Promoting Teachers' Implementation of Culturally and Contextually Relevant Classwide Behavior Plans

Lindsay M. Fallon

Sadie C. Cathcart

Emily R. DeFouw

University of Massachusetts Boston

Breda V. O'Keeffe

University of Utah

George Sugai

University of Connecticut

Author Note. This research was supported in part by a fellowship from the Joseph P. Healey Research Grant Program sponsored by the Office of the President of the University of Massachusetts and the Office of the Vice Provost for Research and Strategic Initiatives & Dean of Graduate Studies of the University of Massachusetts Boston. In addition, the U.S. Department of Education's Institute of Education Sciences supported this research through Grant R324B170010 to the University of Massachusetts Boston. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Address correspondence to Lindsay M. Fallon, Department of Counseling and School Psychology, 100 Morrissey Boulevard, Boston, MA 02125. Email: lindsay.fallon@umb.edu

Abstract

Disproportionality in disciplinary actions for certain racial groups has been well documented for several decades. In an effort to support all students, specifically those who are culturally and linguistically diverse (CLD), many have called for adopting a multi-tiered system of support framework that is considerate of student culture and school context. This framework applies to supporting students' learning and behavior across settings, particularly in the classroom. To bridge existing gaps between theory and practice, this empirical study sought to evaluate if teachers who self-assessed their own use of culturally and contextually relevant practices would implement a classwide behavior plan with high levels of implementation fidelity. Results indicated that teachers who engaged in self-assessment and training did implement the plan with high levels of implementation fidelity, particularly when given performance feedback. Additionally, students tended to display slightly higher rates of academic engagement upon consistent implementation of the plan.

Keywords: diverse learners, implementation fidelity, single case design

Promoting Teachers' Implementation of Culturally and Contextually Relevant Classwide Behavior Plans

In 2014, the National Center for Education Statistics indicated that 50% of students enrolled in public elementary and secondary schools reported a racial identity other than White. Many identified as Black (16%) and Hispanic (25%), while others identified as Asian/Pacific Islander (5%), American Indian/Alaskan Native (1%), or two or more races (3%). Projections for 2026 indicate that percentages for some groups are expected to increase (i.e., Hispanic, Asian/Pacific Islander and two or more races), while the percentage of White students is expected to decrease (i.e., 45%; NCES, 2014). In contrast with the racial identities of students in the United States, most public school teachers identify as White (82%), female (76%) (NCES, 2012). This discrepancy is not problematic in and of itself; however, research results indicate that students from certain racial/ethnic groups are disproportionately referred to the office, and suspended and expelled from schools, possibly due to a cultural mismatch between students and teachers (Fenning & Rose, 2007; Kaufman et al., 2010; McIntosh, Moniz, Craft, Golby, & Steinwand-Deschambeault, 2014; Monroe, 2005; Noltemeyer & Mcloughlin, 2010; Skiba et al., 2011; Skiba, Michael, Nardo, & Peterson, 2002; Skiba, Peterson, & Williams, 1997; Wallace Jr, Goodkind, Wallace, & Bachman, 2008).

Black, Hispanic, and American Indian students are disproportionately referred for disciplinary action outside of the classroom as compared to their White peers (Gregory, Skiba, & Mediratta, 2017). Black students in particular are more frequently suspended and receive suspensions that are longer in duration (Kaufman et al., 2010), often in response to behaviors that are subjective in nature (e.g., insubordination, noncompliance) (Cartledge, Singh, & Gibson, 2008; Skiba et al., 2002). Black students also receive harsher consequences than White students

for the same disciplinary offenses (Browne-Dianis, 2011). These disparities appear to both impact students' perceptions of school climate and correlate to negative outcomes. Specifically, results from a recent large-scale analysis concluded that, as compared to their White peers, Black students at risk for out-of-school suspension reported lower ratings for perceived school equity and school belonging on a state school climate survey, and were at greater risk for adjustment problems (i.e., challenging, disruptive behavior) (Bottiani, Bradshaw, & Mendelson, 2017).

A number of factors may contribute to these patterns of disproportionality. In general, teachers report feeling underprepared to manage classroom behavior (Chesley & Jordan, 2012) that in part may be attributed to deficits in preservice training (Freeman, Simonsen, Briere, & MacSuga-Gage, 2014). To support students who are culturally and linguistically diverse (CLD) in particular, teachers may lack professional development opportunities during which they are encouraged to consider students' culture and the educational context to prevent misinterpretation of student behavior (Sugai, O'Keeffe, & Fallon, 2012). For instance, if a student has his/her head on a desk or uses inappropriate language to escape or avoid a difficult task, a teacher might interpret the behavior as exhibiting lack of motivation, disinterest in learning, and/or signs of a behavior disorder (Delpit, 2012). Other researchers have theorized that cultural mismatch or misunderstanding between teachers and students (Delpit, 2006; Irvine, 1990), racial stereotyping by school staff (Skiba et al., 2011), and/or conscious or unconscious racial bias by teachers (McIntosh et al., 2014) contribute to disproportionality in disciplinary action.

Supporting CLD Learners

To reduce disproportionality in disciplinary actions and support students' learning and behavior in the classroom, numerous researchers, policymakers, and practitioners have called for the development of a culturally responsive, multi-tiered system of support framework in schools

(Bal, Thorius & Kozleski, 2012; Bohanon et al., 2006; McCurdy, Mannella & Eldridge, 2003; Vincent, Swain-Bradway, Tobin, & May, 2011). To accomplish this goal, a few steps appear critical. First, enhanced teacher training and professional development that focuses on universal prevention is needed (i.e., Tier 1; Skiba et al., 2002). Second, a shared focus on all educators supporting both students' academic achievement and social behavior competence is necessary (Algozzine, Putnam & Horner, 2012; Sutherland, Lewis-Palmer, Stichter, & Morgan, 2008). Third, consideration of the school context, classroom ecology and students' culture are important when identifying, designing and implementing supports (e.g., Cartledge et al., 2008; Green, 2005). Finally, classroom practices should be evidence-based (Horner et al., 2005), implemented with high levels of implementation fidelity (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008), and produce student progress data that can be monitored over time (Cook & Odom, 2013).

