International Journal of School & Educational Psychology, 2, 71–84, 2014

ISSN 2168-3603 print/ISSN 2168-3611 online
DOI: 10.1080/21683603.2013.876951

Routledge
Taylor & Francis Group

ARTICLES

Teacher Progress Monitoring of Instructional and Behavioral Management Practices: An Evidence-Based Approach to Improving Classroom Practices

Linda A. Reddy and Christopher M. Dudek
Graduate School of Applied and Professional Psychology, Rutgers University, Piscataway, New Jersey, USA

In the era of teacher evaluation and effectiveness, assessment tools that identify and monitor educators' instruction and behavioral management practices are in high demand. The Classroom Strategies Scale (CSS) Observer Form is a multidimensional teacher progress monitoring tool designed to assess teachers' usage of instructional and behavioral management strategies in elementary school. The present article briefly describes the CSS methodology and psychometric properties. The CSS consists of a three-part assessment: (a) direct classroom observation, (b) Strategy Rating Scales of instruction and behavioral management, and (c) a classroom checklist. A teacher case example is presented to illustrate the CSS's clinical utility in schools. Implications for school psychological practice are outlined.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/usep.

Keywords: Teacher behavior, assessment, professional development

The evaluation of teacher performance and classroom practice is a common praxis worldwide. International recognition of teachers' influences on student achievement and the desire to increase instructional quality has led many countries to establish teacher performance assessments and evaluation procedures (Isore, 2009; Organization for Economic Cooperation and Development [OECD], 2009). For example, Chile follows a four-domain evaluation model occurring every four years, while England follows a three-domain model occurring once per year (Avalos & Assasel, 2006; Training and Development Agency for Schools, 2007). Although teacher evaluation systems vary from country to country in terms of method, criteria, and data collection instruments, they share two common purposes: (a) the monitoring of teacher performance to promote maximal student learning and (b) the improvement of teacher practice via identifying strengths and growth areas (Isore, 2009).

In the United States, improving teacher performance through rigorous teacher evaluation has received recent national attention. Classroom observations are a common method worldwide for teacher evaluation and one of the central assessments for identifying and monitoring effective teacher practices in the United States (Cantrell & Kane, 2013). The recent Measures of Effective Teaching (MET) study found that four brief direct observations conducted by more than one observer yields the highest reliability of teacher practices (Cantrell & Kane, 2013).

Although the MET study results offer some promising directions, historically the teacher evaluation process in the United States has yielded little or no effect on teaching practice despite its purported role and responsibility for

Submitted April 6, 2013; accepted November 13, 2013.
Correspondence should be addressed to Linda A. Reddy, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854-8085, USA. E-mail: LReddy@RCI.Rutgers.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/usep.

directing teachers' professional development (e.g., Kauchak, Peterson, & Driscoll, 1985; Porter, Youngs, & Odden, 2001). Previous studies have documented teacher performance evaluations (i.e., observations) as typically infrequent, occurring as little as once per year in some states (Scheeler, Bruno, Grubb, & Seavey, 2009). Several studies also characterize principals, the key implementers of teacher evaluation, as inaccurate raters of teacher behavior, thus raising questions of accuracy and integrity of evaluation results (e.g., Dwyer & Stufflebeam, 1996; Peterson, 1995; Porter et al., 2001). A recent publication entitled the "Widget Effect" brought widespread attention to the failure of teacher evaluation systems across the nation (Weisberg, Sexton, Mulhern, & Keeling, 2009). This landmark report highlighted that teacher evaluation systems do not practically differentiate levels of individual teacher performance and, unfortunately, are not linked to targeted professional development (Weisberg et al., 2009).

Improving teacher performance through professional development has also become a national focus in the United States, yet these programs have yielded mixed and at times questionable outcomes. Few large-scale studies have directly measured the effects of professional development on teacher learning and professional growth (e.g., Carlisle, Correnti, Phelps, & Zeng, 2009; Goldschmidt & Phelps, 2010). Studies examining the effects of professional development describe these programs as short in duration, lacking in follow-up support, and ineffective in promoting teacher practice change (e.g., Desimone, Porter, Garet, Yoon, & Birman, 2002; Goldenberg, & Gallimore, 1991; Sparks, 1983; Ward, 1985; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Furthermore, sustainability research has demonstrated that teachers do not generalize and transfer the information taught or learned in professional development courses into their classrooms (Riley-Tillman & Eckert, 2001; Rose & Church, 1998). Taken together, there is a critical need for research and assessments that are linked to professional development efforts.

One direction for consideration is the application of progress monitoring for teachers' classroom practices. Progress monitoring is the scientific practice of assessing students' academic performance on a regular basis for the purposes of determining instructional outcomes, building instructional programs for at-risk students, and monitoring student improvement (National Center on Student Progress Monitoring, 2006). Progress monitoring has been used almost exclusively for tracking students' academic and behavioral performance. To date, few teacher assessments exist that identify and monitor educators' professional practices (e.g., Reddy, Fabiano, Dudek, & Hsu, 2013b; Reddy, Fabiano, & Jimerson, 2013).

To this end, the present article describes a new classroom observational measure, the Classroom Strategies Scale (CSS)-Observer Form, a multidimensional teacher progress monitoring assessment for monitoring educators' classroom practices. A teacher case example is presented to illustrate the clinical application of the CSS. Implications of teacher progress monitoring for school practice are offered.

CLASSROOM STRATEGIES SCALE

The CSS-Observer Form is grounded in models of effective teaching from over 50 years of research (e.g., Brophy & Good, 1986; Gage, 1978; Marzano, 1998; Marzano, Pickering, & Pollock, 2001; Wittrock, 1986; Walberg, 1986). This body of work has highlighted general features of effective instructional practice linked to positive student learning (e.g., Bennet, 1988; Creemers, 1994; Good & Brophy, 1980; Harris, 1998; Hattie, Biggs, & Purdie, 1996; Scheerens, 1992; Walberg, 1986; Wang, 1991; Wang, Haertel, & Wahlberg, 1993). Under the umbrella of effective teaching, the CSS has been conceptualized to include dimensions of instructional and classroom management practices (e.g., Alberto & Troutman, 2003; Horner, Sugai, Todd, & Lewis-Palmer, 2000, 2005; Kounin, 1970; Schloss & Smith, 1998; Stage & Quiroz, 1997; Walker, Ramsey, & Gresham, 2003).

Based on research, the CSS was developed as a user-friendly multidimensional assessment of instructional and behavioral management strategies. The CSS generates scores that: (a) assess educators' use of empirically supported instructional and classroom behavioral management strategies, (b) identify practice goals for improvement, (c) monitor educators' progress towards practice goals following intervention, (d) provide evidence for professional development and supports (e.g., professional learning committees), and (e) help refine school-wide teacher professional development plans.

Development of the CSS

Guided by contemporary test theory (e.g., Anastasi & Urbina, 1997; Benson, 1998; Crocker & Algina, 1986; Kane, 2002, 2008), the CSS was designed specifically for school personnel for routine educational practice. The central goal was on maximizing the intended score utility for school personnel to inform educator practice change (Kane, 2002).

