A Randomized Controlled Trial of MTSS-B in High Schools: Improving Classroom Management to Prevent EBDs

Catherine P. Bradshaw, PhD[1,2] iD, Elise T. Pas, PhD[2],
Katrina J. Debnam, PhD[1], and Sarah Lindstrom Johnson, PhD[3]

Abstract

This study presents findings from a 58 high-school group-randomized controlled trial testing the effectiveness of training in a multitiered system of supports for behavior (MTSS-B) framework, which was leveraged to reduce students' risk for emotional and behavior disorders. The trial tested the impact of MTSS-B, which included (a) training in the broader MTSS-B framework that went beyond the existing Tier 1 (school-wide PBIS) training offered by the state; (b) project-provided coaching and technical assistance supports; and (c) integration and training in evidence-based behavioral or social-emotional programs at Tiers 2 and/or 3. We reported effects of MTSS-B on implementation of positive behavior supports across all three tiers using the Schoolwide Evaluation Tool (SET) and Individual Student Systems Evaluation Tool (ISSET), as well as on external observations of teachers' use of classroom management strategies. Results indicated significant effects on multiple SET subscales and significant reductions in teachers' use of reactive behavior management.

Although promoting a safe and supportive school environment has been a high priority in U.S. schools, many high schools still struggle to reduce risk factors and prevent the harms associated with substance use, violence, bullying, and related mental health problems (Musu et al., 2019; U.S. Department of Education, 2018). Of particular concern are supports for students with emotional and behavioral disorders (EBDs) and those at risk for developing these problems, given the high prevalence of social, emotional, and mental health issues among school-aged youth and adolescents (Perou et al., 2013). Not only can prevention efforts reduce the potential onset of EBDs, they can also work to ameliorate the symptoms and prognosis for students with EBDs, which can be exacerbated in settings challenged by poor classroom management and school climate concerns (Lloyd et al., 2019).

Toward that end, there is an increasing movement in the special education field toward the multitiered system of supports (MTSS) framework for providing universal, targeted, and intensive evidence-based supports to prevent the onset or escalation of academic and behavioral issues (Bradshaw et al., 2019). Although MTSS-B has been examined in elementary schools, there has been limited research in high schools using a rigorous randomized design, particularly

with consideration of impacts on the classroom setting (e.g., Flannery et al., 2014; Horner et al., 2010; Lee & Gage, 2020). The need for MTSS-B research at the high school level is especially great, given elevated rates of bullying and suspension, and the long-term consequences of school dropout (Musu et al., 2019). Moreover, there are relatively few evidence-based prevention programs for high-school students at risk for EBD (Lloyd et al., 2019).

The current study aimed to address these gaps by testing the extent to which training and coaching in MTSS-B resulted in the implementation of a continuum of evidence-based social and behavioral programs and practices (EBPs). We were particularly interested in the impact on classrooms, as the theory of change associated with MTSS-B suggests that school-wide programming should result in observable improvements in teacher classroom management and in

[1]University of Virginia, Charlottesville, USA
[2]Johns Hopkins University, Baltimore, MD, USA
[3]Arizona State University, Tempe, USA

Corresponding Author:
Catherine P. Bradshaw, School of Education and Human Development, University of Virginia, Bavaro Hall, 112-D, 417 Emmet Street South, P.O. Box 400260, Charlottesville, VA 22904, USA.
Email: cbradsha@jhsph.edu

turn ensure that the relatively limited Tier 2 and 3 supports can be allocated to students at risk for EBD. Together, these findings have the potential to inform future scale-up of EBPs in high schools, particularly for students at risk for EBD. This line of research is especially relevant to the special education field, given the significant investment in federal funding for the scale-up of Positive Behavioral Interventions and Supports (PBIS; www.pbis.org) through the Office of Special Education Programs' National Technical Assistance Center.

Background on the Maryland Safe and Supportive Schools (MDS3) Project

Maryland was considered a "national exemplar" for the successful implementation of the Tier 1 (school-wide) elements of the MTSS-B framework (Sugai & Horner, 2006). At the time the trial was conducted, PBIS was the term commonly used within the state as the specific framework through which schools in Maryland implemented the universal or school-wide elements of MTSS-B (Bradshaw et al., 2012, 2014). Many of the successes associated with PBIS implementation in Maryland had occurred at the elementary or middle school level and had focused on universal (school-wide) supports, rather than more intensive targeted and indicated interventions (Bradshaw et al., 2012, 2014). Similarly, the most rigorous research supporting the impacts of the Tier 1 behavioral model had been with regard to the school-wide supports in elementary schools specifically (e.g., Bradshaw et al., 2010; Bradshaw, Koth, et al., 2008, 2009; Horner et al., 2009), although more recent research demonstrates positive impacts in secondary schools (e.g., Pas, Ryoo et al., 2019; also see a review by Lee & Gage, 2020). As a result, there was a desire to expand both into high schools and beyond the Tier 1 supports, providing support for the implementation Tiers 2 and 3 (Bradshaw et al., 2014).

Maryland leveraged the U.S. Department of Education's Safe and Supportive Schools opportunity to scale-up the Tier 1 PBIS supports and related EBPs in high schools using the full, three-tiered framework, and rigorously tested the effects in a school-level randomized controlled effectiveness trial (RCT; see Bradshaw et al., 2014). As part of the MDS3 Project, the research team worked with the Maryland State Department of Education (MSDE) to develop the web-based MDS3 School Climate Survey System, which included an integrated set of self-reported school climate indicators (i.e., with versions for students, staff, and parents) covering the broad constructs of safety, engagement, and environment (Bradshaw et al., 2014) and live reporting system; school climate data augmented other data (e.g., office disciplinary referrals, suspensions, fidelity) typically collected and used by school schools to inform decisions and implementation. The MDS3 School Climate Survey was administered annually across the 4 years of the study in all 58 comparison and intervention schools. Project-hired MDS3 coaches received

training in a set of EBPs and delivered training and coaching to randomized intervention schools, to support the implementation of the multitiered framework and a menu of EBPs across Tiers 1, 2, and 3. In contrast, the comparison schools received the regular state-provided Tier 1 school-wide PBIS training and were able to utilize data from the MDS3 School Climate Survey system to inform the implementation of various practices, consistent with the state's PBIS implementation framework. These schools did not receive training on the EBPs or how to apply school climate data to decision-making; nor did these receive MDs3 coaching supports across the advanced tiers.

The menu of EBPs offered to intervention schools included the Olweus Bullying Prevention Program (Olweus et al., 2007; Tier 1 universal bullying curriculum), LifeSkills Training for High Schools (Botvin et al., 2006; Tier 1 drug prevention curriculum), Check-In/Check-Out (CI/CO; Hawken & Horner, 2003; Tier 2 behavior and engagement intervention), Check & Connect (Sinclair et al., 2005; Tiers 2–3 student mentoring and engagement intervention), and Cognitive-Behavioral Intervention for Trauma in Schools (CBITS; Stein et al., 2003; Tiers 2–3 intervention). At the time of the study, all of the EBPs offered were listed on one or more prevention program clearinghouses (e.g., Blueprints for Violence Prevention, What Works Clearinghouse) and selected for inclusion on the menu of EBPs based on feedback and priorities set by the state (Bradshaw et al., 2014). Schools selecting one or more of these EBPs received training from a certified/qualified trainer, along with implementation supports from an MDS3 coach. For additional information on the EBPs and MDS3 model, see Bradshaw et al. (2014).

Overview of the Current Study

The primary purpose of this study was to examine the effects of training and ongoing coaching support in the MTSS-B model on implementation of school-wide programming and classroom-based practices over the course of the 3-year RCT. Consistent with the more recent shift in terminology to MTSS-B by the field and by the U.S. Department of Education, we refer to the model tested in the RCT as MTSS-B, which encompasses the school-wide PBIS framework and integrated EBPs to address behavioral and social-emotional needs. The intervention included the following core elements: (a) training in the broader, three-tiered MTSS-B framework beyond the existing Tier 1 (PBIS) training offered by the state; (b) project-provided coaching and technical assistance focused on the use of the MDS3 school climate data to inform decision-making; and (c) integration and training in evidence-based behavioral or social-emotional programs at Tiers 2 and/or 3. In contrast, the comparison schools only received training in Tier 1 positive behavior supports from the state, in a business as usual condition, and had access to the MDS3 school climate data.