Supporting student classroom behavior. To provide more specific recommendations to support CLD students in particular, Fallon, O'Keeffe and Sugai (2012) conducted a systematic literature review that resulted in identification of a number of empirically supported, culturally and contextually relevant behavior management practices. These practices included increasing positive, equitable interactions with students, setting high expectations for the class, teaching social skills, and using students' culture and language in instruction (Fallon et al., 2012), and were integrated into a model for positively supporting the behavior of CLD learners using Positive Behavioral Interventions and Supports (PBIS; Sugai et al., 2012). This multi-tiered framework included recommendations for enhancing contextual and cultural relevance of PBIS *outcomes* (e.g., translate school-wide social expectations and behaviors into languages of students and families), *data* (e.g., develop data summarization, analysis, and presentation

procedures that are considerate of cultural and contextual factors and learning histories), *systems* (e.g., ensure that membership of school leadership team is representative of the cultural groups of the school and community) and *practices* (e.g., develop lesson plans, posters, practices, activities, etc. with language, images, messages, etc. that are appropriate across cultural groups and school contexts). By integrating these considerations into PBIS, teachers might improve outcomes for students across all cultural and linguistic backgrounds.

Subsequently, Fallon, O'Keeffe, Gage, and Sugai (2015) surveyed school staff members (n = 330) nationwide to assess the social validity of the recommendations proposed by Sugai et al. (2012) and Fallon et al. (2012). Survey results indicated that school staff found the recommendations to be acceptable, feasible, and potentially efficacious in support of the behavior of all students, particularly CLD learners. These recommendations were then used to answer a call for more reliable and valid measures to accurately assess teachers' use of culturally relevant practices (Debnam, Pas, Bottiani, Cash, & Bradshaw, 2015). Although there exists a few tools that measure a single construct of culturally and contextually relevant practice or a teachers' beliefs about his/her ability to deliver culturally responsive instruction (e.g., *Culturally-Responsive Teaching Self-Efficacy Scale*, Siwatu, 2007; *Double Check Self-Reflection Tool*, Hershfeldt et al., 2009; *Multicultural Efficacy Scale*, Guyton & Wesche, 2005), these tools do not comprehensively assess instructional and behavior/social *practices*, use of *data* in assessing progress and decision making, and access to training and support *systems* (Sugai et al., 2012; Vincent et al., 2011). As such, a teacher self-assessment tool, the *Assessment of Culturally and Contextually Relevant Supports* (ACCReS; available from first author upon request) was developed. Items included recommendations suggested for student behavior as well as academic

achievement, teachers' use of data in the classroom, and teachers' access to professional development opportunities.

ACCReS content validity. The teacher self-assessment items were created and sent for review to 10 subject-matter experts in the field including researchers who had published extensively in the areas of behavior management (n = 3), academic achievement (n = 3), or CLD learners (n = 4). All were university professors. Six held the rank of full professor, one was an associate professor, and three held the rank of assistant professor. To recruit these individuals, the current study's authors created a list of possible experts, then the first author reviewed each nominee's most recent curriculum vitae to confirm he/she had an extensive publication record (> 5 publications) related to one of the content areas identified above. Experts were then contacted via email and asked to participate in reviewing the items. All who were contacted agreed to review and were paid a small stipend for their efforts (i.e., $100 stipend). Experts reviewed items and identified (a) the factor with which the item aligned (i.e., academic practice, behavior practice, use of data, access to professional development), (b) how certain the reviewer was of his/her factor assignment, and (c) perceived relevance of the item to that factor. Data were then analyzed to determine the content and face validity of the tool. Generally, the majority of items were (a) correctly identified, (b) rated with certainty, and (c) reported to be highly relevant to the tool. The self-assessment was then revised to incorporate feedback from subject-matter experts and continues to be subject to validation activities for use in school settings, including initial pilot testing its use in intervention research in schools.

Intervention with fidelity. In addition to assessment, intervention is critical to addressing behavioral concerns in the classroom. As previously described, comprehensive

classwide behavioral intervention should include consideration of empirical recommendations for instructional practice, as research has linked achievement and social behavior (Algozzine, Wang, & Violette, 2011; Marchant & Anderson, 2012) in addition to the use of data to support decision-making (Eagle, Dowd-Eagle, Snyder, & Holtzman, 2015). Implementation of a comprehensive classwide intervention should be consistent and complete (i.e., with high levels of implementation fidelity) (Sanetti & Kratochwill, 2009). Implementation fidelity data are important in determining the impact of a comprehensive intervention on student outcomes (e.g., academic engagement, disruptive behavior; Sanetti, Chafouleas, Fallon**,** & Jaffery, 2014). That is, if implementation is low, educators may require support to improve intervention implementation to maximize outcomes.

Implementation support. When educators struggle to deliver classroom-based interventions, ongoing implementation support may be useful (Sanetti & Collier-Meek, 2015). Implementation support can include a variety of activities ranging from requesting that an educator self-monitor his/her own implementation (Simonsen, MacSuga, Fallon, & Sugai, 2013), to providing an educator with comprehensive performance feedback related to implementation (Sanetti et al., 2014). Performance feedback includes offering an educator praise for steps of the intervention implemented completely, and suggestions for how to improve implementation of intervention steps regularly missed (Collier-Meek, Fallon, Sanetti, & Maggin, 2013). Implementation support can range from low (e.g., self-monitoring) to high (e.g., performance feedback) levels of intensity. Researchers propose providing more intensive strategies to educators once less intensive strategies have proven ineffective (Sanetti & Collier-Meek, 2015). Ultimately, implementation support is intended to bolster educators' intervention delivery to support students' learning and behavior in the classroom.

Purpose of Study

For educators seeking to improve support provided to all students, particularly CLD learners, we recommend a comprehensive classwide behavior support plan and the provision of implementation training for educators. In this study, teachers were asked to self-assess their practice with the *ACCReS* tool. Teachers' self-assessment results were subsequently analyzed and a universal classwide behavior support plan was developed based on items teacher participants consistently rated the lowest. Teachers were then trained to implement the plan and asked to self-monitor implementation. When implementation fidelity fell below 80%, teachers received performance feedback. In addition to teachers' implementation of the classwide plan, students' academic engagement and disruptive behavior were tracked. Specifically, the purpose of the current study was to address the following research questions:

1. Will engaging in self-assessment and subsequent training on a comprehensive classwide plan increase teachers' use of culturally and contextually relevant strategies in the classroom?
2. Will teachers implement the classwide plan consistently and comprehensively while self-monitoring implementation? If needed, will performance feedback increase teachers' use of culturally and contextually relevant strategies in the classroom?
3. Will students' displays of academic engagement increase and disruptive behavior decrease upon teachers' implementation of the classwide plan (i.e., use of culturally and contextually relevant strategies in the classroom)?