The CSS was iteratively developed through several methods: (a) expert input, (b) consumer input, (c) extensive field testing (i.e., pilot 1 $n = 100$; pilot 2 $n = 317$; pilot 3 $n = 100$) and (d) a set of data analytic methods. The CSS domains and items were guided by expert input through a comprehensive review of peer-reviewed publications, other related tests, as well as input from a national advisory board that included experts in instruction, behavior management, and measurement. The consumer advisory board provided critical feedback to the specific domains and items, as well as item ambiguity and possible bias. Face/content validity of

the CSS was established in part through the expert and consumer advisory boards independently rating on a 4-point Likert-type scale (i.e., 1 *not at all matches* to 4 *very much matches*) the degree to which each item matched the proposed domain. The boards were also asked to provide feedback on new domains and items and the CSS intended use and score utility for assessing practices and informing changes practices (i.e., professional development). Additionally, several statistical methods were employed to refine and revise the CSS domains and items such as item-to-total correlations, pooled mean item variances across observation (level of disagreement), as well as confirmatory factor analysis within observation using recommended fit indices (Jackson, Gillapsy, & Purc-Stephenson, 2009) and information-theory-based indices of relative fit (Bowen & Guo, 2012; see Reddy et al., 2013b for details).

Dimensional Structure and Scoring

The CSS consists of three parts that include empirically supported instructional and behavioral management strategies (see Table 1). For Part 1 Classroom Observation, observers tally each time eight instructional and behavior management strategies are used during an observation (lesson) period and whether the strategy used was for individual students or groups of students (i.e., two or more students; see Table 1). Following the direct observation, observers complete the Part 2 Strategy Rating Scales, which consist of Instructional Strategies (IS) and Behavioral Management Strategies (BMS) Scales. The IS scale includes 28 items that comprise a total scale, two composite scales, and five subscales. The Instructional Methods Composite scale (17 items; maximum frequency score of 119) consists of the Direct Instruction (8 items; maximum score of 56), Adaptive Instruction (4 items; maximum score of 20) and Student Focused Instruction (5 items; maximum score of 42) subscales. The Academic Monitoring/Feedback Composite scale (11 items; maximum score of 77) consists of the Promotes Student Thinking (5 items; maximum score of 35) and Academic Performance Feedback (6 items; maximum score of 42) subscales (see Table 1).

The BMS scale includes 26 items that compose a total scale, two composite scales, and four subscales. The Behavioral Feedback Composite scale (12 items; maximum frequency score of 84) consists of Praise (5 items; maximum score of 35) and Corrective Feedback (7 items; maximum score of 49) subscales. The Proactive Methods Composite scale (14 items; maximum score of 91) consists of Prevention Management (8 items; maximum score of 56) and Directives (6 items; maximum score of 42) subscales (see Table 1).

After each classroom observation period, observers rate how often (*Frequency Rating*) teachers used specific instructional and behavioral management strategies on a 7-point Likert 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 scale (1 "never used," 3 "sometimes used," 7 "always used"). The Part 2 Rating Scales produce both frequency scores and discrepancy scores. For the Part 2 Strategy Rating Scales, item discrepancy scores are computed as follows: | recommended frequency − frequency ratings |.

Absolute value discrepancy scores indicate if any change (regardless of direction) was needed as measured by the observer using the CSS. Larger discrepancy score values indicate greater amounts of change are needed in the practices measured by the CSS. In the current study, both frequency and discrepancy scores were separately analyzed. Absolute value discrepancy scores are calculated at the item level for the IS and BMS scales, for classroom observations 1 and 2 separately. IS and BMS scale scores are then calculated for observations 1 and 2 separately by summing these discrepancy scores of the associated items. The scale scores are added from observation 1 to the corresponding discrepancy scale scores in observation 2, and then divided by 2 to obtain the average absolute value discrepancy score across both observations.

After completing Parts 1 and 2, the observer then completes the Classroom Checklist (Part 3). The Classroom Checklist assesses the presence of 14 specific items or procedures in the classroom (see Table 2).

The CSS-Observer Form can be used for one or multiple observations. In the current case example, two observations were conducted for each administration of the CSS-Observer Form. CSS scores were calculated in accordance with multiple observation procedures. For Part 1, the eight teacher strategies were averaged across observations 1 and 2 during the baseline phase and across observations 9 and 10 during the posttest phase. For Part 2, both the frequency and absolute value discrepancy scores were first calculated at the item level for the IS and BMS scales, for classroom Observations 1, 2, 9, and 10 separately. IS and BMS scale scores were then calculated separately for each observation by summing the discrepancy scores of the associated items. The respective scale scores from Observation 1 were then added to the corresponding scale scores in Observation 2, and then divided by 2 to obtain the average absolute value discrepancy score across both observations. This process was repeated for Observations 9 and 10 to obtain the posttest phase CSS scores. Effect size calculations between baseline and posttest utilized the averaged totals for both the Part 1 Strategy Counts and Part 2 Strategy Rating Scales comparisons.

Observer Training and Reliability

Given that credentials for administrative positions in the United States vary from state to state, and in many cases multiple credentials can be used, the CSS-Observer Form training is designed to encompass various observer backgrounds ranging from no teaching experience to high levels of teaching experience. The CSS-Observer Form

TABLE 1
Definitions of the Three-Part CSS Assessment

Part 1 Strategy Counts	*Definitions*
Concept Summaries	A 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	A 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 responses (e.g., explain answers, repeat key points, brainstorm ideas, and show answers on the board).
Academic Praise Statements	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate academic performance.
Academic Corrective Feedback	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for incorrect academic performance.
Clear One to Two Step Directives	A 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	A 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 or more than two steps.
Behavioral Praise Statements	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for appropriate behavior.
Behavioral Corrective Feedback	A teacher gives a verbal or nonverbal statement or gesture to provide feedback for inappropriate behavior.
Total	The sum of the frequency of the eight teacher behaviors.
Part 2: Instructional Strategies Scales	*Definitions*
Total Scale	The Total Instructional Strategies scale reflects the overall use of Instructional Methods and Academic Monitoring/Feedback.
Instructional Methods composite scale	How classroom instruction occurs. Measures teachers' use of teacher-directed student-directed methods, or differentiated instruction. This includes how a teacher incorporates active learning techniques such as hands-on learning and collaborative learning in the presentation of lessons as well as how a teacher delivers academic content to students.
Adaptive Instruction subscale	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 subscale	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 subscale	Strategies teachers use to deliver academic content or convey information to students. These practices include direct instruction techniques, modeling, identifying, and summarizing.
Academic Monitoring/Feedback composite scale	How teachers monitor students' understanding of the material and provide feedback on their understanding. These strategies assess students' thinking and encourage students to examine their own thought processes. Teachers guide students' understanding by encouraging students, affirming appropriate application of the material, and correcting misperceptions.
Promotes Students' Thinking subscale	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, and why).
Academic Performance Feedback subscale	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.
Part 2: Behavioral Management Strategies Scales	*Definitions*
Total Scale	The Total Behavioral Management Strategies scale reflects the overall use of Proactive Methods and Behavior Feedback.
Preventative Methods composite scale	Strategies teachers use to promote positive behaviors in the classroom and reduce the likelihood of negative behaviors. These strategies include prompts, routines, reviewing rules, and presenting instructions or requests in a clear manner.
Proactive Methods subscale	Verbal and nonverbal strategies teachers use to prevent student disengagement and problem behaviors from occurring in the classroom. These practices assess how teachers create a positive classroom environment.

(*continued*)

TABLE 1 – (*Continued*)

Part 2: Behavioral Management Strategies Scales	*Definitions*
Directives subscale	Strategies teachers use for issuing directions or instructions to students and behavioral expectations in the classroom.
Behavioral Feedback composite scale	How teachers respond to students' appropriate and inappropriate behaviors. This includes the usage of praise to encourage positive behaviors and corrective feedback to redirect negative behaviors.
Praise subscale	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 subscale	Verbal and nonverbal strategies teachers use to correct students' inappropriate behavior. These practices assess how teachers respond to negative behavior in the classroom.

