baseline 2 [5 weeks later], and postintervention [5 weeks later]). CSAS Strategy Counts totals and Strategy Rating Scales were first calculated separately for each observation. CSAS scores were averaged across observations within each time period, generating a single set of scores per teacher for each of the three time points. CSAS scores from Baseline 1 and Baseline 2 were then averaged together for the waitlist control teachers. As previously noted, the two IS and BMS Rating Scales discrepancy scores were computed by subtracting the observed frequency from the recommended frequency and taking the absolute value of the difference: | recommended frequency—observed frequency ratings .| Discrepancy scores were first calculated at the item level and then summed to create discrepancy scale scores for the IS and BMS totals. We then followed the aforementioned aggregate procedures for combining across observations within time period, generating a single set of discrepancy scores per teacher for each of the three time points. Discrepancy scores from Baseline 1 and Baseline 2 were then averaged together for the waitlist control teachers.

Standardization method. The primary outcomes for this study, the change scores between baseline assessment and postintervention assessment of the Strategy Counts strategies, were standardized to account for the distributions and variability of goals selected among teachers. For example, for IS Counts, the most frequently identified instructional goals were concept summaries (i.e., 20 times) and academic praise (i.e., 10 times), followed by academic response opportunities (i.e., 9 times) and academic corrective feedback (i.e., 7 times). For BMS Counts, the most frequently identified goals included behavioral praise (i.e., 19 times) and behavioral corrective feedback (i.e., 12 times), followed by clear directives (i.e., 7 times) and vague directives (i.e., 5 times). We standardized the change scores by subtracting the postintervention score from the baseline mean score and dividing by the baseline standard deviation for each individual target behavior. For vague directives, teacher goals focused on reducing the use of this strategy and thus, this strategy was reverse coded to make the interpretation of this score consistent with that of the other standardized scores, with positive change scores representing improvement. The standardized change scores for each target were then averaged across targets if more than one was identified. Postintervention results were then analyzed using a one-sample *t* test (two-tailed) and Cohen's *d* (effect sizes; Cohen, 1988). Effect sizes of .20 to .49 were small, .50 to .79 medium, and .80 and greater were large (Cohen, 1988).

Results

Baseline Descriptives

Table 3 presents the means, standard deviations, and effect sizes for baseline and postintervention assessments for strategy counts and IS and BMS Rating Scales. Prior to receiving instructional coaching, teachers were independently observed on average using concept summaries 1.8 times (range = 0–5.75), academic response opportunities 43.96 times (range = 11.75–102.00), academic praise 13.70 times (range = 1.25–35.00), and academic corrective feedback 4.46 times (range = 0–13.25). Also, prior to receiving instructional coaching, teachers were independently observed on average using clear directives 20.02 times (range = 6.75–47.50), vague directives 2.89 times (range = .50–7.50), Behavior Praise 4.13 times (range = 0–16.50), and behavior corrective feedback 11.41 times (range = 1.50–30.50). Overall effect sizes were small for each IS count (.24–.33). Effect sizes for the behavior management strategy counts ranged from no effect (clear directives effect

Table 3. Descriptive Statistics of CSC 4 Session Model Using CSAS Strategy Counts and IS and *BMS Rating Scale* Discrepancy Scores.

	Baseline assessment	Postintervention assessment	Effect size
ISs counts			
Concept summaries	1.80 (1.61)	2.39 (2.54)	.24
Academic response opportunities	43.96 (21.66)	38.48 (16.21)	.32
Academic praise	13.70 (8.42)	16.14 (9.18)	.33
Academic corrective feedback	4.46 (3.32)	3.64 (2.77)	.33
IS standardized change score	0.28 (1.08)		.26
Behavior management counts			
Clear directives	20.02 (10.82)	20.06 (12.33)	.01
Vague directives	2.89 (2.19)	2.13 (3.06)	.36
Behavior praise	4.13 (4.34)	10.39 (7.81)	.83
Behavior corrective feedback	11.41 (6.95)	8.56 (7.17)	.50
Behavior management standardized change score	0.66 (1.56)		.42
IS Rating Scale total discrepancy scores	20.15 (11.81)	12.81 (9.82)	.88
BMS Rating Scale total discrepancy scores	26.19 (14.36)	17.48 (12.98)	.68

Note. CSC = Classroom Strategies Coaching; CSAS = classroom strategies assessment system; IS = instructional strategies; BMS = behavioral management strategies.

size of .01) to a large effect (i.e., behavior praise effect sizes of .83; Cohen, 1988).