The current study employed a single case research design to investigate experimentally the above research questions.

Method

Participant Recruitment and Setting

The first author solicited participation by contacting the superintendent of an urban district in the Northeast. Once the superintendent approved the study, district principals were contacted via email. One principal of a middle school in the district expressed interest. The middle school included Grades 6-8. Of the 275 students enrolled, 47.3% of students were female, 60% identified as a race other than White (Black = 30.9%; Hispanic = 10.5%; Asian = 9.8%; Multiracial = 7.6%), 26.5% were students with disabilities, 16.7% spoke English as a second language, and 28.0% were considered economically disadvantaged.

The principal met with the first author about the study's purpose. Afterward, the principal and the first author presented details about the study to teachers at a staff meeting. During the meeting, two sixth grade and a seventh grade teacher volunteered to participate in the study with students in the one class period for which they were noticing the most disruptive behavior. Once informed consent was secured for teacher participants, parents in participating classrooms were notified about the study and given the option to opt their child out of study participation; however, no parents selected this option.

Participating teachers included one Caucasian male, one Black female, and one Hispanic female. Teachers ranged in age from 28 – 40 (M = 35.7 years) and had three to eight years of teaching experience (M = 5.6 years). All had at least a master's degree and were certified as general education teachers. Each participating class had 23 or 24 students and reflected the demographic trends of the school population. The number of females in each class ranged from 9 – 14 (39 - 61%). Furthermore, students who identified as (a) Black ranged from 6 – 13 (30 – 57%), (b) Caucasian ranged from 6 – 11 (26 – 48%), (c) Asian ranged from 1 – 4 (4 – 17%), (d) Hispanic/Latino ranged from 1 – 3 (4 – 13%), and Multiracial ranged from 0 – 3 (0 – 13%).

The consultant in the current study, a doctoral-level school psychologist and board certified behavior analyst, was a faculty member at a nearby university. Data collectors were two doctoral students in school psychology with advanced coursework in behavior assessment. Training for data collectors involved three steps. First, the consultant trained data collectors on how to use the data collection forms in detail, providing data collectors with opportunities to ask clarifying questions and discuss various scenarios in data collection. Second, data collectors practiced conducting direct observation with video clips of classwide student behavior prior to conducting classroom observations. Third, data collectors conducted a classroom observation with the consultant. Disagreements in the observation data between consultant and data collectors were discussed. Once acceptable interobserver agreement (i.e., > 90%) on training had been achieved, data collectors completed observations independently and with the consultants for interobserver agreement.

Independent Variables

Self-Assessment, Plan Development, and Plan Implementation. The independent variable in the current study contained multiple components. First, each teacher participant completed the *ACCReS* online (via www.SurveyMonkey.com) independently. The first author reviewed results from the three teachers' responses and identified items universally rated lower than other items (e.g., "*My lessons reflect students' interests and preferences*."). These items were used to create action steps (e.g., "*Ask students to provide examples related to the content of the lesson that reflect their interests, experiences and/or preferences*."). These action steps were compiled to create a universal classwide behavior support plan.

Data collectors observed teachers to determine which steps were indeed lacking in the classroom. For each teacher, nine of the 13 steps were not being implemented consistently in the

classroom, and differed by teacher (see Appendix A for steps each teacher was trained to implement). After baseline data collection, each teacher received comprehensive training on the plan. The completion of the self-assessment, training, and implementation of the plan served as the primary multi-component independent variable.

Self-Monitoring. Once teachers received training, teachers were asked to self-monitor their implementation of the plan daily within the class period they identified during recruitment. Teachers indicated if each step of the plan was implemented fully, partially, or not at all that day. Data collectors collected teachers' self-monitoring data once per week.

Self-Monitoring + Performance Feedback. When a teacher's implementation was determined to be below 80% for three consecutive days, he/she was eligible to enter the Self-Monitoring + Performance Feedback Phase. In this phase, the teacher continued to self-monitor, but also receive a daily email before implementing the plan each day. The email included (a) a graph depicting the percentage of the plan implemented during the last observation, (b) a graph depicting the percentage of the plan implemented over the last five observations, (c) praise for three randomly selected steps for which the teacher was implementing fully, and (d) feedback to increase implementation for three randomly selected steps for which the teacher was not implementing completely. Teachers were asked to respond to the email when it was read, and a read receipt was requested in the email. Mr. Johnson (88.9%) and Ms. Watson (85.7%) opened all but one email, while Ms. Garcia (100%) opened every performance feedback email immediately before the next observation during this phase.

Dependent Variables

To evaluate the impact of the independent variables, individual teacher and classwide student data were collected. Data collectors gathered observation data in classrooms three times

per week throughout the study. All observations took place for 40 min during the instructional period identified by the teacher during recruitment (i.e., the class period with the most disruptive student behavior): Mr. Johnson during Period 1 English class, Ms. Watson during Period 2 Social Studies class, and Ms. Garcia during Period 4 English class. Teacher social validity data were collected as well.

Treatment fidelity. Data collectors used a researcher-derived measure of implementation fidelity to evaluate adherence (i.e., steps implemented as planned). Each step of the plan was listed and rated using the following scale*: implemented fully, implemented partially or with deviation, not implemented at all,* or *no opportunity to implement*. Similar measures have been used in previous research studies to provide accurate estimates of implementation (e.g., Sanetti & Collier-Meek, 2015). For 30% of observations sessions across all phases, two data collectors observed so that interobserver agreement could be calculated by dividing the total number of intervals for which there was agreement between two raters by the total number of intervals, then multiplying by 100 (i.e., percent agreement) (Cooper, Heron, & Heward, 2007). Acceptable agreement was documented for treatment fidelity for Mr. Johnson (M = 94.6%; R = 85.7 – 100%), Ms. Watson (M = 93.3%; R = 77.8 – 100%), and Ms. Garcia (M = 95.5%; R = 88.9 – 100%).

Classwide student behavior. Data collectors also used a researcher-derived measure of classwide academic engagement and disruptive behavior using a momentary time sampling procedure with 15-sec intervals for the entire observation (Ferguson, Briesch, Volpe, & Daniels, 2012). Data collectors employed an individual-random observation scheme, where observers randomly selected a different student in the class to observe during each consecutive 15-sec interval (Briesch, Hemphill, Volpe, & Daniels, 2015). Such measures have been previously used

in research and demonstrated to provide accurate estimates of student behavior (Chafouleas, Sanetti, Kilgus, & Maggin, 2012).