Part 3 Classroom Checklist Items	
1. Different methods/mediums of instruction are present in the classroom (e.g., blackboard, overhead projector, smart board, student clickers).	7. Tables/desks are arranged for students to easily view and participate in the lesson.
2. Learning aids are present in the classroom (e.g., number chart, vocabulary list, critical thinking questions).	8. Classroom lesson or activity schedules are clearly posted.
3. Learning materials are present in the classroom (e.g., pencils, rulers, construction paper).	9. Assignments (e.g., homework, readings, tests) are clearly posted.
4. Learning materials and areas in the classroom are labeled.	10. Student work, artwork, and accomplishments are displayed in the classroom.
5. A procedure or routine exists for students to organize their desks, backpacks, or learning materials.	11. Methods for tracking student academic and/or behavioral progress (e.g., homework-tracking chart, rule-following chart, sticker/star chart) are present.
6. Classroom (e.g., floors, walls, table) is clean and uncluttered.	12. Classroom-wide reward system is present (e.g., ticket bin for a pizza party).
	13. Classroom rules are posted.
	14. Classroom rules specify positive behaviors that students "should do" rather than "not do."

training consists of a four-step process gradually increasing exposure to content knowledge, and observation skills related to the CSS. First, observers watched a training video that introduced CSS observation procedures, provided an overview of how ratings are completed, and then showed several classroom examples of teachers displaying specific behaviors assessed by the CSS.

Second, the observers received two didactic training sessions (2 hours each) from a CSS Trainer/Master Coder that included discussion of definitions and criteria. Observers were oriented to the scientific literature guiding the development of the CSS and the recommended frequencies of strategies to ensure observers operated with the same knowledge base for judging the Recommended Frequency of the CSS Part 2. Training on the Recommended Frequency of strategies was informed by the effective instruction literature that spans over 60 years (e.g., Brophy & Good, 1986; Creemers, 1994; Gage, 1978; Hattie et al., 1996; Horner et al., 2000; Kounin, 1970; Marzano, 1998; Marzano et al., 2001; Walberg, 1986; Wang, 1991). For example, the academic and behavioral literatures have indicated that praise statements should be used frequently and consistently (e.g., Albert, Heward, & Hippler, 1999; Beaman & Wheldall, 2000; Sutherland & Wehby, 2001). In particular, praise should be used at a ratio of 3:1 to corrective feedback (i.e., reprimands).

Third, the observers practiced coding classroom videos using the CSS and practice results were reviewed by a CSS Trainer/Master Coder. Specific feedback and additional instruction was provided to observers by the CSS Trainer/Master Coder to further orient them to the CSS definitions and criteria. Finally, observers were required to pass a video coding criterion test on the CSS. Independent observers coded five classroom videos using the CSS. Observers were certified as reliable when their scores reached the minimum interrater reliability level of 80% with CSS Trainer/Master Coders.

Psychometrics Characteristics of the CSS

Psychometric properties of the CSS-Observer Form (version 2.0) were examined in a previous investigation of 317 general education teachers from 73 elementary schools located in New Jersey and New York (Reddy, Fabiano, Dudek, & Hsu, 2013a, 2013b for details). Grade level assignment was stratified across kindergarten to fifth grade and included 60 teachers in kindergarten, 48 in first grade, 64 in second grade, 60 in third grade, 41 in fourth grade, and 44 in fifth grade. A total of 67 observers, composed of principals ($n = 44$) and research staff ($n = 23$) administered the CSS. Principals conducted the CSS on 168 teachers in the sample and research staff performed the CSS on 149 teachers in the sample. Teachers received two 30-min observations with the CSS in which scores from both observations were aggregated together according to CSS procedures for multiple observations.

Factor structure

The Part 2 IS and BMS scales are theoretically and factor analytically derived (confirmatory factor analysis) within

TABLE 2
Descriptive Statistics and Effect Sizes for the CSS Part 1 Strategy Counts

	Baseline		*Posttest*		
Eight Strategies	*Mean*	*SD*	*Mean*	*SD*	*Effect size*
Concept Summaries	1.50	0.71	6.00	2.83	6.36
Academic Response Opportunities	21.50	2.12	38.00	4.24	7.78
Academic Praise	10.50	0.71	35.00	8.49	34.64
Academic Corrective Feedback	0.50	0.71	0.00	0.00	−0.71
Clear Directives	17.00	8.49	17.00	11.31	0.00
Vague Directives	0.00	0.00	0.00	0.00	0.00
Behavioral Praise	4.00	0.00	15.00	1.41	11.00
Behavioral Corrective Feedback	9.50	4.95	2.00	0.00	−1.52

Values before dashes in Table 2 are minus signs.

classroom observations. The CSS factor structure was examined with over 12 confirmatory factor analyses using generalized least squares estimation (SPSS's AMOS Version 19 software, Arbuckle, 2010). As described in Reddy et al. (2013b), several fit indices including χ^2/df, Root Mean Square Error of Approximation (RMSEA), adjusted goodness of fit index (AGFI), and goodness of fit index (GFI) recommended by Jackson et al. (2009) were used to test the fit to the data. CFA fit indices met acceptable benchmarks for all scales, providing evidence for the CSS Total scales, Composite scales, and subscales. In addition, CSS preferred factor models were compared to alternative models using information-theory-based indices of relative fit, including Akaike Information Criterion (AIC), Brown-Cudeck Criterion (BCC), and Schwarz Bayesian Information Criterion (BIC), described by Bowen and Guo (2012). Overall, results indicated that CSS four- and five-factor models yielded good fit to the data and superior fit to the data in comparison to alternative models using information-theory-based indices of relative fit (see Reddy et al., 2013b).

CSS reliability

The CSS was found to demonstrate good internal consistency (Cronbach alphas of 0.92–0.93) across Parts 1 through 3. For the Part 1 Total Strategy Counts, internal consistency was 0.92. For the Part 2 IS and BMS Total scales, internal consistency estimates were 0.91 and 0.92, respectively.

Interrater reliability data was collected on a random sample of 82 cases from the larger sample in the psychometric investigation. Interrater reliability was measured using Pearson's product moment correlation coefficients and percent agreement (Jackson, Gillapsy, & Purc-Stephenson, 2009). Overall, good interrater reliability data was found for all three parts of the CSS. For example, interrater reliability for the Part 1 Teacher Strategies was $r = 0.94$ (percent agreement 92%) and for the Part 2 IS and BMS Strategy Rating Scales was $r = 0.80$ and $r = 0.72$ (percent agreement 92% and 88%). Likewise, the interrater reliability for the Part 3 Classroom Checklist was $r = 0.86$ (percent agreement 91%). The interrater reliability estimates of the CSS align with accepted values for other classroom observation assessments such as the measures used in the Measures of Effective Teacher Project (Cantrell, 2013; Kane & Staiger, 2012) and the Classroom Assessment Scoring System (CLASS; Pianta, Le Paro, Hamre, 2008).

The CSS evidenced fair to good test-retest reliability (approximately 2 to 3 weeks) in a sample of 57 classrooms. For example, an r of 0.70 (percent agreement 81%) was found for the Part 1 Total Behaviors, rs of 0.86 and 0.80 (percent agreement 93% and 85%) for the Stage 2 IS and BMS Total scales, and r of 0.77 (percent agreement was 81%) for the Stage 3 Classroom Checklist. Differential item functioning analyses (partial correlations) have revealed that the Part 2 Strategy Rating Scale items evidence freedom-to-item bias for teacher age, educational degree, and years of teaching experience (Reddy et al., 2013b).