Research Question 1: Primary Outcomes Postintervention

The primary outcome measures used were the standardized Instructional and BMS Counts recorded by the independent observers at postintervention. For IS Counts, results from one-sample *t* tests revealed that the standardized change scores between baseline assessment and postintervention assessment (M = .28, SD = 1.08) were not statistically different from zero, t (29) = 1.42, p = .167, d = .26, suggesting no changes in teachers' usage of the IS counts at postintervention. On the contrary, for BMS Counts, the standardized change scores between baseline assessment and postintervention assessment (M = .66, SD = 1.56) were significantly greater than zero, t (29) = 2.32, p = .028, d = .42.

Research Questions 2 and 3: Secondary Outcomes Postintervention

Observers' IS and BMS Rating Scale discrepancy scores were also used to assess need for change in instruction and behavior management practices. For IS discrepancy scores, paired samples *t* tests revealed that teachers had significantly smaller discrepancy scores at postintervention assessment (M = 12.81, SD = 9.82) than at baseline assessment (M = 20.15, SD = 11.81), $t(31) = 4.97$, $p < .001$, d = .88. Similarly for BMS discrepancy scores, teachers had significantly smaller discrepancy scores at postintervention assessment (M = 17.48, SD = 12.98) than at baseline assessment (M = 26.19, SD = 14.36), $t(30) = 3.84$, $p < .001$, d = .68. As shown on Table 3, the effect sizes for IS and BMS discrepancy scores represent large (d of .80) and medium (d of .50) effect sizes, respectively.

Table 4 presents the descriptive statistics of the TCES. Overall, teachers rated their satisfaction with the CSC Model as very high (M = 89.92, SD = 8.06). Notably, the means for each item did not go below 5.86 (Item 10, SD = 1.43) and ranged as high as 6.96 (Item 1, SD = .19).

Discussion

This article illustrated the benefits of integrating teacher formative assessment in instructional coaching to support improvement of elementary school teachers' universal classroom practices. The theory, emerging evidence, and key components of a brief instructional coaching model that includes formative assessment were described. Formative assessment data as measured by the CSAS-O indicated practice needs and goals, and provided useful performance feedback throughout the process. Using a sample of 32 teachers, we found promising changes in teachers' behavior management targets after receiving the CSC intervention and high levels of satisfaction with the CSC intervention. Findings suggest that the CSAS is sensitive to change in teaching practices over a short duration of instructional coaching and that the CSAS appears to be a useful observational assessment for facilitating teachers' use of evidence-based practices.

Table 4. Descriptive Statistics of the *Teacher Coaching Evaluation Scale* (TCES).

Item description	*M* (*SD*)	Range
Total score	89.93 (8.07)	71–98
The professional I have worked with showed a high level of interest and concern for me and my classroom.	6.96 (0.189)	6–7
The professional I have worked with showed a high level of skill.	6.82 (0.612)	4–7
The intervention strategies that have been used with my classroom fit my needs.	6.71 (0.460)	6–7
The intervention strategies that have been used with the classroom fit with my teaching style.	6.68 (0.612)	5–7
I feel better prepared to work with children with challenging behaviors in my classroom.	6.07 (1.05)	4–7
I feel better prepared to teach children in my classroom.	6.21 (0.957)	4–7
I have applied the skills I have learned to working with other children in my classroom.	6.57 (0.634)	5–7
The professional has helped me find ways to apply the content of our discussions to specific classroom situations.	6.68 (0.612)	5–7
I would work with staff from this project again.	6.61 (0.786)	4–7
The demands placed on me by the assessment components (e.g., completing forms, interviews) of this project were reasonable.	5.86 (1.433)	2–7
The demands placed on me in all other areas of the project except assessment have been reasonable.	6.11 (1.10)	4–7
The issues that originally prompted my participation in the project have been much improved.	5.89 (1.22)	3–7
I am satisfied with my progress.	6.36 (0.951)	4–7
Overall, I am very satisfied with the quality of the coaching I received.	6.61 (0.629)	5–7