Two data collectors observed 30% of observations sessions across all phases for the purpose of calculating interobserver agreement by dividing the total number of intervals for which there was agreement between two raters by the total number of intervals, then multiplying by 100 (i.e., percent agreement) (Cooper et al., 2007). Acceptable agreement was documented for student academic engagement for Mr. Johnson (M = 97.7%; R = 91.6 – 100%), Ms. Watson (M = 99.1%; R = 95.5 – 100%), and Ms. Garcia (M = 97.9%; R = 94.5 – 100%), and disruptive behavior for Mr. Johnson (M = 99.2%; R = 98.5 – 100%), Ms. Watson (M = 99.3%; R = 97.6 – 100%), and Ms. Garcia (M = 97.9%; R = 94.5 – 100%) (Kratochwill et al., 2010).

Social validity. Teachers were asked to complete an adapted version of the Usage Rating Profile – Intervention Revised (URP-IR) (Briesch, Chafouleas, Neugebauer, & Riley-Tillman, 2013). Items on URP-IR aim to reliably assess the acceptability (α =.96), understanding (α =.90), and feasibility (α =.85) of implementing a school-based intervention by having educators rate the extent to which he/she agrees or disagrees with 35 statements (e.g., “This intervention is a good way to handle the child’s behavior problem”) using a six-point Likert scale (1 = *strongly disagree* to 6 = *strongly agree*).

Design

Study procedures followed an experimental single case research design, specifically, a randomized A-B-C-D multiple baseline across participants design. This research design allows an opportunity for an effect replication within and across at least three or more participants, settings, or behaviors to control for threats to internal validity (e.g., maturation, history)

(Kratochwill et al., 2010). All participants entered the Student Baseline Phase (Phase A) at the same time to collect baseline data for students' academic engagement and disruptive behavior.

After five observations, all teachers completed the *ACCReS* and began the Implementation Baseline Phase (Phase B). After three observations in this phase, Mr. Johnson and Ms. Watson demonstrated low, stable implementation. Mr. Johnson was randomly selected (using www.random.org) to receive training and enter the Self-Monitoring Phase (Phase C). Ms. Watson and Ms. Garcia remained in the Implementation Baseline Phase (Phase B). After five additional observations, Ms. Watson and Ms. Garcia demonstrated low and either stable or decreasing implementation, and were eligible for training to enter the Self-Monitoring Phase (Phase C). Ms. Watson was randomly selected and Ms. Garcia remained in the Implementation Baseline Phase (Phase B) for an additional three observations before entering the Self-Monitoring Phase (Phase C). Once each teacher implemented the intervention below 80% for three observation periods, the teacher became eligible to be staggered into the Self-Monitoring + Performance Feedback Phase (Phase D).

This design meets *What Works Clearinghouse (WWC)* standards for a multiple baseline design as (a) the independent variable was systematically manipulated, (b) interobserver agreement was assessed across at least 20% observations, and (c) at least three replications were used to demonstrate an effect across participants (Kratochwill et al., 2010). However, this study meets standards with reservations since there were not six, but at least three, data points per phase for the primary dependent variable (i.e., implementation fidelity) (Kratochwill et al., 2010).

Procedures

As previously noted, data collectors observed students' behavior (and later, teachers' implementation fidelity) for 40 min across phases during each observation for three days a week each week over the course of three months. After student baseline data were collected during the Student Baseline Phase (Phase A), teachers completed the *ACCReS* online via Survey Monkey (www.surveymonkey.com) before the first observation in the Implementation Baseline Phase (Phase B). The first author reviewed teachers' responses to identify items rated lower than other items universally (i.e., across the three teachers). As assessment items pertain to educators' use of behavioral supports, academic practices, and data, the first author selected items rated low from each of these three sections to target creating action steps to populate a universal classwide behavior support plan. Items on the *ACCReS* also assess professional development, but these items were not used to inform the creation of the universal classwide plan used in the current study, as the plan targeted action steps that could be implemented daily.

For items rated low by all three teachers on the *ACCReS,* the first author generated corresponding action step(s) to include in the universal classwide plan (see Appendix A). Once action steps were generated, data collectors observed teachers again, the Implementation Baseline Phase (Phase B) was initiated, and data collectors observed student behavior (academic engagement and disruptive behavior). During this phase, teachers had not yet been trained on the steps of the plan, yet observers assessed for teachers' baseline adherence to steps of the plan while observing to determine if teachers were implementing steps of the plan despite provided a low rating for the corresponding item(s) on the *ACCReS.* The first author noted that each teacher inconsistently implemented nine of the 12 steps of the intervention. Immediately before each teacher was staggered into the Self-Monitoring Phase (Phase C), the first author provided training to that teacher to implement these nine steps in practice.

The training entailed a detailed description of each intervention step and verbal modeling for how teachers might implement the step in practice. Teachers were given comprehensive verbal descriptions for how to self-monitor their implementation using the treatment fidelity form researchers used during observations. Finally, teachers were asked if they had any questions before beginning to implement the plan the next day in class. In total, this training took 30-45 min.

Once teachers received training, they were asked to implement the plan the next day and to self-monitor implementation (Phase C). As described above, once implementation fell below 80% for three observation periods (as evidenced by data collectors' implementation fidelity ratings), teachers were eligible to receive performance feedback emails. They were once again staggered into this condition (Phase D). Email content included a graph of implementation fidelity data from the previous observation and the past five observations, praise for three steps implemented consistently, and suggestions to improve three steps not implemented consistently. Data collectors continued to track students' behavior and teachers' implementation fidelity. At the completion of the study, teacher participants were given a $60 Amazon gift card as compensation for their voluntary involvement in this study.

Data Analysis

For data reflecting teachers' implementation fidelity and students' behavior, graphed data (see Figure 1 and 2) and descriptive results (mean, standard deviation; see Table 1 and 2) were analyzed by phase. Visual analysis of graphed data was supplemented with the calculation of TauU (Parker, Vannest, Davis, & Sauber, 2010) using an online calculator (singlecaseresearch.org). Tau-U was used to provide a quantitative estimate of study effects by participant across phases. Guidelines used for interpreting Tau-U results aligned with those

applied in other single case design research studies (small = 0–0.65, moderate = 0.66–0.92, large = 0.93–1.00) (Sanetti, Collier-Meek, Long, Byron, & Kratochwill, 2015).