CSS validity

The CSS evidences concurrent and divergent validity, as well as predictive validity. In a study with 125 classrooms, the CSS was compared to the Classroom Assessment Scoring System (CLASS), a well-established measure of teacher and classroom quality (Pianta, La Paro, & Hamre, 2008). As hypothesized, the CSS corresponded with logically related CLASS domains (e.g., Behavior Management) and it did not correspond with domains hypothesized to be unrelated (e.g., Language Modeling). Thus, the CSS has been found to have good convergent and discriminant validity with the Classroom Assessment Scoring System (Reddy, Fabiano, & Dudek, 2013). Using a series of two-level hierarchical linear modeling, the CSS IS scale discrepancy scores uniformly predicted student mathematics and language arts statewide testing scores for 663 third, fourth, and fifth graders (Reddy, Fabiano, Dudek, & Hsu, 2013c).

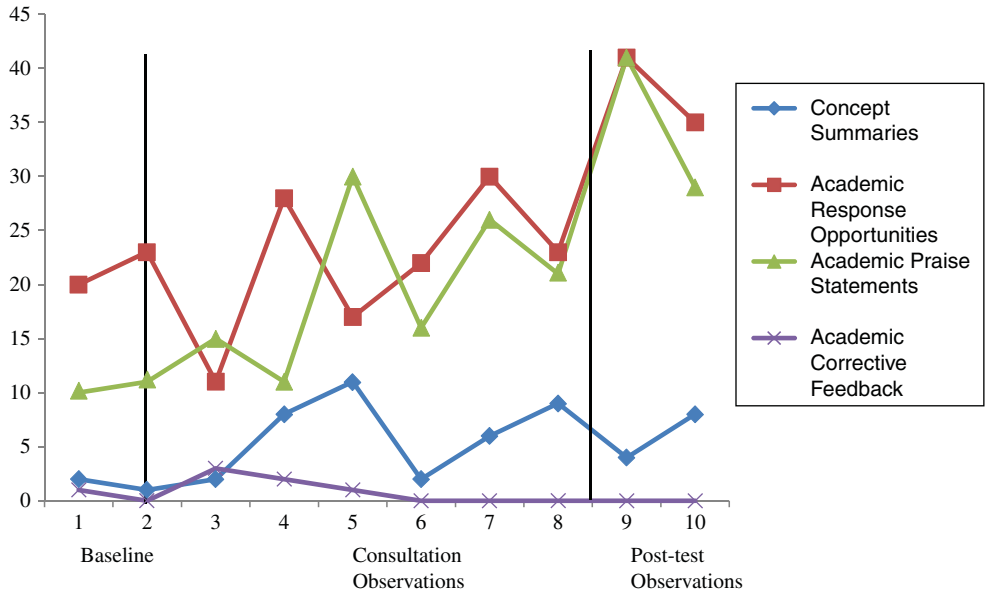

FIGURE 1 Visual analysis of the CSS Part 1 Teacher Instructional Strategies.

We offer the following teacher case example to illustrate the application of the CSS as a progress monitoring instrument for teachers' professional practice.

ILLUSTRATIVE CASE STUDY

A four-session modified collaborative consultation model (Bergan & Kratochwill, 1990; Reddy, Fabiano, Barbarasch, & Dudek, 2012) was used in the described case. During consultation, the consultant administered the CSS during six 30-min lessons to provide individualized visual performance feedback (VPF) to the teacher. Independently trained observers administered the CSS in two 30-min observations prior to consultation (i.e., baseline) and after the completion of the consultation (posttest).

Two data analytic methods were used to assess the CSS sensitivity to change following consultation. First, time-series graphs are presented to display changes in teacher practices using the CSS Part 1 eight teacher strategies (see Figures 1 and 2) and Part 2 IS and BMS Frequency and Discrepancy scales from baseline to posttest (see Figures 3 through 6). Second, single-case design effect sizes[1] were computed to provide an estimate of the practical changes in the teacher's classroom practices as measured by the CSS (Busk & Serlin, 1992).

The Case of "Jane"

Jane is a 41-year- old, Caucasian female teacher with a bachelor's degree in elementary education. She has 19 years of experience as a teacher in elementary school settings and has worked in her current position as a second-grade teacher for the past five years. Her classroom was composed of 25 general education students. Although no students in her class are classified with a specific learning disability, Jane reported academic and behavioral concerns for four students.

Consultation

Consultation was conducted by a supervised doctoral student in school psychology. The consultant and teacher met for four 30-min sessions once per week over the period of four weeks. The consultant administered the CSS and graphed CSS scores (i.e., visual performance feedback; VPF) between consultation sessions (i.e., sessions 1 and 2; sessions 2 and 3; sessions 3 and 4). The VPF provided the teacher feedback on her progress toward her practices goals (e.g., Figure 1) and were reviewed during consultation sessions 2, 3, and 4. After each consultation session, the consultant faxed and e-mailed the teacher a memo outlining what was discussed during the meeting.

During session 1, the consultant and Jane discussed her overall use of instructional and classroom behavioral management strategies. Jane and the consultant collaboratively reviewed the CSS Part 1 eight teacher strategies and

1. Busk and Serlin's (1992) single-subject ES was used to assess change in teacher behavior from baseline to posttests. The ESs were calculated by subtracting the mean of the treatment phase from the mean of the baseline phase and dividing the sum by the standard deviation of the baseline. The number of data points per phase was used in these computations rather than the number of participants. This method is sometimes referred to as the *No Assumptions* approach because there are no assumptions made about the normality of the distribution or the equality of variances.

identified initial practice goals. Jane chose to work on her usage of Concept Summaries, Academic Praise, and Behavioral Praise. The consultant and Jane discussed increasing the rate of these strategies during lessons. The meeting concluded with the consultant arranging times to observe the classroom for two lessons (i.e., math and literacy). The consultant then conducted two observations using the CSS.

During session 2, Jane's three goals were confirmed. The consultant and Jane first briefly reviewed the VPF of the CSS Part 1 eight teacher strategies collected by the consultant. Discussion then focused on the three identified strategy goals with a particular emphasis on the two instructional goals of Concept Summaries and Academic Praise. The consultant defined each strategy, modeled how to use each strategy, and provided Jane with a tip sheet with examples and suggestions on how to implement each strategy. Jane and the consultant established a plan to increase her usage of Concept Summaries and Academic Praise that would build upon Jane's strengths as a teacher. Following the session, the consultant conducted another two classroom observations using the CSS.

In the third session, the consultant and teacher briefly reviewed the VPF of the CSS Part 1 eight teacher strategies.