We found that the delayed CSC intervention significantly improved teachers' standardized BMS Counts, resulting in a practical medium effect. Likewise, individual behavior management strategy counts yielded a range of effect sizes. For example, large, medium and small positive effects were found for increased use of behavior praise, as well as reduction in behavior corrective feedback and vague directives. Similarly, statistically significant and practical reductions in teachers' BMS rating scale discrepancy scores (i.e., need for change in classroom practices) were detected at postintervention by independent observers. These findings parallel outcomes from studies in the behavioral consultation literature that focus on enhancing teachers' classroom behavior management strategies, specifically studies focused on improving the use of specific-behavioral praise. Several studies have found behavioral consultation as an effective method for increasing teachers' use of specific-behavioral praise (e.g., Briere et al., 2015; Dufrene et al., 2012; Reinke et al., 2008). Similar to the current study, Reinke et al. (2008) focused on multiple BMS targets by also aiming to have teachers' reduce their use of behavioral reprimands (i.e., behavioral corrective feedback in the CSC). As the CSC Model is based on a behavioral consultation framework and behavior management literature, the current study's findings offer some support for the use of an instructional coaching approach for improving teachers' behavior management practices.

No significant improvements were found for the standardized IS Counts, although small positive effects were observed for individual ISs (e.g., academic response opportunities, academic praise, and academic corrective feedback). In contrast to the current study, a previous study by Stitcher, Lewis, Richter, Johnson, and Bradley (2006) found peer coaching to increase teachers' use of key opportunity to respond (OTR) variables. Specifically, the study found peer coaching effective for increasing teacher prompts and contingent feedback, which relate directly to the CSC and CSAS academic response opportunities and academic praise. Similar to the current study, Stitcher et al. (2006) sought to change multiple OTR variables at once. Although the current study sought to increase the rates of academic response opportunities and academic praise, the CSC Model did not specifically operate on these two related OTR variables in isolation from behavior management strategies, nor did the current study seek to increase these two strategies in accordance with the OTR sequence specified in the Stitcher et al. study. Interestingly, postintervention assessment revealed statistically significant and practical reductions in teachers' IS rating scale discrepancy scores, suggesting a reduced need for change in instructional practices. One potential explanation for the juxtaposition of findings between the CSAS Strategy Counts and *Instructional Strategies Rating Scales* can be attributed to type of information collected by each metric. The IS Counts (and by

extension the BMS Counts) collect raw frequency information on teachers' use of strategies and do not account for quality of implementation, which is present in the IS and BMS Rating Scale items. For example, teachers' frequency of academic praise is tallied in the Strategy Counts, whereas the specificity and immediacy of the praise are assessed in Strategy Rating Scales. Although a statistical improvement was not detected in the raw frequencies of the IS Counts, it appears that improvements in quality related dimensions were detected as measured by the reduced need for change in the IS Strategy Rating scale discrepancy scores. Alternatively, ISs may be more contextually dependent on the lesson being taught in comparison with classroom BMSs, which may be independent or universal to all lesson contexts (Fabiano et al., in press; Reddy et al., 2017).

Results from the current study reflect similar outcomes to those found in the first RCT of the CSC intervention, which examined the efficacy of the same four session model (Fabiano et al., in press). In the larger RCT, the intervention group teachers were found to have significant improvements in the standardized BMS Counts and BMS Rating Scale discrepancy scores compared with waitlist controls at postintervention (i.e., the second baseline for control group teachers in the current study). In addition, teachers in the RCT also possessed high ratings of satisfaction with the CSC intervention. The current study administered the same dosage of the CSC intervention to waitlist control teachers following a longer baseline period. Although the current study reflects Schmidt's (2009) definition of conceptual replication versus direction replication given the change in study design, the similarity in findings between the two studies provides some additional support for the efficacy of the CSC in promoting changes in teachers' classroom practices.