Results

Treatment Fidelity

During Implementation Baseline (Phase B), all teachers implemented the classroom plan with low levels of implementation fidelity (see Figure 1 and Table 1). Upon the introduction of the Self-Monitoring Phase (Phase C), two of the three teachers' levels of implementation fidelity substantially increased, while one teacher's levels slightly increased. When performance feedback was introduced with self-monitoring (Phase D), all teachers' levels of implementation fidelity increased.

Mr. Johnson demonstrated stable, but low levels of implementation fidelity in the Implementation Baseline Phase. During the Self-Monitoring Phase, Mr. Johnson's implementation fidelity immediately increased, indicated by an upward trend. Specifically, for Mr. Johnson, average treatment fidelity increased 29.8% from the Implementation Baseline to the Self-Monitoring Phase, a large effect (Tau-U = 1.00). During the Self-Monitoring + Performance Feedback Phase, Mr. Johnson's levels of implementation fidelity decreased slightly but remained stable and above Implementation Baseline. Additionally, average implementation fidelity increased 42.4% from the Implementation Baseline to Self-monitoring + Performance Feedback indicating a large effect (Tau-U = 1.00).

Ms. Watson implemented the classroom plan with variable and low levels of implementation fidelity during the Implementation Baseline phase. During the Self-Monitoring Phase, Ms. Watson's implementation fidelity increased slightly while remaining variable. For Ms. Watson, average implementation fidelity increased 5.7% from the Implementation Baseline

to Self-Monitoring Phase, a small effect (Tau-U = 0.34). During the Self-monitoring + Performance Feedback Phase, Ms. Watson's implementation fidelity demonstrated an increasing trend and average implementation fidelity increased 17.5% from the Implementation Baseline Phase, a moderate effect (Tau-U = 0.74).

Finally, Ms. Garcia initially implemented the intervention with moderate, but variable levels of implementation fidelity during the Implementation Baseline Phase. During the Self-Monitoring Phase, the level of implementation fidelity increased substantially. For Ms. Garcia, average implementation fidelity increased 33.0% from the Implementation Baseline to Self-Monitoring Phase, a moderate effect (Tau-U = 0.88). During the Self-Monitoring + Performance Feedback Phase, Ms. Garcia's implementation fidelity increased, but became more variable. For Ms. Garcia, average implementation fidelity increased 43.4% from the Implementation Baseline to the Self-Monitoring + Performance Feedback Phase, a moderate effect (Tau-U = 0.91).

Student Outcomes

In addition to implementation fidelity improvements across phases, students' behavior also improved throughout the study (see Table 2). For Mr. Johnson's class, academic engagement increased from Student Baseline to the Implementation Baseline Phase (+ 5.1%). Subsequently, an increase was noted from the Implementation Baseline to the Self-Monitoring Phase (+ 5.7%), resulting in a small effect (Tau-U = 0.38), and Implementation Baseline to Self-Monitoring + Performance Feedback Phase (+ 8.2%), resulting in a small effect (Tau-U = 0.51). Levels of disruptive behavior remained constant from Student Baseline to the Implementation Baseline Phase (+ 0.03%). The Tau-U estimate for the difference between academic engaged behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a moderate effect (Tau-U = 0.88). Subsequently, levels

of disruptive behavior was constant from the Student Baseline to Implementation Baseline (+0.3%). Levels were also constant from Implementation Baseline to the Self-Monitoring Phase (+ 0.9%), indicating a small effect (Tau-U = 0.13). Also, levels increased slightly from Implementation Baseline to Self-Monitoring + Performance Feedback Phase (+ 1.0%), resulting in a small effect (Tau-U = 0.07). The Tau-U estimate for the difference between disruptive behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a very small effect (Tau-U = 0.12).

Ms. Watson's class demonstrated higher levels of academic engagement in comparison to Mr. Johnson's and Ms. Garcia's class from Baseline Student to the Implementation Baseline Phase (+ 9.4%). In addition, Ms. Watson's class' levels of academic engagement increased from Implementation Baseline to the Self-Monitoring Phase (+ 2.9%), indicating no effect (Tau-U = 0.05), and Implementation Baseline to the Self-Monitoring + Performance Feedback Phase (+ 4.4%), a small effect (Tau-U = 0.08). The Tau-U estimate for the difference between academic engaged behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a moderate effect (Tau-U = 0.65). Levels of disruptive behavior remained relatively constant from Student Baseline to the Implementation Baseline Phase (- 0.4%), and from Implementation Baseline to the Self-Monitoring Phase (- 0.4%), indicating a small effect (Tau-U = 0.33). Levels of disruptive behavior slightly decreased from the Implementation Baseline Phase to the Self-Monitoring + Performance Feedback Phase (-1.3%), demonstrating a small effect (Tau-U = 0.14). The Tau-U estimate for the difference between disruptive behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a very small effect (Tau-U = 0.14).

Finally, Ms. Garcia's class demonstrated increased academic engagement during

Implementation Baseline Phase as compared to the Student Baseline (+ 11.4%). Additionally, academic engagement increased slightly from Implementation Baseline to the Self-Monitoring Phase (+ 2.9%), indicating no effect (Tau-U = 0.05), and from the Implementation Baseline to Self-Monitoring + Performance Feedback Phase (+ 0.6%), indicating no effect (Tau-U = 0.08). For disruptive behavior, a more substantial decrease was noted from Student Baseline to the Implementation Baseline Phase (- 4.2%). The Tau-U estimate for the difference between academic engaged behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a small effect (Tau-U = 0.35). Levels of disruptive behavior increased from the Implementation Baseline to the Self-monitoring Phase (+ 2.7%), a small effect (Tau-U = 0.44). Despite this increase, disruptive behavior decreased from the Implementation Baseline to Self-Monitoring + Performance Feedback Phase (- 2.7%), indicating no effect (Tau-U = 0.02). The Tau-U estimate for the difference between disruptive behavior during Student Baseline + Implementation Baseline Phase and Self-Monitoring + Performance Feedback Phase indicated a very small effect (Tau-U = 0.16).