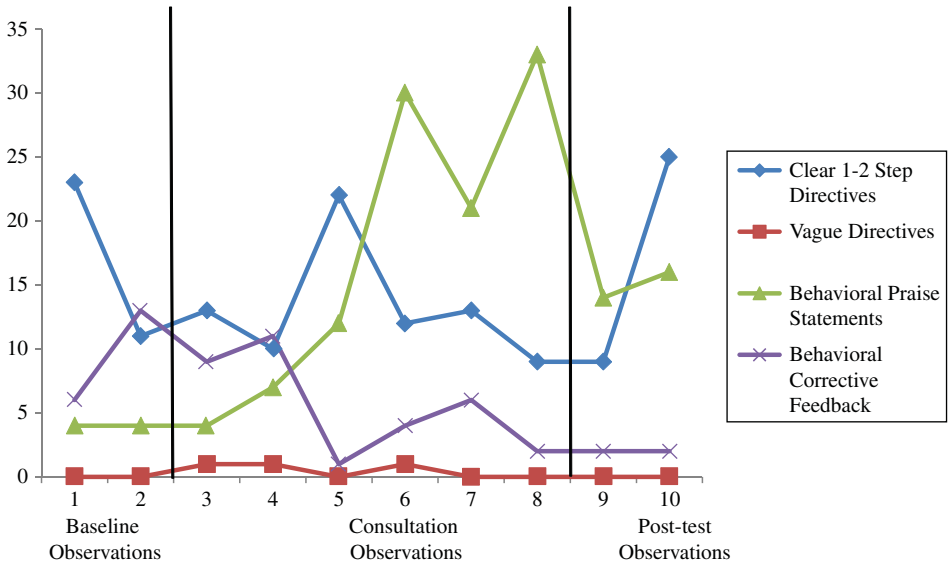

FIGURE 2 Visual analysis of the CSS Part 1 Teacher Behavioral Management Strategies.

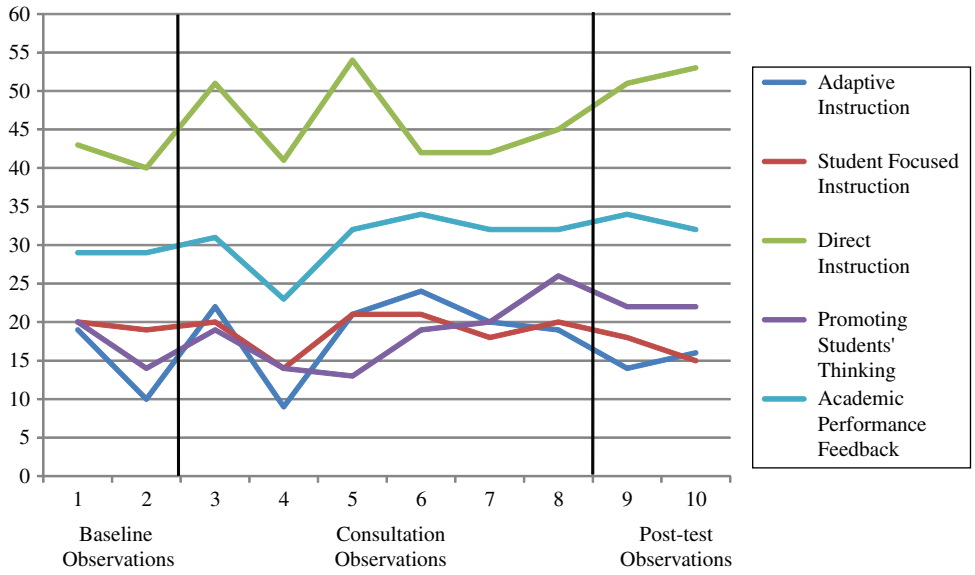

FIGURE 3 Visual analysis of the CSS Part 2 IS Subscale Frequency Scores.

The consultant provided Jane with positive reinforcement for her efforts to improve the two instructional strategy goals. Jane and the consultant reviewed the implementation plan for Concept Summaries and Academic Praise, as well as discussing sustainability. The session then focused on creating a plan for increasing Jane's rate of Behavioral Praise. Similar to session 2, the consultant defined and modeled Behavioral Praise and discussed strategies for implementation. Following the session, the consultant conducted another two classroom observations using the CSS.

In the final session, the consultant and Jane reviewed her progress on the CSS eight teacher strategies using VPF with a focus on the three identified goals. The consultant noted that Jane's increased usage of Academic Praise enabled her to quickly adopt and increase Behavioral Praise. Likewise, improvements in the teacher's use of Concept Summaries were discussed. Jane and the consultants reviewed the goals of the consultation process and discussed plans for sustainability.

Outcomes

Visual analysis of the CSS Part 1 scores presented in Figures 1 and 2 revealed positive improvements on the level (quantity) of Jane's use of praise for both academic performance and appropriate behavior. As consultation progressed, Jane's praise statements also increased (i.e., specifically labeling behaviors, immediacy) as measured by the Academic Performance Feedback subscale and Behavioral Praise subscale (Figures 3 and 4). Visual analysis also revealed that Jane's increased use of Academic Praise could be coupled with Academic Praise with Academic Response Opportunities. Although increasing Academic Response Opportunities was not an identified goal, Jane's usage of Praise and Academic Response Opportunities became synchronous near the end of the consultation process (Figure 1). We postulate that a feedback loop between Jane and her students occurred as Jane worked on implementing more Academic Praise into her repertoire.

At the beginning of the consultation process, Jane's usage of Academic Praise was relatively low compared to her Academic Response Opportunities usage. As Jane began to increase her usage of Academic Praise, her students were positively reinforced for engaging and interacting with Jane during the lesson. Over time, more students began interacting during the lesson to receive praise from Jane. Jane was similarly reinforced as her usage of Academic Praise prompted more engagement from the class and pleasurable exchanges with her students. Thus, providing her students with Academic Response Opportunities and following up with Academic Praise became a positive teaching sequence for Jane.

Additionally, Jane's increased usage of Behavioral Praise resulted in decreased usage of Behavioral Corrective Feedback. Praising a student for displaying appropriate behavior reinforces the appropriate behaviors in other students and subsequently reduces the need for corrective feedback. This finding was consistent with numerous studies showing that praise for appropriate behavior is an effective antecedent strategy for preventing problem behaviors in the classroom (e.g., Gable, Hester, Rock, & Hughes, 2009; Leflot, van Lier, Onghena, & Colpin, 2010).

Visual analysis also depicted increased usage (quantity) of the Part 1 Concept Summaries behavior (Figure 1). This increase in Concept Summaries paralleled an increase on

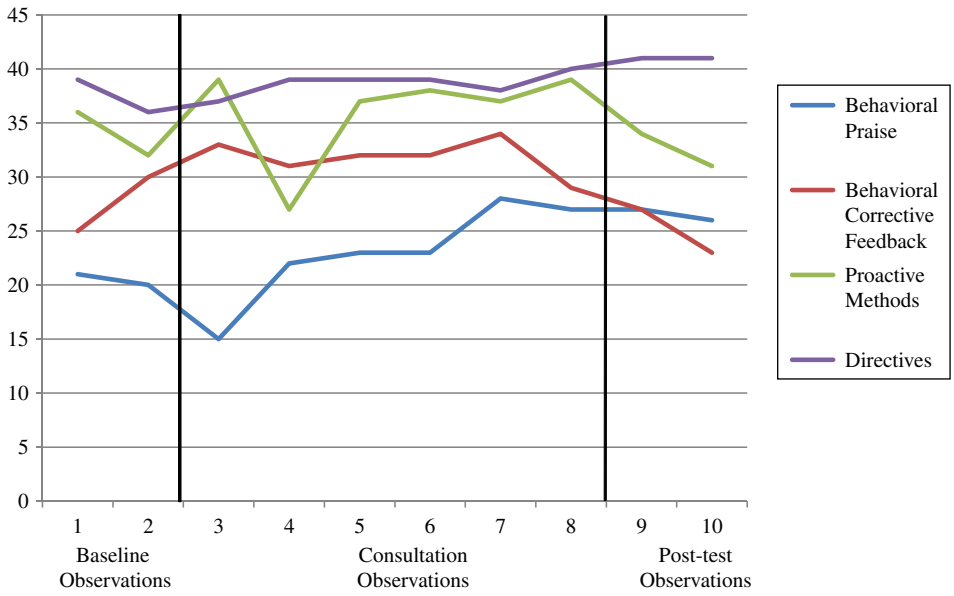

FIGURE 4 Visual analysis of the CSS Part 2 BMS Subscale Frequency Scores.

the Direct Instruction subscale (Figure 3). Summarizing important information is an effective strategy for promoting student academic outcomes and falls under Direct Instruction models of teaching (e.g., Marzano et al., 2001).