Table 1. School Demographics.

Demographic variables	Intervention condition (n = 31)		Comparison condition (n = 27)		Full sample of schools (N = 58)	
	M	SD	M	SD	M	SD
School characteristics						
School enrollment	1,331.6	488.9	1,226.8	445.1	1,282.8	467.9
School attendance (%)	92.8	2.0	93.1	1.5	92.9	1.8
Student mobility (%)	18.6	10.9	18.0	8.6	18.3	9.8
Free/reduced meals (%)	35.3	16.4	33.3	16.7	34.4	16.4
Special education (%)	5.1	0.3	7.8	4.0	6.3	2.9
Minority (%)	45.7	25.7	44.5	25.3	45.2	25.3
Suspension rate (%)	28.6	17.5	26.9	14.3	27.8	16.0
HSA: Algebra passing rate	87.8	8.1	90.0	7.3	88.8	7.8
HSA: Biology passing rate	85.6	8.1	83.4	9.6	84.6	8.8
HSA: English passing rate	85.5	7.5	86.6	6.9	86.0	7.2
Teacher characteristics						
Student–teacher ratio	19.8	3.0	19.8	3.4	19.8	3.1
Standard or advanced certification (%)	94.3	4.4	95.3	3.5	94.8	4.0

Note. T tests indicated no significant differences on any of these indicators across condition (at $p < .05$). HSA = high school assessment (i.e., state standardized academic test).

The first aim leveraged the full sample of 58 schools to examine the implementation of the core components of Tier 1 school-wide positive behavior supports, as indicated by the Schoolwide Evaluation Tool (SET; Sugai et al., 2001). We also examined the implementation of the more intensive EBPs, measured via the Individual Student Systems Evaluation Tool (ISSET; Lewis-Palmer et al., 2005). We were particularly interested in determining whether scores on the SET and ISSET subscales and overall scales improved over time for intervention schools, relative to the comparison schools, indicating that the MTSS-B training and coaching impacted the quality of implementation of the Tier 1, 2, and 3 supports, over and above the training by the state to the comparison schools. Collecting these data in both conditions also served as a contamination check or treatment contrast in the comparison condition. To further triangulate these implementation findings, we also explored the level of implementation of specific EBPs offered through the MTSS-B framework among the 31 intervention schools (only) to examine uptake and implementation levels over the course of the study.

The second aim of this study was to examine the proximal impacts of the MTSS-B intervention on observed teacher behavioral management practices in the classrooms to determine whether there was growth in proactive and positive practices and declines in reactive and negative practices for MTSS-B intervention schools, in contrast to the comparison schools. Importantly, the teachers' classroom management practices were measured via classroom observations conducted by trained observers who were unaware of the schools' intervention condition. The third aim examined how baseline

implementation, measured by the SET and ISSET, related to changes in teacher behaviors over time. All three of these primary aims were realized using hierarchical linear modeling (HLM). In summary, this effectiveness study of MTSS-B training and coaching is particularly novel, as it focuses on high schools and incorporates external observational data collected at both the classroom and school levels to examine the extent to which training and coaching supports resulted in observable changes in the classroom context and uptake of the EBPs and school-wide practices.

Method

Participants

The MDS3 Project included 58 high schools (traditional/comprehensive; Grades 9–12) across 12 Maryland school districts. The participating schools had a diverse student population with a student racial/ethnic minority composition of 45.18% and a mean student enrollment of 1,282.81 ($SD = 467.93$). Approximately 34.37% ($SD = 16.43\%$) of students in these schools received free and reduced-price meals (FARMs) and 9.60% received special education services ($SD = 3.10\%$). The student–teacher ratios were approximately 20 students to teachers ($M = 19.81$, $SD = 3.14$) and most teachers had a standard or advanced teaching certification ($M = 94.78\%$, $SD = 23.99$). A series of t tests on the baseline school demographic, school climate, and observational data indicated no significant differences (at $p < .05$) across condition, indicating baseline equivalence (see Table 1 for descriptives by intervention condition).

Procedures

Recruitment. MSDE leadership led the recruitment and enrollment of 12 school systems and the 58 high schools through formal presentations to local superintendents and their staff regarding the MDS3 Project. Districts were selected by MSDE based on need, readiness, and willingness to participate. All districts approached about participation consented. District and school participation in the trial was voluntary, and all district and school-based administrators provided written consent for participation and were informed verbally and in writing of the RCT design and all data collection procedures. As this project did not collect individual student- or staff-level identifiers, it was deemed exempt by the researchers' institutional review board.

Study design. The study employed a group (school-level) RCT design (Gottfredson et al., 2015; Murray, 1998). The university-based research team conducted random assignment, whereby 31 of the 58 high schools were randomly assigned to receive training and ongoing coaching support in the MTSS-B model and use of the school climate data to determine the need for EBPs at Tiers 1 and 2 and/or 3. The remaining 27 schools served as comparisons with access to Tier 1 PBIS training from the state and access to the school climate survey system (i.e., business as usual). See Bradshaw and Pas (2011) for additional details about the standard model for Tier 1 PBIS training by the state, which was available to the schools in the comparison condition. Schools were assigned into paired matches based on their district- and school-level demographic characteristics and baseline data (e.g., school climate, enrollment, suspension, academic proficiency); one school in each pair was randomly assigned to each group, to ensure balance across conditions. A slightly higher proportion of schools was assigned to the intervention condition to increase the power for implementation analyses, as is the focus here. The random assignment was performed by a statistician engaged in the project only for this purpose. Consistent with the effectiveness design (Gottfredson et al., 2015), overall project management was led by the MSDE, implementation was led by Sheppard Pratt Health System (SPHS), and randomization and data collection activities were led by the university-based research team (see Bradshaw et al., 2014).

Over the summer following the end of the RCT, the comparison schools received access to project materials and were offered a training; however, no additional data collection efforts, coaching, or systematic supports were provided to schools in either condition. As such, the effectiveness study was designed to be a traditional cluster/school-level RCT (Murray, 1998); there was no waitlist or follow-up at the end of the 3 years in the RCT. The value-added design testing the benefits of the training in the MTSS-B framework, over and above the Tier 1 PBIS framework provided by the state, precludes causal interpretations of changes that might occur for schools assigned to the comparison condition, as there is no true control condition. There was no attrition of schools from the RCT. Although there was some turnover in school staff and leadership, we did not track individual data (or turnover) within the school buildings.

Training and implementation. To support the intervention schools, master- or doctoral-level coaches were hired and trained by SPHS to serve as MDS3 project coaches, following the PBIS National Technical Assistance Center's Model. Coaching was provided to intervention schools by 12 MDS3 coaches, of which eight were female and nine had a master's degree (or higher). All MDS3 coaches either had previously been teachers (i.e., nine) and/or were mental health providers (i.e., one school psychologist, two social workers, three school counselors). MDS3 coaches were assigned to approximately three to five schools each and provided support to the schools across all 3 years post randomization. Through their assigned coach, the MTSS-B intervention schools received individually tailored training and coaching regarding data-based decision-making and how to identify and allocate the necessary resources to implement an integrated continuum (e.g., Tiers 1, 2, and 3) of EBPs. The MDS3 coaches (a) led trainings in the MTSS-B framework, including the teaming process and use of data to inform decision-making; (b) provided ongoing coaching in the implementation of the core Tier 1 school-wide PBIS foundational features, integrating in school climate data as a new data element for data-based decision-making, before moving into a focus on Tier 2 and 3; and (c) provided training and/or coordinated trainings in the Tier 1, 2, and 3 EBPs (i.e., with certified trainers if they were not certified in that EBP). To support integration of the project-provided supports with other district efforts, the MDS3 coaches regularly attended school district meetings and trainings; these formal and informal connections were intended to help increase district-level integration and sustainability of the supports provided in the intervention schools. The MDS3 Project provided the necessary resources (e.g., training, materials, ongoing coaching) to implement one or more of the EBPs in the intervention schools.