Social Validity

Overall, teachers responded favorably to most items on the URP-IR. Specifically, the mean teacher rating overall was 4.43 (*SD* = 1.20), with a mean of 4.63 for acceptability (*SD* = 0.97), 4.89 for understanding (*SD* = 0.33), 4.06 for feasibility (*SD* = 1.51), and 4.22 for systems support (*SD* = 1.09). Mr. Johnson reported that the classroom plan was very acceptable (*M* = 5.33; *SD* = 0.71), very understandable (*M* = 5.00; *SD* = 0.00), very feasible (*M* = 5.17; *SD* = 0.41), and required some systems support (*M* = 3.67; *SD* = 1.15). Ms. Watson reported that the classroom plan was somewhat acceptable (*M* = 4.78; *SD* = 0.44), very understandable (*M* = 5.00; *SD* = 0.00), somewhat feasible (*M* = 4.00; *SD* = 1.79), and required some systems support (*M* =

4.67; *SD* = 0.58). Ms. Garcia reported that the classroom plan was somewhat acceptable (*M* = 3.78; *SD* = 0.97), understandable (*M* = 4.67; *SD* = 0.58), somewhat feasible (*M* = 3.00; *SD* = 1.51), and required some systems support (*M* =4.33; *SD* = 1.53).

Discussion

Schools are becoming increasingly diverse settings serving students from a variety of racial, language, and ethnic backgrounds. Researchers, educators and policymakers are proposing recommendations for behavioral strategies that support all students, particularly CLD learners. These efforts are critical considering the extent to which research results indicate disproportionality related to school discipline for certain groups over the past several decades (e.g., Skiba et al., 2002). To address this issue, many researchers and practitioners have proposed culturally and contextually relevant approaches to behavior management framed within a multi-tiered system of support model (Bal et al., 2012; Bohanon et al., 2006; McCurdy et al., 2003). This framework is particularly flexible to support students' behavior schoolwide (Vincent et al., 2011), but also can be used in the classroom by encouraging teachers to implement empirically-supported, Tier 1 culturally relevant behavior and academic practices, as well as use data to inform decision-making (Sugai et al., 2012). Although previous research in this area has been largely theoretical, the results of this experimental study are promising,

This study is the first to empirically evaluate the impact of self-assessment and teacher training in the use of culturally and contextually relevant strategies in the classroom. It took place in a middle school, a setting in which limited intervention research is available (Lane, Oakes, Carter, & Messenger, 2015). Results from the current study provide preliminary data related to the impact of performance feedback on teachers' implementation fidelity of a

classwide plan, as well as the impact of the plan on students' academic engagement and disruptive behavior.

On average, all teachers implemented more components of the classwide plan during the Self-Monitoring Phase than during Implementation Baseline, and even more during Self-Monitoring + Performance Feedback versus both the Implementation Baseline and Self-Monitoring Phases alone (see Table 1). Specifically, Figure 1 illustrates that teachers' implementation was relatively low and stable during Implementation Baseline. During the Self-Monitoring Phase, all teachers demonstrated an increase in level, although results were more pronounced for Mr. Johnson and Ms. Garcia. Mr. Johnson also demonstrated an increasing trend in this phase. These preliminary data suggest that a receiving brief training and self-monitoring plan implementation appeared to be more effective than self-assessing alone. Teachers did not report during or after the study exactly when and how they self-monitored their implementation. Teachers' data sheets were merely collected weekly.

During the Self-Monitoring + Performance Feedback Phase, Mr. Johnson and Ms. Garcia's implementation remained relatively high and stable, while Ms. Watson demonstrated an increasing trend. Tau-U effect size estimates were high for Mr. Johnson from Implementation Baseline to the Self-Monitoring Phase and Implementation Baseline to the Self-Monitoring + Performance Feedback Phase. For Ms. Watson and Garcia, Tau-U was larger for Implementation Baseline to the Self-Monitoring + Performance Feedback Phase. Overall, teachers appeared to implement more components of the plan when receiving performance feedback in addition to self-monitoring their implementation. Performance feedback emails included visual graphs of implementation that may have impacted teachers' behavior. Emails were sent before class and

may have also served as a prompt to provide specific steps of the intervention, increasing the number of steps implemented overall in the phase.

Although performance feedback seemed to make a difference to teachers' implementation, a large difference in student outcomes was not noted from the Self-Monitoring Phase to the Self-Monitoring + Performance Feedback Phase. Average academic engagement increased slightly for Mr. Johnson and Ms. Garcia, but decreased for Ms. Garcia (see Table 2). Disruptive behavior was slightly higher when Mr. Johnson received performance feedback, but declined for Ms. Watson and Ms. Garcia during the Self-Monitoring + Performance Feedback condition. Tau-U estimates indicate there was a larger change in academic engagement from Implementation Baseline to Self-Monitoring + Performance Feedback condition for Mr. Johnson, but small to no effect was noted for differences between other phases across participants.

Although an increase in academic engagement data was only noted during the Self-Monitoring + Performance Feedback Phase for two teachers (Mr. Johnson and Ms. Watson), evaluation of the effort involved in sending performance feedback is important and justified. Although Ms. Garcia's class did not have an increase in academic engagement, an overall decrease in disruptive behavior was observed during the Self-Monitoring + Performance Feedback condition. Future research should replicate this study and expand these phases to determine if a clearer effect for academic engagement and disruptive behavior is observed over time.

Limitations

Several limitations are important to consider when evaluating the results from this study. First, teachers volunteered after a meeting with the study's first author and the school principal, potentially introducing a selection bias (i.e., teachers who volunteered might be more receptive

to training than other teachers). Furthermore, although all teachers in the current study had a master's degree (which is aligned with national teacher trends, n = 56%; NCES, 2012), participant characteristics did not necessarily reflect the demographic of national teacher trends otherwise. Rather than White female teacher participants, the three teachers who volunteered included a male and two women of color, possibly impacting self-assessment responses and subsequent implementation of suggested practices. Also, as this was a single case design study, only three teachers were needed to demonstrate three replications of a treatment effect across participants (Kratochwill et al., 2010). Yet, future research might seek to replicate the method with a larger sample of teachers with diverse characteristics to increase the generalizability of findings.

Also, ceiling and floor effects were evident when tracking academic engagement and disruptive behavior across phases. As such, student outcome dependent variables may not be as sensitive to change as other possible variables may have been given the current research questions. However, lack of substantive change in rates of academic engagement and disruptive behavior across classrooms throughout the study also indicates that implementing the classwide plan did not appear to adversely impact student outcome data. If in future studies academic engagement and disruptive behavior are tracked as dependent variables, researchers might screen classrooms initially to ensure high levels of disruptive behavior and low levels of academic engaged behavior are present to more adequately warrant classwide intervention.