Time series graphs of the CSS Part 2 IS and BMS discrepancy scale scores also revealed positive results for Jane (Figures 5 and 6). As noted, IS and BMS discrepancy scale scores reflect teachers' need for change on specific practice domains. The larger the discrepancy scale scores, the greater need for change in that specific classroom practice. Jane's practice goals of increased use of Concept Summaries, Academic Praise, and Behavioral Praise as measured in Part 1 were noted. Based on these goals, a greater need for change at baseline would have been reflected in Jane's discrepancy scale scores for the IS Academic Performance Feedback subscale and the BMS Behavioral Praise subscale (Figures 5 and 6).

Throughout the consultation process, Jane's need for change (discrepancy scores) in the domains (subscales) of Academic Feedback, Behavioral Praise, and Behavioral

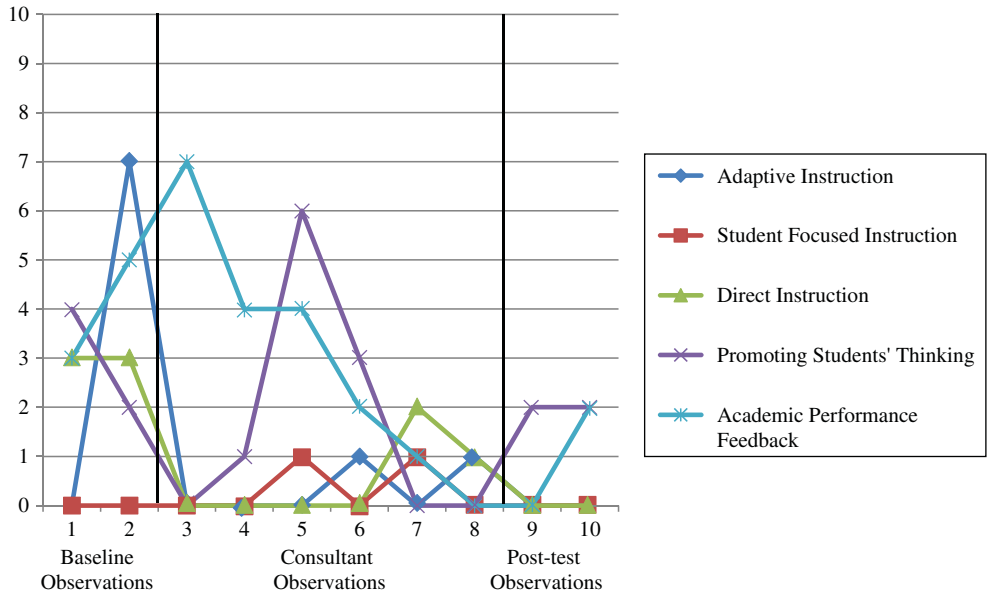

FIGURE 5 Visual analysis of the CSS Part 2 IS Subscale Discrepancy Scores.

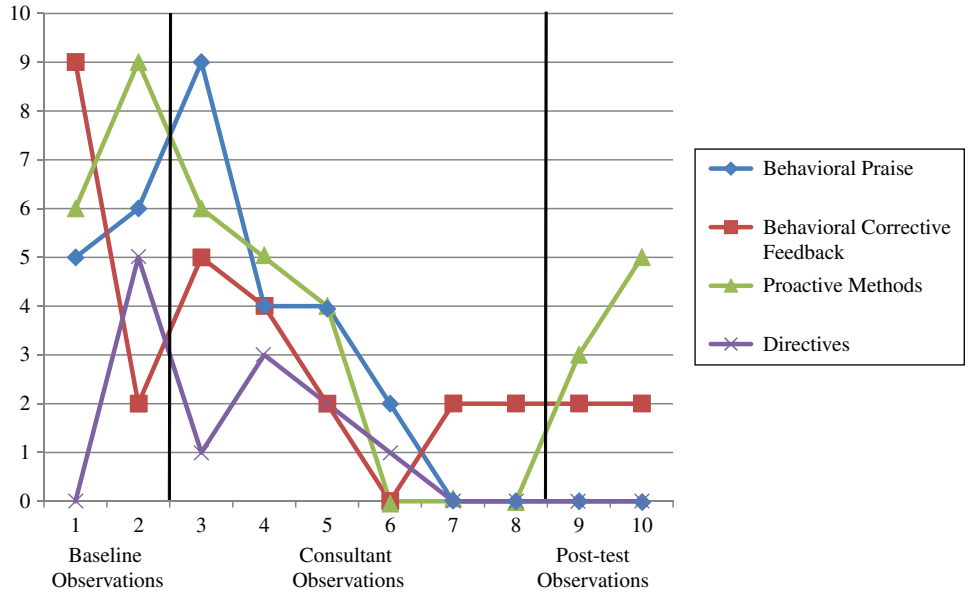

FIGURE 6 Visual Analysis of the CSS Part 2 BMS Subscale Discrepancy Scores.

Corrective Feedback gradually *decreased* and were comparatively lower at posttest. On the CSS, Praise and Corrective Feedback for behavior share key quality indicators that make them effective teacher strategies (i.e., specificity and immediacy). As Jane improved her rate of Behavioral Praise (Part 1), she subsequently improved the quality of her praise statements overall. This improvement in quality on Behavioral Praise (as measured on Part 2) generalized to Corrective Feedback (as measured on Part 1) and resulted in a decreased need for change on the BMS Corrective Feedback subscale (Figure 6).

As mentioned, single-case design effect sizes were computed to assess the practical significance of the teacher's change in classroom practices between baseline(s) and posttest CSS administrations (Busk & Serlin, 1992). Effect size (ES) comparisons between the baseline and posttest scores for the CSS Part 1 eight strategies and Part 2 Strategy Rating scales (IS and BMS) are presented in Tables 2 to 4. ESs were interpreted as follows: effect sizes of 0.20 to 0.49 were considered small, 0.50 to 0.79 medium, and 0.80 and above large (Cohen, 1988).

As shown on Table 2, Jane produced large positive changes in her use of Academic Praise (ES = 34.65) and Behavioral Praise (ES = 11.0). As noted, when Jane's use of Behavioral Praise increased, her need to use Corrective Feedback was reduced, which resulted in an ES of −1.52. Jane also successfully increased her usage of Concept Summaries (ES = 6.36).

As shown on Table 3, ESs were also calculated for the Part 2 IS and BMS scales (frequency ratings). As Jane's instructional goal of Part 1 Academic Praise improved (i.e., increased), the related Part 2 IS Academic Performance Feedback subscale evidenced an increase in feedback at posttest (ES = 4.0). Similarly, the Direct Instruction subscale yielded a positive effect size at posttest (ES = 4.95) in relation to Jane's increased usage of the Part 1 Concept Summaries. Jane's success in increasing her goal of Behavioral Praise (Part 1) reflected increased use in the Part 2 BMS Behavioral Praise subscale (ES = 8.49). However, the Part 2 BMS Corrective Feedback subscale evidenced a decrease in use (ES = −0.71). Although this was not ideal, we hypothesize that the decrease in Corrective Feedback subscale may have occurred due to the overall decrease in the usage of this strategy.