While in the schools, the coaches engaged in a range of activities, including coaching schools on the review of fidelity and outcomes data, data-based decision-making, teaming, and engaging in targeted practices. They also provided training and support specific to the EBPs (for additional details, see Bradshaw et al., 2014; Pas et al., 2020). Coaches tracked their school contacts using an electronic coaching log after each school visit, by recording details about the services provided to the school (e.g., total time, activity types, specific EBP supported). Briefly, those coaching log data indicated that, on average, schools received a total of 248.39 ($SD = 138.22$) hr spread across the

3 years. Coaches spent, on average, 82.06 ($SD = 49.02$) hr in each school in Year 1, 85.37 ($SD = 50.78$) hr in Year 2, and 80.96 ($SD = 59.77$) hr in Year 3.

Data collection procedures. Intervention and comparison schools were monitored over the same 3-year period. Data were collected at baseline (i.e., fall of the first year of the study), and then in the spring of each of the subsequent three interventions years (i.e., four waves of data collections in total). We collected data on the implementation of the MTSS-B framework at Tiers 1 to 3 in schools in both conditions, to detect the potential for contamination and examine the treatment contrast; this is particularly important for the PBIS/MTSS-B model, given schools could have accessed program elements independent of formal training through this project (Bradshaw et al., 2008).

Measures

SET. The SET (Sugai et al., 2001) was the most commonly used PBIS implementation measure at the time of this study and has the longest history of use and research. It assesses seven core components of the universal, school-wide components of PBIS. For this study, the SET was completed by an external evaluator hired and trained by the research team who conducted brief interviews, toured the school, and reviewed materials during one day to rate the extent to which each of the 58 items and subcomponents of Tier 1 PBIS were in place (i.e., not implemented = 0, partially implemented = 1, fully implemented = 2). These 58 items comprised seven scales on the SET: A: Expectations Defined (Cronbach's αs for the current sample; $\alpha = .81$), B: Behavioral Expectations Taught ($\alpha = .85$), C: System for Rewarding Behavioral Expectations ($\alpha = .82$), D: System for Responding to Behavioral Violations ($\alpha = .43$), E: Monitoring and Evaluation ($\alpha = .60$), F: Management ($\alpha = .91$), and G: District-Level Support ($\alpha = .65$; see Horner et al., 2004, for additional details on the SET). Each scale score was calculated by dividing the number of earned points on scale items by possible total points and multiplying by 100. The scores range from 0% to 100%, with higher scores indicating greater implementation fidelity. An overall summary score was then computed by averaging all seven scale scores (i.e., overall SET score), which also ranges from 0% to 100% ($\alpha = .93$). Baseline data were used to calculate all alphas reported here. See Bradshaw, Reinke, et al. (2008) and Pas, Johnson, et al. (2019) for greater detail. Descriptive SET data are provided in Table 2.

ISSET. The ISSET (Lewis-Palmer et al., 2005) was administered simultaneously with the SET by the external observer. The ISSET utilized brief interviews and review of materials developed and used for intervention planning and implementation of EBPs (Debnam et al., 2012; Lewis-Palmer et al., 2005). The ISSET includes 46 items organized into four subscales: (a) *Schoolwide Interventions* includes questions about universal interventions targeting specific areas (e.g., bullying), the link between offered interventions and school-wide behavior expectations, and the data-based decision-making process (14 items; α for the current sample = .95); (b) *Foundations* includes items about the referral procedures for additional supports (11 items, $\alpha = .69$); (c) *Targeted Interventions* asks about three specific interventions being offered at Tiers 2 and 3 and the process for identifying students eligible to receive each intervention and implementing those interventions (13 items, $\alpha = .93$); and (d) *Intensive Individualized Interventions* includes elements about the school's functional behavioral assessment process and staff qualifications (eight items, $\alpha = .60$). Again, items are scored on a 3-point scale (i.e., 0–2), and a percent for all scales is calculated. An overall ISSET score is created by averaging the four subscale scores ($\alpha = .92$). All alphas reported here are baseline data. Descriptive ISSET data are provided in Table 2.

SET/ISSET assessors were unaware of the schools' intervention status within the RCT; they also were unaware whether the schools were receiving ongoing training and coaching supports, and moreover had minimal information about the purpose of the MDS3 Project, or details on MTSS-B or the EBPs. Assessors were trained in administering the SET and ISSET tools specifically for research purposes. Thus, their focus was on using the assessment guide to gather data rather than provide any technical assistance or consultation. SET and ISSET subscale scores from the three spring data collections during active intervention years were utilized as outcomes of interest; baseline overall scores were also included as predictors of teacher classroom behavioral management practices. For additional information on these measures, including psychometrics properties and training, see Debnam et al. (2012) and Pas, Johnson, et al. (2019).

Assessing School Settings: Interactions of Students and Teachers (ASSIST). The ASSIST (Rusby et al., 2001) is an observational measure that includes event-based tallies (i.e., counts of specific behaviors) of teacher classroom management strategies. Prior research has documented the reliability and validity of the ASSIST in high schools (see Pas et al., 2015). Data were collected at each school in 25 classrooms over 3 days by trained observers. Observers followed a written protocol for identifying classrooms, including that they first observed *all* language arts teachers and then randomly selected teachers, blocking on core subject area (i.e., math, science, social studies) from the school schedule to reach the needed 25 teachers. They ensured that teachers were only observed once. Observers did not have any knowledge about the intervention, school, or teachers to bias their selections. Data were entered on a Samsung tablet using the Pendragon mobile data collection software.

Table 2. SET and ISSET Scores Across the 3-Year RCT.

| | Baseline | | | | Year 1 | | | | Year 2 | | | | Year 3 | | | |
| | Intervention | | Comparison | | Intervention | | Comparison | | Intervention | | Comparison | | Intervention | | Comparison | |
Scales	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
SET Subscales																
A: Expectations Defined	46.77	41.70	46.30	44.78	72.58	37.28	59.26	43.93	82.26	27.53	68.52	37.72	92.74	20.61	60.19	42.89
B: Expectations Taught	47.74	35.19	42.59	35.04	62.90	31.33	53.70	39.04	78.06	24.42	57.41	31.45	81.61	15.94	64.44	29.00
C: System for Rewarding Behavioral Expectations	33.87	33.19	35.19	38.49	46.24	35.15	40.74	36.20	61.83	35.80	48.15	32.14	63.98	30.15	54.32	29.81
D: System for Responding to Behavioral Violations	81.61	16.55	82.59	12.59	90.32	11.69	90.37	10.91	93.23	7.91	91.85	7.36	92.90	10.39	94.81	9.75
E: Monitoring and Evaluation	83.47	20.00	68.98	25.09	82.66	23.20	75.46	23.89	91.94	14.63	80.56	26.48	90.32	16.99	76.85	28.10
F: Management	69.35	32.13	56.71	38.09	70.36	33.46	58.80	38.03	86.29	18.64	62.04	40.17	81.45	26.39	69.21	37.22
G: District-Level Support	61.29	42.25	50.00	43.85	69.35	40.16	55.56	42.37	83.87	29.96	64.81	38.77	83.87	32.63	62.96	42.95
SET Overall Score	60.59	23.70	54.62	26.70	70.63	23.24	61.98	28.00	82.50	16.21	67.62	25.91	83.84	14.55	68.97	25.24
ISSET Subscales																
ISSET: Foundations	68.50	15.47	63.13	22.10	71.11	11.94	69.70	16.44	76.25	18.06	66.67	22.27	78.59	14.71	77.78	13.24
ISSET: Targeted	37.47	31.75	24.34	25.11	61.29	32.21	72.36	30.15	79.53	25.52	63.25	31.52	79.78	26.72	77.92	22.97
ISSET: Intensive	76.81	24.00	80.36	14.70	86.49	14.97	85.19	12.15	91.94	9.44	88.43	15.76	90.73	8.97	92.13	8.60
ISSET: Schoolwide	58.06	29.02	49.73	23.37	55.88	31.24	59.79	30.04	62.79	27.29	54.76	33.02	70.85	31.53	67.20	28.62
ISSET Overall Score	60.24	19.52	54.39	12.66	68.69	16.94	71.76	17.10	77.62	14.33	68.28	19.74	79.99	15.06	78.76	13.55

Note. SET = Schoolwide Evaluation Tool; ISSET = Individual Student Systems Evaluation Tool; RCT = randomized controlled effectiveness trial.