Furthermore, teachers implementing the universal plan were making an effort to engage in more culturally and contextually relevant practice. Rather than measuring academic engagement and disruptive behavior, another dependent variable may be more responsive to teachers' change in practice (e.g., perception of connectedness to teacher or lesson). Future

research may explore this possibility and identify alternative dependent variables that seem to be more impacted by teachers' treatment fidelity to these types of classwide plans. In addition, although the justification for the current study includes reported trends related to disproportionality in school discipline, we did not directly measure variables related to race or disproportionality in the current study, which should be explored in subsequent research. Finally, breaks and scheduling interruptions limited data collection during certain phases, affecting adherence to single case design standards (Kratochwill et al., 2010).

Conclusion

Findings from the current preliminary study indicate engaging in self-assessment and subsequent training on a comprehensive classwide plan seemed to be associated with increases in teachers' use of culturally and contextually relevant strategies in the classroom. Teachers implemented the classwide plan relatively consistently and comprehensively while self-monitoring implementation; however, the introduction of performance feedback increased slightly teachers' implementation of the plan. Results related to the relationship between teachers' implementation and students' behavior were less clear. Overall, some evidence suggested a relationship between increased academic engagement and decreased disruptive behavior upon teachers' implementation of the classwide plan, but future research should further explore this relationship or consider an alternative student dependent variable to monitor over time.

Educators needing assistance to support the behavior of all students in his/her class, particularly CLD learners, might consider engaging in self-assessment and seeking technical support to change his/her classroom practice. Teachers might seek this assistance from support staff in the school (e.g., school psychologist, lead teacher, behavior coach) and self-monitor

progress toward changes implemented over time. Collecting student data may also assist teachers in determining if changes in classroom practice are having the desired effect. Future research should evaluate if these procedures can be implemented effectively by a wider range of teachers.

References

Algozzine, B., Putnam, R., & Horner, R. H. (2012). Support for teaching students with learning disabilities academic skills and social behaviors within a response-to- intervention model: Why it doesn't matter what comes first. *Insights on Learning Disabilities, 9,* 7-36.

Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavioral Interventions, 13,* 3-16.

Bal, A., Thorius, K. A. K., & Kozleski, E. B. (2012). *Culturally responsive positive behavioral support matters.* Tempe, AZ: The Equity Alliance. Retrieved from http://www.equityallianceatasu.org/sites/default/files/CRPBIS_Matters.pdf

Bohanon, H., Fenning, P., Carney, K. L., Minnis-Kim, M. J., Anderson-Harriss, S., Moroz, K. B., & … Pigott, T. D. (2006). Schoolwide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions*, *8*, 131-145.

Bottiani, J. H., Bradshaw, C. P., & Mendelson, T. (2017). A multilevel examination of racial disparities in high school discipline: Black and white adolescents' perceived equity, school belonging, and adjustment problems. *Journal of Educational Psychology*, 109, 532-545.

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, *51*, 81-96.

Briesch, A. M., Hemphill, E. M., Volpe, R. J., & Daniels, B. (2015). An evaluation of observational methods for measuring response to classwide intervention. *School Psychology Quarterly*, *30*, 37-49.

Browne-Dianis, J. (2011). Stepping back from zero tolerance. *Educational Leadership, 69,* 24-28.

Cartledge, G., Singh, A., & Gibson, L. (2008). Practical behavior-management techniques to close the accessibility gap for students who are culturally and linguistically diverse. *Preventing School Failure, 52,* 29–38.

Chafouleas, S. M., Sanetti, L. M. H., Kilgus, S. P., & Maggin, D. M. (2012). Evaluating sensitivity to behavioral change using Direct Behavior Rating Single-Item Scales. *Exceptional Children, 78,* 491-505.

Chesley, G. M., & Jordan, J. (2012). What's Missing from Teacher Prep. *Educational Leadership*, *69*, 41-45.

Collier-Meek, M. A., Fallon, L. M., Sanetti, L. M. H., & Maggin, D. M. (2013). Focus on implementation: Strategies for problem-solving teams to assess and promote treatment fidelity. *Teaching Exceptional Children, 45,* 52-59.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children*, *79*, 135–144.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed). Upper Saddle River, NJ: Prentice Hall.

Debnam, K. J., Pas, E. T., Bottiani, J., Cash, A. H., & Bradshaw, C. P. (2015). An examination of the association between observed and self- reported culturally proficient teaching practices. *Psychology in the Schools*, *52*, 533-548.

Delpit, L. (2006). Other people's children: Cultural conflict in the classroom (2nd ed.). New York: New Press.

Delpit, L. (2012). *"Multiplication is for White People" Raising Expectations for Other People's Children.* New York: New Press.

Eagle, J. W., Dowd-Eagle, S. E., Snyder, A., & Holtzman, E. G. (2015). Implementing a multi-tiered system of support (MTSS): Collaboration between school psychologists and administrators to promote systems-level change. *Journal of Educational and Psychological Consultation*, *25*, 160-177.

Fallon, L. M., O'Keeffe, B. V., Gage, N., & Sugai, G. (2015). Brief report: Assessing attitudes toward culturally and contextually relevant school-wide positive behavior support strategies. *Behavioral Disorders 40,* 251-260.

Fallon, L. M., O'Keeffe, B. V., & Sugai, G. (2012). Consideration of culture and context in School-wide Positive Behavior Support: A review of current literature. *Journal of Positive Behavior Interventions, 14,* 209-219.

Ferguson, T. D., Briesch, A. M., Volpe, R. J., & Daniels, B. (2012). The influence of observation length on the dependability of data. *School Psychology Quarterly*, *27*, 187-197.

Fenning, P., & Rose, J. (2007). Overrepresentation of African American students in exclusionary discipline the role of school policy. *Urban Education*, *42*, 536-559.

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, *37*, 106-120.

Green, T. D. (2005). Promising prevention and early intervention strategies to reduce

overrepresentation of African American students in special education. *Preventing School Failure, 49,* 33–41.

Goldring, R., Taie, S., and Riddles, M. (2014). *Teacher attrition and mobility: Results from the 2012–13 teacher follow-up survey* (NCES 2014-077). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Gregory, A., Skiba, R. J., & Mediratta, K., (2017). Eliminating disparities in school discipline: A framework for intervention. Special issue on equity in school. *Review of Research in Education, 41,* 253–278. DOI: 10.3102/0091732X17690499

Guyton, E., & Wesche, M. (2005). The multicultural efficacy scale: Development, item selection, and reliability. *Multicultural Perspectives, 7*(4), 21–29. doi:10.1207/s15327892mcp0704_4

Hershfeldt, P. A., Sechrest, R., Pell, K. L., Rosenberg, M. S., Bradshaw, C. P., & Leaf, P. J. (2009). Double-Check: A framework of cultural responsiveness applied to classroom behavior. *Teaching Exceptional Children Plus, 6 (2),* 2-18. Retrieved from http://web.ebscohost.com

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Irvine, J. J. (1990). *Black students and school failure: Policies, practices, and prescriptions.* New York: Greenwood Press.