ESs for the CSS Part 2 IS and BMS discrepancy scores mirrored visual analysis results (Table 4). Jane's need for change (discrepancy score) on the IS Academic Performance Feedback subscale evidenced a large reduction at posttest (ES = −2.12; positive outcome). The decrease in the need for change was the result of Jane successfully increasing her usage of CSS Part 1 Academic Praise and improving quality aspects related to effective praise statements. Similarly, The BMS Praise subscale also yielded a robust negative effect size at posttest (ES = −7.78). Jane's goal of increasing rate of Concept Summary (Part 1) also resulted in a large reduction in her need for change on the Part 2 Instructional Delivery subscale (ES = −3.0).

DISCUSSION

This article highlights the theoretical and empirical basis of a user-friendly observational tool, the CSS-Observer Form for assessing teacher classroom practices. Grounded in effective instruction and behavioral management literatures, the CSS-Observer Form has been iteratively and rigorously developed and pilot tested with more than 400 classrooms. The initial work on the CSS provides good reliability and validity evidence as a tool for assessing and informing teacher classroom practices. Likewise, the CSS-Observer Form offers a promising addition to the small collection of teacher evaluation assessments in education worldwide.

The clinical utility of the CSS-Observer Form scores for assessing individual teachers' use of evidence-based instructional and behavioral management practices, formulating specific practice goals, and monitoring educators'

TABLE 3
Descriptive Statistics and Effect Sizes for the CSS Part 2 IS and BMS Subscale Frequency Scores

	Baseline		*Posttest*		
Strategy Rating Scales	*Mean*	*SD*	*Mean*	*SD*	*Effect size*
IS					
Adaptive Instruction	14.50	6.36	15.00	1.41	0.08
Student-Focused Learning	19.50	0.71	16.50	2.12	−4.24
Direct Instruction	41.50	2.12	52.00	1.41	4.95
Promotes Student Thinking	17.00	4.24	22.00	0.00	1.18
Academic Performance Feedback	29.00	0.00	33.00	1.41	4.00
BMS					
Behavioral Praise	20.50	0.71	26.50	0.71	8.49
Behavioral Corrective Feedback	27.50	3.54	25.00	2.83	−2.59
Proactive Methods	34.00	2.83	15.00	1.41	−6.72
Directives	37.50	2.12	41.00	0.00	1.65

TABLE 4
Descriptive Statistics and Effect Sizes for the CSS Part 2 IS and BMS Subscale Discrepancy Scores

	Baseline		*Posttest*		
Strategy Rating Scales	*Mean*	*SD*	*Mean*	*SD*	*Effect size*
IS					
Adaptive Instruction	3.50	4.95	0.00	0.00	−0.71
Student-Focused Learning	0.00	0.00	0.00	0.00	0.00
Direct Instruction	3.00	0.00	0.00	0.00	−3.00
Promotes Student Thinking	3.00	1.41	2.00	0.00	−0.71
Academic Performance Feedback	4.00	1.41	1.00	1.41	−2.12
BMS					
Behavioral Praise	5.50	0.71	0.00	0.00	−7.78
Behavioral Corrective Feedback	5.50	4.95	2.00	0.00	−0.71
Proactive Methods	7.50	2.12	0.00	0.00	−3.54
Directives	2.50	3.54	0.00	0.00	−0.71

progress towards practice goals is noted. The illustrative case example demonstrates the benefits of applying progress monitoring principles to the assessment and intervention of teacher classroom practices within a collaborative consultation model. The case of Jane underscores the importance of targeted assessment, specific practice goals, as well as visual performance feedback for promoting instructional improvement plans (goals) for educators. While the connection with teacher progress monitoring is critical for improving teacher classroom practices (Office for Standards in Education, 2006), few countries link reviewed performance with ongoing professional development opportunities (OECD, 2009). Margo, Benton, Withers, and Sodha (2008) noted some of the many problems facing teacher evaluation reform in England that include training inconsistencies and the inadequacy of professional development. They recommended strengthening the link between continuing professional development and increased monitoring (observations) during the review process. Similarly, Pochard (2008) highlighted the French evaluation system's lack of connection between professional development and teacher needs identified by the evaluation system.

There has been international recognition that school psychologists are uniquely positioned to contribute to the measurement and professional development of teachers' classroom practices and improvement of student academic outcomes (Farrell, Jimerson, Kalabouka, & Benoit, 2005). While the roles and functions of school psychologists vary from country to country, there is consensus on the need for school psychologists to use evidence-based approaches to assess and inform teachers' best practices. For more than 30 years, the field has discussed the critical role of school psychologists as instructional and behavioral consultants for teachers (Bergan & Kratochwill, 1990; Rosenfield, 2008). Leaders have called for an increased emphasis on classroom-wide best practices such as collaborative consultation and system-level interventions informed by data-based decision making with teachers and school administrators (Shapiro, 2006; Ysseldyke, 2005). The importance of progress monitoring and instructional and behavioral consultation are explicitly emphasized in the National Association of School Psychologist's *School Psychology: A Blueprint for Training and Practice III* (2006): "School psychologists should be instructional consultants who can assist parents and teachers to understand how students learn and what effective instruction looks like" (p. 13). Taken together, teacher progress monitoring is an important and underutilized practice in schools (Reddy, Fabiano, & Jimerson, 2013).

Thus, access to and implementation of validated, easy-to-use tools that measure educators' use of evidence-based practices for professional improvement plans are warranted. We believe that measures like the CSS-Observer Form can help school personnel in collaboratively improving teachers' classroom practices and student academic outcomes.

CONCLUSION

The CSS-Observer Form is a promising tool for school personnel to assess and guide educators' classroom instructional and behavioral management practices. Initial reliability and validity evidence offers a good foundation for school assessment of classroom practices. However, additional validation work is needed to fully maximize the CSS's utility for educational practice. Studies utilizing the CSS have been conducted in the northeastern United States. These findings may not generalize to other geographic regions, grade levels, teachers with particular training, or special education settings within the United States. Similarly, these findings may not generalize to international settings where the practice of education encompasses

different training, credentialing, and vastly different cultural contexts. Additional research on the CSS's predictive validity toward student academic outcomes, such as growth in achievement in the United States and countries abroad, is warranted. Also, studies that further examine the CSS's sensitivity to change following consultation would offer insight on the process of change in teacher practice and the sustainability of practice changes over time.

FUNDING

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through Grant R305A080337 to Rutgers University. The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education

ABOUT THE AUTHORS

Linda A. Reddy, PhD, is a Professor of School Psychology at Rutgers University. Her research interests include classroom assessment, ADHD assessment and intervention, and test development and validation. She is currently the principal investigator of two measurement development grants funded by the US Department of Education and author of numerous articles and books. She is a Fellow of APA and the President of APA Division 16 (School Psychology).

Christopher M. Dudek, MEd, is a research analyst on the School System Improvement Project, funded by the US Department of Education (Dr. Reddy, PI). His research interests include the development and validation of school assessments linked to targeted professional development efforts. He is a coauthor of numerous professional presentations and publications.

REFERENCES

Albert, S. R., Heward, W. L., & Hippler, B. J. (1999). Teaching middle school students with learning disabilities to recruit positive teacher attention. *Exceptional Children*, *65*, 253–270.

Alberto, A. A., & Troutman, A. C. (2003). *Applied behaviour analysis for teacher* (6th edition). Upper Saddle River, NJ: Merrill Prentice Hall.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Arbuckle, J. L. (2010). *IBM SPSS Amos 19 User's Guide, 587–590*. Crawordville, FL: Amos Development Corporation.