Data collectors were trained in four stages: an initial didactic session, on-site practice, on-site interobserver agreement or reliability, and on-site recalibration. Observers were required to meet 80% interobserver agreement in three practice classrooms prior to starting observations and again in on-site recalibrations conducted during active data collection. Interobserver agreement was calculated by dividing the total number of agreements by the total number of agreements and disagreements (Pas et al., 2015). Observers engaged in reliability testing until the criterion was met. For the current study, the average interobserver agreement for three classroom observations during the initial training was 88.23%. The range was 80% to 99%. Interobserver agreement rates were examined again during active data collection and were approximately 87.00%. Moreover, prior work on the ASSIST indicates the intraclass correlations on the ASSIST dimensions assessed across multiple within-classroom observations ranged from .72 to .81, with an average of .75, thus suggesting high reliability of the instrument within teachers/classrooms. We have also generally found limited evidence of systematic variation in ASSIST scores by time of day. See Pas et al. (2015) and Gaias et al. (2019) for further description of the ASSIST training and reliability and validity data.

Observers tallied discrete instances of teacher classroom management behaviors that occurred during a 15-min time frame, including (a) proactive behavioral management, (b) approval, (c) reactive behavior management, and (d) disapproval. Specifically, the ASSIST *proactive behavioral management* tally was defined as including all demonstrations of expectations provided verbally (e.g., explaining, reminding, commanding, prompting) and physically (e.g., modeling) prior to a problem behavior emerging. The *approval* tally was defined as recognition of students' performance through providing verbal praise, approving gestures (e.g., thumbs up), a tangible item, or physical contact (e.g., pat on the back). *Reactive behavior management* included teacher cues to redirect inappropriate behavior (e.g., touch, gesture, proximity, comment) but excluded disapprovals (see Pas et al., 2015, for additional information). *Disapproval* was the threat of or actual use of a tangible punitive consequence (e.g., detention), providing verbal criticism or using sarcasm, or gestural or physical contact demonstrating dissatisfaction with behavior. These four tallies were then aggregated into two summary scores. A *positive management* summary score was created by totaling the average frequencies of proactive behavioral management and approval. A *reactive management* summary score was created by totaling the average frequencies of reactive behavior management strategies and disapprovals. Observed counts of specific teacher behaviors were collected in each teacher's classroom and averaged across classrooms within each school to generate a single, school-level score for analysis in the current study.

School archival data. Data on suspensions (i.e., number of suspension events divided by total student enrollment) and school enrollment were obtained from the MSDE for the year prior to each school's first year of involvement, as a baseline indicator of school behavioral concerns and size. These two variables were included as control variables in the multilevel models.

MDS3 coach ratings of implementation of evidence-based interventions. In the summer of 2013 (i.e., Year 3 for 29 of the schools and Year 2 for two schools), coaches rated the phase of implementation the school had achieved for each of the offered EBPs, on a 7-point rating of phases of implementation scale based on a implementation phases framework (see Bradshaw, Debnam, et al., 2009), ranging from: (a) *Exploration* (i.e., identifying the need for change, learning about possible interventions that may be solutions, learning about what it takes to implement the innovation effectively, developing stakeholders and champions, deciding to proceed), (b) *Training* (i.e., EBP training provided to school personnel), (c) *Installation* (i.e., establishing the resources needed to use and implement an innovation with fidelity to achieve positive outcomes for students), (d) *Initial Implementation* (i.e., the first use of intervention practices by newly trained teachers and others school staff and district to support the new teaching), (e) *Full Implementation* (i.e., the skillful use of an innovation that is well integrated into teachers' repertoire and routinely supported by building and district administrations.), (f) *Innovation* (i.e., the advances in knowledge and skills that come from evaluated changes in how teachers and others make use of a science-based intervention), and (g) *Sustainability* (i.e., persistent and skillful support for teachers and staff who are using an innovation effectively, with each cohort of teachers achieving better results than the last; this is sometimes referred to as "regeneration" defined as "the set of procedures that allow a system to continually compare valued outcomes against current practice and modify practices to continue to achieve valued outcomes as the context changes over time"). Consistent with this scale, coaches rated at least some implementation of the following EBPs: Olweus Bullying Prevention Program (14 schools; 45.2%), LifeSkills (14 schools; 45.2%), CI/CO (28 schools; 90.3%), Check & Connect (25 schools; 80.6%), and CBITS (15 schools; 48.4%). The internal reliability for the items on these items is adequate ($\alpha = .81$). Coach ratings of the schools' implementation of the EBPs on average were moderately correlated (range of 0.37–0.49, $ps < .05$) with SET and ISSET scores, demonstrating convergent validity but also a unique contribution of this measure.

Analyses

We first conducted descriptive analyses on each SET and ISSET subscale and on the coach reports of implementation

of evidence-based interventions (see Table 1). We then conducted multilevel modeling using the HLM software (Raudenbush et al., 2008) to test the intervention effects on implementation fidelity (based on the SET and ISSET subscale scores in Aim 1) and observations of teacher practices (for Aim 2). Finally, we examined the association between school-level implementation fidelity and observations of teacher practices. All analyses included four data points (i.e., fall baseline in Year 1 and 3 subsequent springs). Specifically, two-level repeated measures models for continuous outcomes were fit for SET and ISSET scores and were modeled with the normal distribution; therefore, we report beta coefficients for these outcomes. Poisson models were fit for the teacher ASSIST (count) tally data; therefore, the Poisson distribution was used, which accounted for the fact that scores could not go below 0 and were unbounded on the high end of the range. The variance of the tallies exceeded the mean, so overdispersion was also accounted for (Cameron & Trivedi, 1998). Poisson regressions produced log-based coefficients that were exponentiated to present more easily interpretable event rate ratios (ERRs). Values less than 1 were desirable for negatively worded ASSIST tallies (e.g., reactive management summary); values greater than 1 were desirable for positively worded ASSIST tallies (e.g., positive management summary).

Repeated measures outcomes were modeled at Level 1. Time point (i.e., 0, 1, 2, and 3) was the only, uncentered, predictor at this level. Random error was freed both for the intercept and for slope of time equations at Level 2. Baseline measures of overall SET, overall ISSET, suspension rate, and enrollment were modeled to predict the intercept and slope, or change in implementation over time, and were grand-mean centered. The intervention status (i.e., 0 = comparison, 1 = intervention; uncentered) also predicted the slope. The effect of the intervention status on the implementation outcomes (Aim 1) and on classroom practices (Aim 2) was of interest. We also examined the association between baseline SET and ISSET scores and observations of teacher practices over time (Aim 3). Population averages with robust standard errors were used for the results, to maximize generalizability (Raudenbush et al., 2008). Spybrook's (2008) deltas (i.e., Δ) were calculated using adjusted HLM coefficients for the intervention effect, divided by the pooled standard deviation of the outcome, using the interpretations of Cohen's *d* (Cohen, 1992), where an effect of up to .20 was considered small, from .20 to .50 was considered moderate, and above .50 was considered large.

Results

Descriptive Analyses: Implementation of MTSS-B and the Selected EBPs

Prior to conducting the HLM analyses, we performed a series of descriptive analyses to better understand the overall pattern of MTSS-B implementation, as indicated by the SET and ISSET data (see Table 2 for SET and ISSET subscales scores). Other than Scale D (responding to behavioral violations), all baseline SET subscale scores were below the 80% threshold for high fidelity. In Year 1, the same was true, but additionally, the intervention group had > 80% on Scale E (monitoring and evaluation). By Year 2, on average, intervention schools met the 80% SET threshold for high fidelity implementation for most scales, whereas the comparison group continued to only meet 80% fidelity on Scales D and E. Baseline ISSET subscale scores were also initially low and showed incremental increases in both groups, but all of the average subscale scores, except for *Intensive Individualized Interventions* scale, were below 80% for both groups in all study years (see Table 2).