Practical Implications

The current study's outcomes have implications for teachers' classroom practices and PD initiatives. Descriptive statistics in the current study suggest usage of evidence-based practices by teachers, however, some practices were not used in accordance with recommendations from research. For example, behavior praise at the baseline assessment for the teachers evidenced a 1:4 ratio of behavior praise to behavior corrective feedback. This ratio contrasts the longstanding recommended ratio of 3:1 behavior praise to behavioral corrective feedback in general education and ratio of 5:1 in special education (e.g., Partin, Robertson, Maggin, Oliver, & Wehby, 2010). Previous investigations examining teachers' use of behavior praise have documented low frequency of implementation in general, and decreased usage as grade levels increase, despite the effectiveness of the strategy for reinforcing students' behavior (Brophy & Good, 1986; Reddy et al., 2013a). Using behavior praise as an example, the current study's findings in conjunction with past research reflects the literature documenting research-to-practice gaps for evidence-based practices in instruction and classroom behavior management (e.g., Gersten, Chard, & Baker, 2003; Horner & Sugai, 2009; McLeskey & Billingsley, 2008).

The current study's results add to the literature base supporting instructional coaching as an effective form of PD that can lead to changes in teachers' classroom practices. In this study, the CSC intervention helped teachers reverse the behavior praise to behavior corrective feedback ratio and effectively use both strategies. Similarly, previous studies using instructional coaching and behavioral consultation methods have demonstrated indirect service delivery approaches that are tailored and specific to teachers' classroom context are effective for bridging the research to practice to gap (Briere et al., 2015; Dufrene et al., 2012; Reinke et al., 2008). Future policy initiatives seeking to increase the delivery of evidenced-based practices should consider using instructional coaching interventions as a vehicle for delivery. Furthermore, methods for sustaining teachers' use of evidence-based strategies during the school year are needed. The results of the present CSC intervention have demonstrated an effective method for helping teachers' enhance their use of practices—specifically classroom BMSs—that can be employed as part of routine and ongoing PD activities throughout the school year.

In addition, in regard to teacher PD, the improvements in the behavioral management practices as measured by the CSAS may offer insight into the professional needs that educators consider most desirable for targeted interventions. Requests for additional training and supports in classroom management across grade levels by teachers have been documented in PD research (e.g., Coalition for Psychology in Schools and Education, 2006; Public Agenda, 2003; Reinke, Stormont, Herman, Puri, & Goel, 2011). Relatedly, new classroom teachers often begin teaching with minimal training in classroom management (Begeny & Martens, 2006) and minimal research exists documenting effective methods for promoting new classroom teachers use of classroom behavior management practices (Briere et al., 2015). The present findings of improved behavior management practices may be indicative of a general trend related to a lack of training and knowledge, for both new and experienced teachers, related to effective classroom behavior management strategies. Although not explored in the current study, instructional coaching models such as the CSC offer a solution for providing both new and experienced teachers with training and supports in implementing evidence-based behavior management practices.

The present findings also suggest observational measures such as the CSAS may be a valuable tool to provide individualized teacher feedback and follow-up support, tailored to a teacher's repertoire of current practice (e.g., Reddy &

Dudek, 2014; Reddy, Fabiano, Dudek, & Barbarasch, 2012). One advantage of a measure such as the CSAS is that it can be administered formatively to track teachers' use of specific strategies over time and across content areas. At this time, there is a need for additional research on the degree to which the CSAS is sensitive to change over time in teacher practices, as well as the reliability of estimated rates of change based on CSAS observational data.

Limitations

Results from this investigation should be interpreted cautiously with regard to setting, participants, and data collection limitations. The design and sample size of this descriptive study preclude firm conclusions about causality between the instructional coaching intervention and change in teacher practice, or about the true magnitude of the observed effect. Teachers in this study were predominantly Caucasian females from New Jersey and New York elementary schools. Although in alignment with state-level statistics, these results may not be generalizable to other settings. Unlike the previous RCT (Fabiano et al., in press) examining the CSC's efficacy in changing teachers' classroom practices, the current study utilized a pre–post design and did not contain a comparison group. We, therefore, cannot rule out that the waitlist control teachers may have been affected by their prior exposure to the teaching strategies on the CSAS, which are key components of the CSC coaching model.