Avalos, B., & Assael, J. (2006). Moving from resistance to agreement: The case of the Chilean teacher performance evaluation. *International Journal of Educational Research*, *45*, 254–266.

Beaman, R., & Wheldall, K. (2000). Teacher's use of approval and disapproval in the classroom. *Educational Psychology*, *20*, 431–446.

Bennet, N. (1988). The effective primary school teacher: The search for a theory of pedagogy. *Teaching and Teacher Education*, *4*, 19–30.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, *1*(7), 10–22.

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

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. New York, NY: Oxford University Press.

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.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study.* Retrieved from www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Carlisle, J. F., Correnti, R., Phelps, G., & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing: An Interdisciplinary Journal*, *22*, 459–486.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Holt, Rinehart, and Winston.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Education Evaluation and Policy Analysis*, *24*, 81–112.

Dwyer, C. A., & Stufflebeam, D. (1996). Teacher Evaluation. In D. Bergliner & R. Calfee (Eds.), *Handbook of educational psychology.* New York, NY: Macmillan.

Farrell, P., Jimerson, S., Kalambouka, A., & Benoit, J. (2005). Teachers' perceptions of school psychologists in different countries. *School Psychology International*, *26*(5), 525–544.

Gable, R., Hester, P., Rock, M., & Hughes, K. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School & Clinic*, *44*(4), 195.

Gage, N. L. (1978). *The scientific basis of the art of teaching*. New York, NY: Teachers College Press.

Goldenberg, C., & Gallimore, R. (1991). Changing teaching takes more than a one-shot workshop. *Educational Leadership*, *49*(3), 69–72.

Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review*, *29*, 432–439.

Good, T. L., & Brophy, J. (1980). *Educational psychology: A realistic approach* (2nd ed.). New York, NY: Holt, Rinehart & Winston.

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

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on students' learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99–136.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2000). Elements of behavioral support plans: A technical brief. *Exceptionality*, *8*, 205–215.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005). School-wide positive behavior support. In L. Bambara & L. Kern (Eds.), *Individualized supports for students with problem behaviors: Designing positive behavior plans* (pp. 359–390). New York, NY: Guilford Press.

Isore, M. (2009). Teacher evaluation: Current practices in OECD countries and a literature review. *OECD Education Working Papers* No. 23. Paris: OECD Publishing. Retrieved from http://www.oecd.org/edu/workingpapers

Jackson, D. L., Gillapsy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23.

Kane, M. (2008). Terminology, emphasis, and utility in validation [Peer Commentary on the paper entitled A suggested change in terminology

and emphasis regarding validity and education]. *Educational Researcher*, *37*(2), 76–82.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, *21*, 31–41.

Kauchak, D., Peterson, K., & Driscoll, A. (1985). An interview study of teachers' attitudes toward teacher evaluation practices. *Journal of Research and Development in Education*, *19*(1), 32–37.

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

Leflot, G., van Lier, P. A. C., Onghena, P., & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the good behavior game. *Journal of Abnormal Child Psychology*, *38*(6), 869–882.

Margo, J., Benton, M., Withers, K., & Sodha, S. (2008). *Those who can?* Windhoek, Namibia: Institute for Public Policy Research (IPPR) Publications.

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-Continent Research for Education and Learning. (Eric Document Reproduction Service No. ED 427 087)

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

National Center on Student Progress Monitoring. (2006). *Common questions for progression monitoring*. Retrieved March 9, 2013, from www.studentprogress.org.

Organization for Economic Cooperation and Development (OECD). (2009). *Teacher evaluation: A conceptual framework and examples of country practices*. Paris, France: Author.

Office for Standards in Education. (2006). *The logical chain: continuing professional development in effective schools*. London, England: Author.

Peterson, D. R. (1995). The reflective educator. *American Psychologist*, *50*, 975–983.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System [CLASS] manual: Pre-K*. Baltimore, MD: Brookes.

Pochard, M. (2008). *Livre vert sur l'évolution du métier d'enseignant* Rapport au ministre de l'Education nationale, Collection des rapports officiels. Paris, France: La Documentation Française.

Porter, A. C., Youngs, P., & Odden, A. (2001). Advances in teacher assessment and their uses. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 259–297).

Reddy, L., Fabiano, G., Barbarasch, B., & Dudek, C. (2012). Behavior management of students with Attention-Deficit/Hyperactivity Disorders 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*. New York, NY: Continuum International.

Reddy, L. A., Fabiano, G., & Dudek, C. (2013). Concurrent validity of the Classroom Strategies Scale-Observer Form. *Journal of Psychoeducational Assessment*, *31*(3), 258–270.

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

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

Reddy, L., Fabiano, G., & Jimerson, S. (2013). Assessment of General Education Teachers' Tier 1 Classroom Practices: Contemporary science, practice and policy. *School Psychology Quarterly*, *28*, 273–276.

Riley-Tillman, T. C., & Eckert, T. (2001). Generalization programming and school based consultation. *Journal of Educational and Psychological Consultation*, *12*, 217–241.

Rose, D. J., & Church, R. J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education*, *8*, 5–35.

Rosenfield, S. (2008). Best practice in instructional consultation and instructional consultation teams. *Best practices in School Psychology*, *V*, 1645–1660.

Scheeler, M. C., Bruno, K., Grubb, E., & Seavey, T. L. (2009). Generalizing teaching techniques from university to K–12 classrooms: Teaching preservice teachers to use what they learn. *Journal of Behavioral Education*, *18*, 189–210.

Scheerens, J. (1992). *Effective schooling: Research, theory, and practice*. London: Cassell.

Schloss, P. J., & Smith, M. A. (1998). *Applied behavior analysis in the classroom*. Boston, MA: Allyn & Bacon.

Shapiro, E. S. (2006). Are we solving the big problems? *School Psychology Review*, *35*(2), 260–265.

Sparks, G. M. (1983). Synthesis of research on staff development for effective teaching. *Educational Leadership*, *41*(3), 65–72.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*, *26*, 333–368.

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relationship between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD: A review. *Remedial and Special Education*, *22*, 113–121.

Sutherland, K. S., Wehby, J. H., & Yoder, P. J. (2002). Examination of the relationship between teacher praise and opportunities for students with EBD to respond to academic requests. *Journal of Emotional and Behavioral Disorders*, *10*(1), 5–13.

Training and Development Agency for Schools. (2007). *Professional standards for teachers*. London, England: The Training and Development Agency for Teachers. Retrieved March 9, 2013, from www.tda.gov.uk/upload/resources/pdf/s/standards_a4.pdf

Walberg, H. J. (1986). Synthesis of research on teaching. In M. Wittorck (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2003). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Wang, M. C. (1991). Productive teaching and instruction: Assessing the knowledge base. *Phi Delta Kappan*, *71*, 470–478.

Wang, M. C., Haertel, G. D., & Walhberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*, 249–294.

Ward, B. (1985). Teacher development: The challenge of the future. In S. Hord, S. O'Neal, & M. Smith (Eds.), *Beyond the looking glass* (pp. 283–312). Austin: Research and Development Center for Teacher Education, University of Texas.

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

Yoon, K. S., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: US Department of Education, Institute of Education Science, National Center for Education and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.edu.gov/ncee/edlabs/regions/southwest/pdf/refl_2007033.pdf

Ysseldyke, J. (2005). Assessment and decision making for students with learning disabilities: What if this is as good as it gets? *Learning Disability Quarterly*, 125–128.