As described above, the coaches provided a rating (1 to 7) for each intervention schools' implementation status of each EBP the school had expressed interest to their coach in adopting; these decisions were guided by the school team's data-based decision-making process, which largely focused on ongoing review and analysis of the MDS3 School Climate Survey data and SET/ISSET data (see Table 2). The most commonly reached phase across all interventions was Training (42.8% of schools), followed by Exploration (22.1%), Initial Implementation (15.6%), and Full Implementation (11.0%). CI/CO was the most commonly discussed program (i.e., by 28 schools) and the only intervention where a coach rated at least one school at every phase of implementation. For CI/CO, four schools reached Exploration (i.e., 12.9% of all 31 intervention schools; 14.3% of the 28 schools discussing CI/CO), nine schools were Trained (i.e., 29.0% of all schools; 32.1% of schools discussing), one school reached Installation (3.2% of all schools; 3.6% of those discussing), six schools reached Initial Implementation (i.e., 19.4% of all schools; 21.4% of schools discussing), seven schools were rated as Fully Implementing (i.e., 22.6% of all intervention schools; 25% of schools discussing), and one school reached the Innovation phase (i.e., 3.2% of all intervention schools; 3.6% of schools discussing). Check & Connect was the next most commonly discussed program (i.e., 25 schools), followed by CBITS (i.e., 15 schools). Three schools each reached Exploration for Check & Connect (9.7% of all intervention schools; 12.0% of schools discussing) and CBITS (20.0% of those discussing). Eleven schools each were trained (i.e., 35.5% of all schools) in Check & Connect (44.4% of those discussing) and CBITS (73.3% of those discussing). For Check & Connect, seven reached Initial Implementation (i.e., 22.6% of all schools; 28.0% of those discussing) and four schools reached Full Implementation (i.e., 12.9% of all schools; 16% of those discussing). For CBITS, just one reached Full Implementation (i.e., 3.2% of all schools and 6.7% of those discussing). The Olweus Bullying Prevention Program and LifeSkills had the lowest uptake, with 14 schools discussing these programs. Four

schools (i.e., 12.9% of all schools; 28.6% of those discussing) reached each phase of Exploration, Training, and Installation for the Olweus Bullying Prevention Program and one school (i.e., 3.2% of all schools; 7.1% of those discussing) had each reached Initial Implementation and Full Implementation of the Olweus program. For LifeSkills Training, five schools (i.e., 16.1% of all and 35.7% of those discussing) reached Exploration and Training. One school reached Installation (i.e., 3.2% of all schools; 7.1% of those discussing) and three schools reached Initial Implementation (i.e., 9.7% of all schools; 21.4% of those discussing) for LifeSkills Training.

Aim 1: Intervention Effects on Implementation Fidelity

Scores on the overall SET score and scales regarding Behavioral Expectation Taught (B), System for Rewarding (C), and System for Responding (D) as well as all ISSET subscales significantly increased across the 3-year RCT (as displayed in the "Time" column in Table 3). There was a significant intervention effect on the SET subscales assessing Expectations Defined (SET A; $\beta=10.17$, $p<.01$; $\Delta=0.25$), Behavioral Expectations Taught (SET B; $\beta=6.82$, $p<.01$; $\Delta=0.21$), District-Level Support (SET G; $\beta=6.90$, $p<.01$; $\Delta=0.17$), and the SET overall score ($\beta=4.80$, $p<.01$; $\Delta=0.19$). The scores on each of these four SET scales increased more for intervention than comparison schools. The Spybrook's delta indicated small to medium effects ranging up to 1/4 standard deviation for each scale. See Table 3 for a full listing of results and see Figure 1 for a depiction of score changes over time. There were no significant intervention effects on the other SET scales or any of the ISSET scales.

Aim 2: Intervention Effects on Classroom Practices

Intervention status had a significant effect on reactive behavior management tallies over time (i.e., significant predictor of slope of time; $\beta=-0.09$; ERR=0.91; $p=.03$; see Table 4). The adjusted average occurrences of reactive behavior management for teachers in comparison schools were 1.51 across the 15-min observation; in intervention schools, the rate was 1.37 instances across the 15 min. Although this difference of just 0.14 in 15 min may seem modest, if one extrapolates the findings to a full hour, it increases to a difference of over half an instance, and in a school day, the differences is closer to three or four instances.

Over the course of the four waves of data collection (i.e., baseline and three outcome years), the frequency of reactive behavior management increased; the increase in this occurrence was slower for the MTSS-B intervention schools in contrast to the comparison schools (see Figure 2). Similarly,

there was a significant intervention effect on the reactive management summary score (i.e., reactive behavior management plus disapproval tallies; $\beta=-0.10$; ERR=0.90; $p=.01$); specifically, the average reactive management summary score increased in the comparison schools but declined in the intervention schools over time. The adjusted average total occurrence for the reactive management summary score was 3.90 in the comparison schools, over the course of the 15-min observations, whereas it was 3.52 in the intervention schools. There were no significant intervention effects on the positive tallies (i.e., proactive behavior management, approvals, or the summary score) or on the disapprovals tally.

Aim 3: Association Between Baseline ISSET and Classroom Practices

Baseline ISSET scores were significantly associated with the change in (i.e., slope of) proactive behavioral management ($\beta=-0.004$; ERR=0.996; $p<.01$) and with the positive management summary score, or proactive behavioral management plus approval tallies ($\beta=-0.003$; ERR=0.997; $p=.02$), but not the intercept. Schools with lowest baseline ISSET scores showed the greatest increases in their average tallied proactive behavioral management and positive management summary score (i.e., proactive behavioral management plus approvals) over time; this suggested a low level of MTSS-B implementation at baseline was associated with a significant increase in teachers' use of proactive behavior management strategies over the course of the trial. Baseline ISSET overall scores were significantly associated with the disapprovals intercept ($\beta=0.01$; ERR=1.01; $p=.04$), but not slope, whereby schools with higher baseline ISSET scores also had a larger number of tallied instances of disapprovals. Baseline ISSET was not associated with the reactive behavior management tally or reactive management summary score (i.e., reactive behavior management plus disapprovals) over time.

Discussion

This study reported intervention effects and implementation findings from an RCT testing the effectiveness of MTSS-B training and coaching in 58 high schools. Following the framework developed by the National Technical Assistance Center on PBIS (see www.pbis.org) on how to best install three-tiered MTSS-B in schools, the intervention schools received ongoing training and coaching regarding data-based decision-making; they focused on building the infrastructure and systems needed to implement multitiered systems of supports for behavior (also see Bradshaw et al., 2019; Lane et al., 2014). This is a novel contribution to the research field, as it examined how training and coaching in MTSS-B promoted implementation of PBIS and related EBPs across all three tiers. Furthermore, this research was

Table 3. HLM Findings for SET and ISSET Scales.

Implementation Outcome Variables	Intercept	Suspensions	Enrollment	SET/ISSET	Time	Intervention	Suspensions	Enrollment	SET/ISSET
SET Subscale Scores									
A: Expectations Defined	51.21*	0.18	0.01	−0.11	4.83	10.17*	0.00	−0.01	−0.01
B: Behavioral Expectations Taught	47.31*	0.05	0.01	0.20	5.82*	6.82*	0.05	0.00	−0.17
C: System for Rewarding Behavioral Expectations	35.26*	0.25	0.00	0.31	6.35*	4.35	0.09	0.00	−0.05
D: System for Responding to Behavioral Violations	84.09*	−0.17	0.01*	0.04	3.78*	−0.07	0.07	−0.00*	−0.02
E: Monitoring and Evaluation	77.28*	0.03	0.01	−0.01	1.90	1.93	0.13	0.00	−0.04
F: Management	62.77*	0.11	0.01	0.03	2.82	3.49	0.17	0.00	−0.03
G: District-Level Support	57.07*	0.49	0.00	0.19	2.95	6.90*	−0.14	0.00	−0.29*
SET Overall Score	59.28*	0.13	0.01	0.09	4.06*	4.80*	0.05	0.00	−0.09
ISSET Subscale Scores									
ISSET: Foundations	65.92*	−0.04	0.01	0.02	3.05*	1.40	0.07	0.00	0.01
ISSET: Schoolwide	53.08*	−0.19	0.02*	0.22	4.53*	0.18	0.19*	0.00	0.01
ISSET: Targeted	39.94*	0.18	0.02*	0.08	14.39*	0.79	0.08	0.00	0.01
ISSET: Intensive	80.03*	−0.42*	0.00	0.06	4.07*	0.46	0.18*	0.00	−0.03
ISSET Overall Score	59.75*	−0.12	0.01*	0.09	6.56*	0.60	0.13*	0.00	0.00