Similarly, because the RCT randomized within school, we cannot rule out the possibility that teachers in the waitlist control group were influenced by their peers in the experimental immediate instructional coaching condition. Although we implemented procedures to reduce contamination between groups and observer impact on teacher behavior (e.g., teachers and observers completed a ground rules sheet; observers did not interact with teachers or students while observing), we cannot rule out that their presence may have had an influence on teachers in this study. However, these risks are likely attenuated by the focus on standardized strategy targets and the written ground rules agreement procedures.

The long-term effect of the CSC Model on teacher practices was not assessed in this study and warrants future investigation. Treatment fidelity and sustainability have been noted as common limitations affecting instructional coaching interventions and PD opportunities for teachers (Martens & Ardoin, 2002). Finally, the current study only focused on teacher-level data as outlined by the grant design. No data at the student level were collected, and therefore, a comparison translating changes in teacher behavior to student outcomes was not available and warrants future research.

Conclusion

Overall, this investigation describes findings on a short-term coaching and formative assessment model for improving teachers' universal classroom practices. This study offer promising and complimentary findings that support the efficacy of the CSC in driving teachers' best practices. Specially, in this investigation, independent observers rated improvements in teacher behavioral management practices following coaching intervention compared with baseline. Coaches used classroom observational data from the CSAS to target practice needs, set goals, monitor implementation, and provide ongoing performance feedback to teachers. This study suggests that the integration of instructional coaching with formative data serve as a promising approach for helping teachers enhance their use of universal practices.

Authors' Note

The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

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 disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080337 to Rutgers, The State University of New Jersey.

ORCID iD

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

References

Acker, M. M., & O'Leary, S. G. (1987). Effects of reprimands and praise on appropriate behavior in the classroom. *Journal of Abnormal Child Psychology*, *15*, 549–557.

Alberto, P., & Troutman, A. (2008). *Applied behavior analysis for teachers* (8th ed.). Columbus, OH: Prentice Hall.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Banilower, E. R. (2002). *Results of the 2001–2002 study of the impact of the local systemic change initiative on student achievement in science*. Chapel Hill, NC: Horizon Research.

Becker, K. D., Darney, D., Domitrovich, C., Keperling, J. P., & Ialongo, N. S. (2013). Supporting universal prevention programs: A two-phased model. *Clinical Child and Family Psychology Review*, *16*, 213–228.

Begeny, J. C., & Martens, B. K. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly*, *21*, 262–285.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy*. New York, NY: Plenum.

Briere, D. E., Simonsen, B., Sugai, G., & Myers, D. (2015). Increasing new teachers' specific praise using a within-school consultation intervention. *Journal of Positive Behavior Interventions*, *17*, 50–60.

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.

Bush, R. N. (1984). *Effective staff development in making schools more effective: Proceedings of three state conferences*. San Francisco, CA: Far West Laboratory.

Carlisle, J. F., & Berebitsky, D. (2011). Literacy coaching as a component of professional development. *Reading & Writing Quarterly*, *24*, 773–800.

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. *Psychological Assessment*, *6*, 284–290. doi:10.1037/1040-3590.6.4.284

Coalition for Psychology in Schools and Education. (2006, August). *Report on the teacher needs survey*. Washington, DC: American Psychological Association, Center for Psychology in Schools and Education.

Coffee, G. C., & Kratochwill, T. R. (2013). Examining teacher use of praise taught during behavioral consultation: Implementation and generalization considerations. *Journal of Educational and Psychological Consultation*, *23*, 1–35.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Darling-Hammond, L. (1993). Reframing the school reform agenda. *Phi Delta Kappan*, *74*, 752–761.

Darling-Hammond, L., Chung Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council. Retrieved from https://learningforward.org/docs/default-source/pdf/nsdcstudytechnicalreport2009.pdf

Denton, C. A., & Hasbrouck, J. (2009). A description of instructional coaching and its relation to consultation. *Journal of Educational and Psychological Consultation*, *19*, 150–175.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181–199.