Kaufman, J. S., Jaser, S. S., Vaughan, E. L., Reynolds, J. S., Di Donato, J., Bernard, S. N., & Hernandez-Brereton, M. (2010). Patterns in office referral data by grade, race/ethnicity,

and gender. *Journal of Positive Behavior Interventions, 12,* 44-54.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lane, K. L., Oakes, W. P., Carter, E. W., & Messenger, M. (2015). Examining behavioral risk and academic performance for students transitioning from elementary to middle school. *Journal of Positive Behavior Interventions*, *17*, 39-49.

Marchant, M., & Anderson, D. H. (2012). Improving social and academic outcomes for all learners through the use of teacher praise. *Beyond Behavior*, *21*(3), 22-28.

McIntosh, K., Moniz, C., Craft, C. B., Golby, R., & Steinwand-Deschambeault, T. (2014). Implementing school-wide positive behavioural interventions and supports to better meet the needs of indigenous students. *Canadian Journal of School Psychology, 29*, 236-257.

McCurdy, B. L., Mannella, M. C., & Eldridge, N. (2003). Positive behavior support in urban schools: Can we prevent the escalation of antisocial behavior? *Journal of Positive Behavioral Interventions, 5,* 158-170.

Monroe, C. R. (2005). Why are" bad boys" always black?: Causes of disproportionality in school discipline and recommendations for change. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *79*, 45-50.

Noltemeyer, A. L., & Mcloughlin, C. S. (2010). Changes in Exclusionary Discipline Rates and Disciplinary Disproportionality over Time. *International Journal of Special Education*, *25*, 59-70.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S.B. (2010). Combining non-overlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42,* 284–299.

Sanetti, L. M. H., Chafouleas, S. M., Fallon, L. M., & Jaffery, R. (2014). Increasing teachers' treatment integrity when implementing a class-wide intervention through performance feedback provided by a school-based consultant: A case study. *Journal of Educational and Psychological Consultation, 24,* 239-260.

Sanetti, L. M. H., & Collier-Meek, M. A. (2015). Data-driven delivery of implementation supports in a multi-tiered framework: A pilot study. *Psychology in the Schools*, *52*, 815-828.

Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C. J., Byron, J. R., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of School Psychology, 53,* 209-229.

Sanetti L. M. H., Kratochwill T. R. (2009). Toward developing a science of treatment integrity: Introduction to a special series. *School Psychology Review, 38,* 445–459.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). A review of evidence based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children, 31,* 351-380.

Simonsen, B., MacSuga, A. S., Fallon, L. M., & Sugai, G. (2013). Teacher self-monitoring to increase specific praise rates. *Journal of Positive Behavior Interventions, 15,* 5-15.

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education, 23,* 1086–1101. doi:10.1016/j.tate.2006.07.011

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*, 85-107.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, *34*, 317-342. DOI: 0042-0972/02/1200-0317/0

Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education and Treatment of Children*, 295-315.

Sugai, G., O'Keeffe, B. V., & Fallon, L. M. (2012). A contextual consideration of culture and school-wide positive behavior support. *Journal of Positive Behavior Interventions, 14,* 197- 208.

Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. L. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *Journal of Special Education, 41,* 223-233.

U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD). (2017). [Graph illustration the Percentage distribution of students enrolled in public elementary and secondary schools, by race/ethnicity: Fall 2004, fall 2014, and fall 2026, table 203.50]. *Digest of Education Statistics*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cge.asp

U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS). (2017). [Graph illustration the Total number of public school teachers and percentage distribution of school teachers, by race/ethnicity and state: 2011–12, table]. *National Center for Education Statistics*. Retrieved from https://nces.ed.gov/surveys/sass/tables/sass1112_2013314_t1s_001.asp

Vincent, C. G., Swain-Bradway, J., Tobin, T. J., & May, S. (2011). Disciplinary referrals for culturally and linguistically diverse students with and without disabilities: Patterns resulting from school-wide positive behavior support. *Exceptionality,* 19, 175–190.

Wallace Jr., J. M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among US high school students: 1991-2005. *The Negro Educational Review*, *59*, 47-62.

Table 1

Teachers' Implementation Across Phases

	Baseline	Intervention		Effect Size Estimates	
	Implementation Baseline (IB)	Self-monitoring (SM)	SM + Performance Feedback (PF)	ES IB to SM	ES IB to PF
Treatment Adherence					
Mr. Johnson	31.6 (4.0)	61.4 (17.5)	74.0 (6.5)	1.00	1.00
Ms. Watson	45.2 (8.7)	50.9 (18.6)	62.7 (18.1)	0.34	0.74
Ms. Garcia	35.6 (18.9)	68.6 (7.0)	79.0 (13.5)	0.83	0.91

Table 2

Student Outcomes Across Phases

	Baseline		Intervention		Tau-U Estimates		
	Student Baseline (SB)	Implement. Baseline (IB)	Self-monitoring (SM)	SM + Performance Feedback (PF)	Tau-U IB to SM	Tau-U IB to PF	Tau-U SB + IB to SM + PF
Academic Engagement							
Mr. Johnson	81.4 (12.8)	86.5 (4.8)	92.2 (1.4)	94.1(1.8)	0.38	0.51	0.88
Ms. Watson	83.8 (10.1)	93.2 (4.1)	96.1 (1.7)	97.6 (1.8)	0.05	0.08	0.65
Ms. Garcia	77.8 (12.0)	89.2 (4.0)	91.9 (2.8)	89.8 (6.5)	0.17	0.08	0.35
Disruptive Behavior							
Mr. Johnson	2.2 (0.9)	2.5 (1.5)	3.4 (2.0)	3.5 (2.3)	0.13	0.07	0.12
Ms. Watson	4.0 (0.4)	3.6 (3.1)	3.4 (1.4)	2.3 (1.8)	0.33	0.14	0.14
Ms. Garcia	10.7 (3.8)	6.5 (4.7)	9.2 (3.3)	3.8 (1.9)	0.44	0.02	0.16