Note. All covariates reflect baseline data. For the SET outcomes, only baseline ISSET overall score was modeled; for ISSET outcomes, only baseline SET overall score was modeled. HLM = hierarchical linear modeling; SET = Schoolwide Evaluation Tool; ISSET = Individual Student Systems Evaluation Tool.
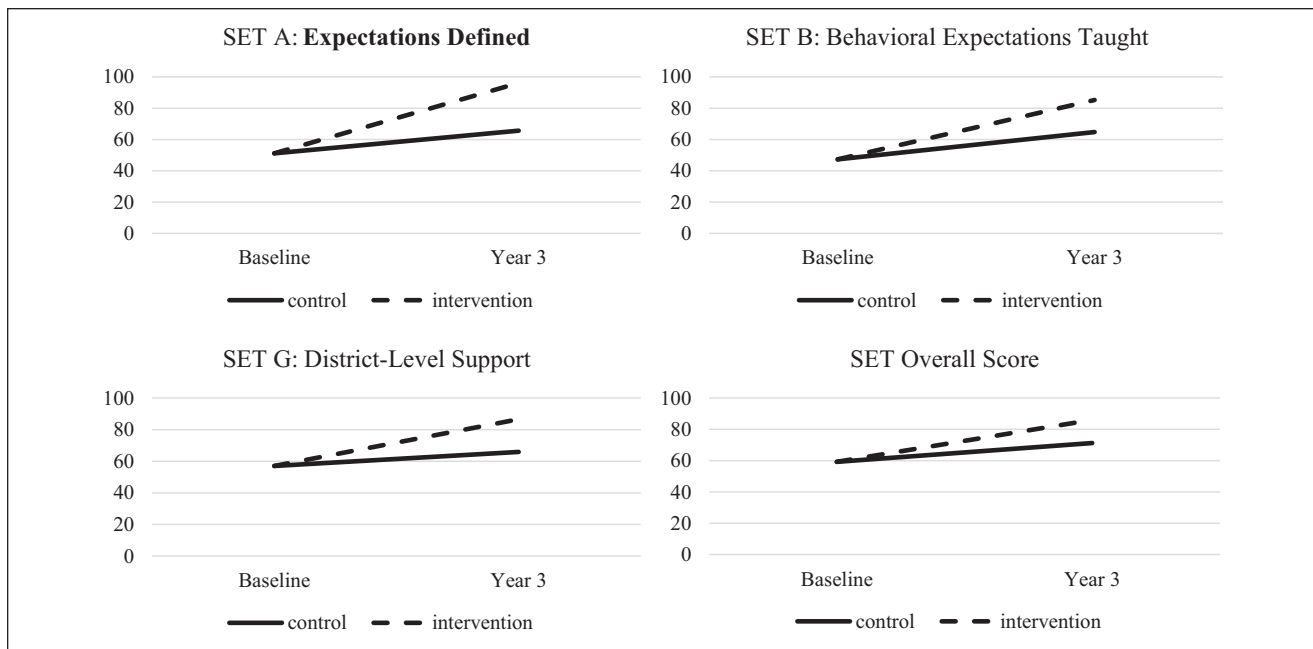*$p < .05$.

Figure 1. Intervention effects on SET subscales over time.
Note. SET = Schoolwide Evaluation Tool.

conducted in the context of a state-wide PBIS scale-up effort (Bradshaw & Pas, 2011), and thus, the implementation was managed by the state in conjunction with an implementation team, separate from the research team. The ongoing state-provided supports included annual trainings and three coaches' meetings per year. There was no direct support provided to schools, although some local school systems additionally offered quarterly, school-based coaches meetings. In contrast, much of the extant PBIS RCT research has focused on universal, school-wide implementation within elementary schools (e.g., see Bradshaw, Koth, et al., 2008, 2009, 2010; Horner et al., 2009; Lee & Gage, 2020). Similarly, the few high school–focused PBIS studies have largely used nonrandomized designs (e.g., Flannery et al., 2014), thereby limiting the extent to which causal inferences could be drawn (Murray, 1998). Given the widespread prevalence of the behavioral, social, emotional, and mental health concerns associated with EBD (Perou et al., 2013), it is important to explicitly study training and coaching to improve the implementation of EBPs across all tiers and to determine the impact on school and classroom practices to reduce the prevalence of EBD.

Results from the current study indicated that intervention schools demonstrated improved implementation fidelity and classroom management practices over the course of the study, and initial school-wide fidelity was predictive of improved teacher practice. Schools in the intervention condition demonstrated improvements in their defining and teaching of behavioral expectations (i.e., the Tier 1 core

foundations that coaches explicitly targeted in Year 1 of the trial). Interestingly, schools in the intervention condition also had better scores on the district-level support scale of the SET, which may have been an indication that MDS3 coaches provided not only study resources but also better connected schools to what was available by the district. A similar finding emerged in a prior RCT of Tier 1 PBIS (Bradshaw, Koth, et al., 2009). Finally, overall school-wide implementation SET scores improved for the intervention schools. It is also worth noting that for several of the SET (and ISSET) subscales, schools in both conditions had significant improvements or growth in their scores over time. This is quite possibly due to the process of receiving information on the SET and/or ISSET implementation, as both types of data were made available to schools in both conditions; however, only schools in the intervention condition received training and coaching from an MDS3 coach on how to use these data. We lacked a true control condition which received no training in Tier 1 supports. As such, we are unable to draw causal conclusions regarding the overall increases on these scales over the course of the trial across schools in both conditions.

With regard to the ISSET scores, despite the growth in ISSET scores on average across several of the subscale scores, there were no significant intervention effects. These null findings for the school-wide and targeted interventions scales may be related to the generally low uptake of the offered EBPs. For example, the most common phase of implementation was Training, whereby approximately half

Table 4. Multilevel Results for ASSIST Tallies.

Implementation Predictor Variables	Proactive		Approvals		Positive Summary		Reactive		Disapprovals		Reactive summary	
	ERR	CI	ERR	CI	ERR	CI	ERR	CI	ERR	CI	ERR	CI
Intercept	4.47*	[4.04, 4.95]	2.06*	[1.83, 2.33]	6.53*	[5.95, 7.17]	1.51*	[1.32, 1.72]	3.27*	[2.77, 3.87]	3.90*	[3.49, 4.36]
Suspensions	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]	1.01	[1.00, 1.02]	1.00	[1.00, 1.01]
Enrollment	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]
SET	1.00	[1.00, 1.01]	1.00	[1.00, 1.01]	1.00	[1.00, 1.01]	1.00	[1.00, 1.01]	1.00	[1.00, 1.01]	1.00	[1.00, 1.01]
ISSET	1.00	[1.00, 1.01]	1.00	[0.99, 1.01]	1.00	[1.00, 1.01]	1.00	[0.99, 1.00]	1.01*	[1.00, 1.02]	1.01	[1.00, 1.01]
Time	1.09*	[1.03, 1.15]	1.03	[0.97, 1.11]	1.07*	[1.02, 1.12]	1.53*	[1.41, 1.66]	0.17*	[0.12, 0.23]	1.06	[0.98, 1.14]
Intervention	0.99	[0.93, 1.06]	0.97	[0.89, 1.06]	0.99	[0.93, 1.04]	0.91*	[0.84, 0.99]	1.27	[0.82, 1.96]	0.90*	[0.83, 0.98]
Suspensions	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.01*	[1.00, 1.01]	1.01	[1.00, 1.02]	1.004*	[1.00, 1.01]
Enrollment	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]
SET	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	1.00	[0.99, 1.01]	1.00	[1.00, 1.00]
ISSET	0.996*	[0.99, 1.00]	1.00	[0.99, 1.00]	0.997*	[0.99, 1.00]	1.00	[1.00, 1.01]	0.99	[0.98, 1.00]	1.00	[1.00, 1.00]

Note. All covariates reflect baseline data. SET and ISSET reflect the overall scores. ASSIST = Assessing School Settings: Interactions of Students and Teachers; ERR = event rate ratio; CI = confidence interval; SET = Schoolwide Evaluation Tool; ISSET = Individual Student Systems Evaluation Tool. *$p < .05$.

of the intervention schools in the study received training in an EBP; however, few of these schools reached the Installation or Implementation phases. Relatedly, the limited uptake of a wide range of EBPs and the relative focus on CI/CO and Check & Connect is consistent with other literature identifying the most common Tier 2 interventions (Bruhn et al., 2013). These two EBPs generated the most interest and had the highest number of schools reaching the Implementation phases. As noted above, on average, all schools within the study showed significant growth on the ISSET scales, although they never reached high fidelity on any of these scales, except the one measuring *Intensive Individualized Interventions*.