Desimone, L. M., Smith, T., & Frisvold, D. (2007). Is NCLB increasing teacher quality for students in poverty? In A. Gamoran (Ed.), *Standards-based and the poverty gap: Lessons from No Child Left Behind* (pp. 89–119). Washington, DC: Brookings Institution Press.

Dudek, C. M., Kettler, R. J., Kurz, A., & Reddy, L. A. (under review). *A multi measure approach to classroom observational assessment: A preliminary investigation of concurrent validity*. Studies in Educational Evaluation.

Dudek, C. M., Reddy, L. A., & Lekwa, A. J. (in press). Measuring teacher practices to inform student achievement in high poverty schools: A predictive validity study. *Contemporary School Psychology*.

Dufrene, B. A., Lestremau, L., & Zoder-Martell, K. (2014). Direct behavioral consultation: Effects on teachers' praise and student disruptive behavior. *Psychology in the Schools*, *51*, 567–580.

Dufrene, B. A., Parker, K. M., Menousek, K., Zhou, Q., Harpole, L. L., & Olmi, D. J. (2012). Direct behavioral consultation in Head Start to improve teacher use of praise and effective instruction delivery. *Journal of Educational and Psychological Consultation*, *22*, 159–186.

Ermeling, B. (2009). Tracing the effects of teacher inquiry on classroom practice. *Teaching and Teacher Education*, *26*, 377–388.

Fabiano, G., Reddy, L. A., & Dudek, C. M. (in press). Use of teacher coaching and formative assessment for improving classroom practices. *School Psychology Quarterly*.

French, V. W. (1997). Teachers must be learners, too: Professional development and national teaching standards. *NASSP Bulletin*, *81*, 38–44.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco, CA: Jossey-Bass.

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.

Garet, M. S., Porter, A. C., Desimone, L. M., Briman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Gersten, R., Chard, D., & Baker, S. (2003). Factors enhancing sustained used of research-based instructional practices. *Journal of Learning Disabilities*, *33*, 445–457.

Gulamhussein, A. (2013). *Teaching the teachers: Effective professional development in an era of high stakes accountability*. Center for Public Education. Retrieved from http://www.centerforpubliceducation.org/system/files/2013-176_ProfessionalDevelopment.pdf

Guskey, T. R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal*, *21*, 24–259.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*, 381–391. doi:10.1080/135406002100000512

Harris, A. (1998). Effective teaching: A review of the literature. *School Leadership & Management*, *18*, 169–183.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. New York, NY: Routledge.

Hoagwood, K., Burns, B. J., & Weisz, J. (2002). A profitable conjunction: From science to service in children's mental health. In B. J. Burns & K. Hoagwood (Eds.), *Community-based interventions for youth with severe emotional disturbances* (pp. 327–338). New York, NY: Oxford University Press.

Horner, R. H., & Sugai, G. (2009). *Is school-wide positive behavior support an evidence-based practice?* Available from www.pbis.org

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*(8), 1–14

Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational Leadership*, *40*, 4–10.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development*. Alexandria, VA: Association for Supervision and Curriculum Development.

Kerns, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools*, *44*, 65–75.

Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oaks, CA: Corwin Press.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rhinehard, & Winston.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, *33*, 279–299.

Lekwa, A. J., Reddy, L. A., Dudek, C. M, & Hua, A. (under review). *Predictive Validity of the Classroom Strategies Assessment System – Observer Form for Growth in Reading and Mathematics. School Psychology Quarterly*.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Simms, J. A. (2013). *Coaching classroom instruction*. Bloomington, IN: Marzano Research Laboratory.

Martens, B. K., & Ardoin, S. P. (2002). Training school psychologists in behavior support consultation. *Child and Family Behavior Therapy*, *24*, 147–163. doi:10.1300/J019v24n01_10

McLeskey, J., & Billingsley, B. S. (2008). How does the quality and stability of the teaching force influence the research-to-practice gap: A perspective on the teacher shortage in special education. *Remedial and Special Education*, *29*, 293–305.