Another way to conceptualize the SET and ISSET subscale scores is in relation to the number of schools that met the 80% criterion for fidelity. Among intervention schools, 68% met the 80% criteria on the SET at Year 3, and 61% met these criteria for the ISSET in Year 3. In contrast, just 48% of the comparison schools reached these criteria on the overall SET and 56% met it for the ISSET score in Year 3. Regarding the individual subscale scores, there was a considerable gap between the percentage of comparison and intervention schools achieving 80% on SET A (set expectations), E (monitoring), and F (management). It seems quite likely that the additional training and coaching provided to the intervention schools helped improve implementation in these areas, over and above the "training as usual" provided by the state. Together, this highlights the "value added" with the more comprehensive training and support provided in the intervention condition compared with the traditional focus on Tier 1 positive behavior supports by the state at the time of the study.

In addition to improving school-wide implementation, the intervention schools also demonstrated improvements in the use of classroom-based management strategies.

Specifically, the average rates of reactive behavior management increased less over time for intervention schools and the total of reactive behavior management plus disapprovals (reactive management summary score) declined in intervention schools, while the rates increased in comparison schools. Although the raw averages presented in the results may seem small, this was a significant difference in just 15 min of observation. If extrapolating to the entire school day, this would reflect a difference of about three instances of reactive strategies; taken across days, weeks, and months, this is much more practically significant. This finding is important because the MDS3 coaches did not work directly with teachers, but rather were focused on systems coaching consistent with the National PBIS Technical Assistance model; therefore, the coaches only indirectly supported teachers. As a result, this finding indicates that improvements in the school-wide MTSS-B approach reached classrooms, in the form of shift away from more reactive and punitive responses to behavior (Pas et al., 2015; Reinke et al., 2016). It is possible that the improvements detected on the SET regarding setting and teaching of behavioral expectations resulted in school-wide improvements of student positive behavior and that teachers, in turn, did not increase their rates of reactive behavior management at the rate of teachers in comparison schools.

Although there were no main intervention effects on the positive and proactive classroom management strategies (i.e., approvals and proactive behavior management), the implementation analyses suggested schools with the poorest initial multitiered implementation, as measured by the ISSET, and thus demonstrating the most room for improvements, had significant improvements on both positive and proactive classroom management strategies even in the absence of such direct teacher coaching. To change teacher proactive behavior management, it is likely that
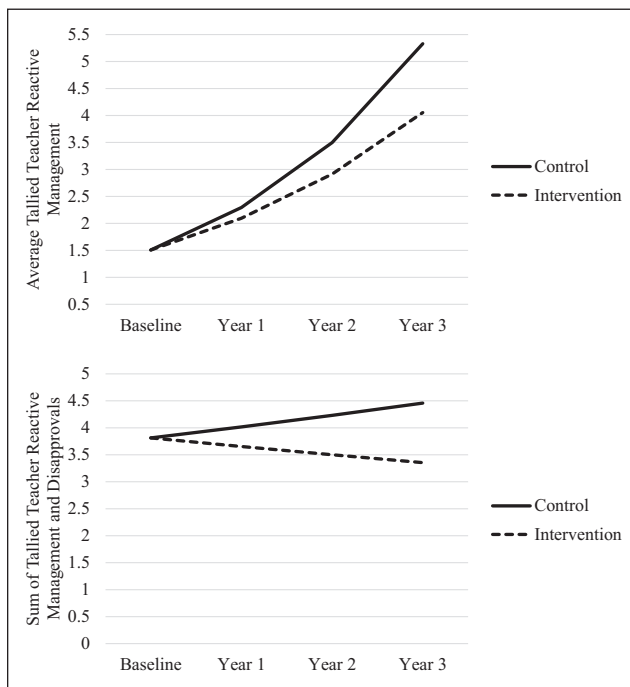
Figure 2. Intervention effects on ASSIST tally changes over time (top: reactive behavior management, bottom: reactive management summary score or reactive management plus disapprovals).
Note. ASSIST = Assessing School Settings: Interactions of Students and Teachers.

classroom-based coaching would be needed. In fact, prior research assessing the effectiveness of classroom-based coaching has demonstrated significant improvements in teachers' use of proactive behavior management strategies (e.g., see Reinke et al., 2008).

Limitations and Future Research

Although this research includes many strengths, most notably the randomized effectiveness design, the multitiered implementation of MTSS-B in high schools, and the assessment of fidelity from external and internal sources across all three tiers, there are some limitations to consider. Implementation is nuanced and complex (Fixsen et al., 2005), and prior PBIS research indicates the importance of examining fidelity measures in detail (i.e., at the subscale level) rather than just at the aggregated ("overall score") level (see Pas, Johnson et al., 2019). Therefore, we included each SET and ISSET subscale score in examining the first aim of this study; this produced 13 implementation main effects findings and may have increased the likelihood of detecting significant effects. However, the findings reported were all at a $p < .01$ and were quite consistent in that the SET, but not ISSET, was found to be positively

impacted by ongoing training and support. As in other studies of the SET and ISSET (see Debnam et al., 2012; Pas, Johnson et al., 2019), the internal consistency of some of the subscales were lower than the preferred .80 threshold; this may be the result of few items on those particular subscales, in conjunction with limited variability on the responses. As such, additional attention is needed to the psychometrics of fidelity measures. We also relied on the coach reports of the implementation of the EBPs; however, additional sources of information on these practices would have provided greater insight into the adoption of the specific EBPs.

Although the conduct of external observations of classrooms is a strength, the anonymity of students did not allow for the analysis of how students, and particularly those with EBD or identified disabilities, were impacted by the classroom practices. We also sampled general education classrooms, which often included students with EBD and those receiving other special education supports. As such, we are unable to formulate conclusions specific to students with EBD. Future research should consider embedding a targeted, student-focused research design within the group-randomized design to more explicitly examine how students with greater emotional and behavioral needs respond to MTSS-B. There was some principal and school staff turnover during the project, but it was independently managed and beyond the control of the project. We did not systematically track it within the study and thus were unable to adjust for it in the analyses. Given the school-level random assignment and the fact that some staff and leadership turnover is common, we do not believe that it was a significant concern in this study.

This research was conducted in the state of Maryland, with training and implementation led by the state and their implementation partner (SPHS; see Bradshaw et al., 2012). It is possible these findings would not generalize to other states with less state-wide infrastructure to support PBIS training and to provide ongoing coaching. Due to the study design, we were unable to evaluate the effectiveness of any one particular EBP or combination of EBPs. Moreover, there was a relatively low rate of uptake of these additional EBPs, particularly when it came to full implementation. While many schools were interested in and received training in the EBPs, few followed through to implementation. This was particularly true for programs addressing bullying (i.e., Olweus Program), substance use and general social skills (i.e., LifeSkills), and addressing trauma (i.e., CBITS). Anecdotally, schools and MDS3 coaches reported the ongoing resources (e.g., time, staff) to implement these programs were a major barrier. Furthermore, it is important to note that coaching is a dynamic process, and while all coaches were trained and supervised by the team of experts at SPHS, which was affiliated with the National PBIS Technical Assistance Center (see Bradshaw et al., 2014), there may be

some MDS3 coaches who were a better match or formed stronger alliances with their schools than others (see Johnson et al., 2016). Furthermore, issues around buy-in and readiness to implement MTSS-B and EBPs, while considered in the coaching process, were not explicitly measured in this study. Finally, we used the term MTSS-B as a broader framework encompassing the traditional Tier 1 PBIS model, along with integrated implementation of EBPs at Tiers 2 and 3. As noted above, the terminology used to describe the intervention was more focused on PBIS at the time the study was conducted (2011); as such, we have aimed to map the description of the model tested in the trial onto more common-day and inclusive terminology, which largely uses the term MTSS-B to describe the fuller three-tiered model including implementation of other EBPs (Bradshaw et al., 2019).