Northwest Evaluation Association. (2011, January). *Technical manual for measure of academic progress & measure of academic progress for primary grades*. Portland, OR: Northwest Evaluation Association.

O'Leary, K. D., Kaufman, K. F., Kass, R. E., & Drabman, R. S. (1970). The effects of loud and soft reprimands on behavior of disruptive students. *Exceptional Children*, *37*, 145–155.

Partin, T., Robertson, R., Maggin, D., Oliver, R., & Wehby, J. (2010). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure*, *54*, 172–178.

Penuel, W., Fishman, B., Yamaguchi, R., & Gallagher, I. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Education Research Journal*, *44*, 921–958.

Public Agenda. (2003). *Where we are now: 12 things you need to know about public opinion and public schools*. New York, NY: Author.

Reddy, L. A., & Dudek, C. (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., & Lekwa, A. J. (2017). Classroom strategies coaching model: Integration of formative assessment and instructional coaching. *Theory Into Practice*, *56*, 46–55. doi:10.1080/00405841.2016.1241944

Reddy, L. A., Fabiano, G. A., & Dudek, C. M. (2012). Concurrent validity of the Classroom Strategies Scale for elementary school—Observer form. *Journal of Psychoeducational Assessment*, *31*, 258–270.

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Barbarasch, G. (2012). Behavior management of students with ADHD using teacher and student progress monitoring. In L. M. Crothers & J. B. Kolbert (Eds), *Understanding and managing behaviors of children with psychological disorders: A reference for classroom teachers*. (pp. 17-48). New York, NY: Continuum International.

Reddy, L. A., Fabiano, G. A., 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. A., Dudek, C. M., & Hsu, L. (2013b). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology*, *51*, 683–700.

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013c). Predictive validity of the Classroom Strategies Scale-observer form on statewide testing. *School Psychology Quarterly*, *28*, 301–316.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavioral Interventions*, *15*, 39–50.

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*, 315–332.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., Goel, N. (2011). Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, *26*, 1–13.

Riley-Tillman, T. C., & Ecker, T. L. (2001). Generalization programming and school-based consultation: An examination of consultees' generalization of consultation-related skills. *Journal of Educational and Psychological Consultation*, *12*, 217–241.

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, *13*, 74–101.

Scheeler, M. C., Rukl, K., & McAffee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, *27*, 396–407. doi:10.1177/088840640402700407

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. doi:10.1037/a0015108

Sheridan, S. M., & Kratochwill, T. R. (2008). *Conjoint behavioral consultation*. New York, NY: Springer.

Sheridan, S. M., Welch, M., & Orme, S. (1996). Is consultation effective? A review of outcome research. *Remedial and Special Education*, *17*, 341–354.

Shernoff, E. S., Lekwa, A. J., Reddy, L. A., & Coccaro, C. (2017). Examining teachers' attitudes and experiences with coaching to inform research-based practice: An iterative developmental design study. *Journal of Educational and Psychological Consultation*, *27*, 459–485.

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review*, *41*, 160–175. doi:10.1002/pits.20625

Stitcher, J. P., Lewis, T. J., Richter, M., Johnson, N. W., & Bradley, L. (2006). Assessing antecedent variables: The effects of instructional variables on student outcomes through in-service and peer coaching professional development models. *Education and Treatment of Children*, *29*, 665–692.

Sugai, G., Horner, R., & Gresham, F. (2002). Behaviorally effective school environments. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Prevention and remedial approaches* (pp. 315–350). Bethesda, MD: NASP.

Tomlinson, C., & McTighe, J. (2006). *Integrating differentiated instruction and understanding by design: Connecting content and kids*. Alexandria, VA: Association for Supervision and Curriculum Development.

Truesdale, T. W. (2003). *The implementation of peer coaching on the transferability of staff development to classroom practice in two selected Chicago public elementary schools*. Chicago, IL: Loyola University of Chicago.

Yoon, K. S., Dunca, T., Wen-Yu Lee, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* [Issues and Answers Report]. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retried from http://ies.ed.gov/ncee/edlabs

Ysseldyke, J. E., & Christenson, S. L. (2002). *Functional assessment of Academic Environment Scale*. Longmont, CO: Sopris West.