Conclusions and Implications

Taken together, the findings of this study suggest training in the MTSS-B framework leads to implementation of Tier 1 elements in high schools, and measurable impacts in the classroom. Research indicates it can take 3 to 7 years for systemic change to occur (Fixsen et al., 2005), and it is possible 3 years was not sufficient time to improve implementation across all tiers; in fact, implementation of PBIS, MTSS-B, and other EBPs may be more protracted in high schools (see Flannery et al., 2014; Pas, Ryoo et al., 2019). Perhaps with a longer assessment window, schools may have more time to solidify the Tier 1 foundational elements and then show enhanced readiness to layer on additional EBPs across all three tiers (Fixsen et al., 2005).

Additional scale-up research is needed to examine EBP implementation in schools (see Fagan et al., 2019; Lloyd et al., 2019) to determine how to the optimize support provided, to ensure high implementation fidelity, and to translate these efforts into improved student outcomes. Although the schools in this study all expressed a high degree of interest in training in these specific EBPs at recruitment into the project, which was also evidenced through the number of intervention schools that requested and received training in each of the offered EBPs, full implementation of the EBPs rarely occurred. Nevertheless, there were significant improvements in the classroom context and teacher behaviors, which likely resulted from the Tier 1 implementation supports through school-wide Tier 1 training, rather than the subsequent introduction of other EBPs within this framework. This improvement in classroom practices, even in the absence of specific EBPs, may prove critical for enhancing both the school context and outcomes for all students, including those with or at risk for EBDs (Bradshaw et al., 2015, 2019; Lloyd et al., 2019). Although the current study design precluded a tracking or disaggregation of data for students with or at risk for EBD, a prior RCT of school-wide PBIS (i.e., Tier 1) at the elementary level indicated significant impacts on at-risk students; in fact, the school-wide PBIS effects were larger for students with more elevated behavioral and social-emotional concerns at baseline, compared with their lower risk peers (Bradshaw et al., 2015). This suggests that even the universal elements implemented in the current trial may translate into significant impacts for students with or at risk for EBD, regardless of the presence of Tier 2 or 3 supports. Moreover, these effects may be even greater for students with or at risk for EBD relative to their peers.

ORCID iD

Catherine P. Bradshaw https://orcid.org/0000-0003-2048-3225

References

Botvin, G. J., Griffin, K. W., & Nichols, T. R. (2006). Preventing youth violence and delinquency through a universal school-based prevention approach. *Prevention Science*, *7*, 403–408.

Bradshaw, C. P., Debnam, K. J., Koth, C., & Leaf, P. J. (2009). Preliminary validation of the implementation phases inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions*, *11*, 145–160. https://doi.org/10.1177/1098300708319126

Bradshaw, C. P., Debnam, K. J., Lindstrom Johnson, S., Pas, E. T., Hershfeldt, P., Alexander, A., Barrett, S., & Leaf, P. (2014). Maryland's evolving system of social, emotional, and behavioral interventions in public schools: The Maryland Safe and Supportive Schools Project. *Adolescent Psychiatry*, *4*(3), 194–206. https://doi.org/10.2174/221067660403140912163120

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N. S., & Leaf, P. J. (2008). The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly*, *23*, 462–473.

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide Positive Behavioral Interventions and Supports: Findings from a group-randomized effectiveness trial. *Prevention Science*, *10*, 100–115. https://doi.org/10.1007/s11121-008-0114-9

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide Positive Behavioral Interventions and Supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, *12*, 133–148. https://doi.org/10.1177/1098300709334798

Bradshaw, C. P., & Pas, E. T. (2011). A statewide scale up of positive behavioral interventions and supports: A description of the development of systems of support and analysis of adoption and implementation. *School Psychology Review*, *40*, 530–548.

Bradshaw, C. P., Pas, E. T., Bloom, J., Barrett, S., Hershfeldt, P., Alexander, A., McKenna, M., Chafin, A. E., & Leaf, P. (2012). A state-wide partnership to promote safe and supportive schools: The PBIS Maryland Initiative. *Administration and Policy in Mental Health and Mental Health Services Research*, *39*(4), 225–237. https://doi.org/10.1007/s10488-011-0384-6

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of school-wide Positive Behavioral Interventions and Supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children*, *31*, 1–26. https://doi.org/10.1353/etc.0.0025

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2015). Examining variation in the impact of school-wide positive behavioral interventions and supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology*, *107*(2), 546–557.

Bradshaw, C. P., Williamson, S. K., Kendziora, K., Jones, W., & Cole, S. (2019). Multi-tiered approaches to school-based mental health, wellness, and trauma response. In D. Osher, R. Jagers, K. Kendziora, M. Mayer & L. Wood (Eds.), *Keeping students safe and helping them thrive: A collaborative handbook for education, safety, and justice professionals, families, and communities* (2 vols., pp. 85–110). Praeger.

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2013). A review of Tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders*, *22*, 171–189. https://doi.org/10.1177/1063426613476092

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Debnam, K. J., Pas, E., & Bradshaw, C. P. (2012). Secondary and tertiary support systems in schools implementing School-wide Positive Behavioral Interventions and Supports: A preliminary descriptive analysis. *Journal of Positive Behavior Interventions*, *14*, 142–152. https://doi.org/10.1177/1098300712436844

Fagan, A., Bumbarger, B., Barth, R., Bradshaw, C. P., Rhoades Cooper, B., Supplee, L. H., & Walker, D. K. (2019). Scaling up evidence-based interventions in U.S. public systems to prevent behavioral health problems: Challenges and opportunities. *Prevention Science*, *20*, 1147–1168. https://doi.org/10.1007/s11121-019-01048-8

Fixsen, D., Naoom, S., Blasé, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). https://fpg.unc.edu/node/4445

Flannery, K. B., Fenning, P., McGrath Kato, M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, *29*(2), 111–124. https://doi.org/10.1037/spq0000039

Gaias, L. M., Lindstrom Johnson, S., Bottiani, J. H., Debnam, K. J., & Bradshaw, C. P. (2019). Examining teachers' classroom management profiles: Incorporating a focus on culturally responsive practice. *Journal of School Psychology*, *76*, 124–139. https://doi.org/10.1016/j.jsp.2019.07.017

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, *16*(7), 893–926. https://doi.org/10.1007/s11121-015-0555-x

Hawken, L., & Horner, R. (2003). Evaluation of a targeted intervention within a school wide system of behavior support. *Journal of Behavioral Education*, *12*, 225–240.

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

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*, 133–144. https://doi.org/10.1177/1098300709332067

Horner, R. H., Todd, A., Lewis-Palmer, T., Irvin, L., Sugai, G., & Boland, J. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *6*(1), 3–12.

Johnson, S., Pas, E., & Bradshaw, C. P. (2016). Understanding and measuring coach-teacher alliance: A glimpse inside the "black box." *Prevention Science*, *17*(4), 439–449.

Lane, K. L., Oakes, W. P., Jenkins, A., Menzies, H. M., & Kalberg, J. R. (2014). A team-based process for designing Comprehensive, Integrated, Three-Tiered (CI3T) models of prevention: How does my school-site leadership team design a CI3T Model? *Preventing School Failure: Alternative Education for Children and Youth*, *58*, 129–142. https://doi.org/10.1080/1045988x.2014.893976

Lee, A., & Gage, N. A. (2020). Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychology in the Schools*, *57*(5), 783–804. https://doi.org/10.1002/pits.22336

Lewis-Palmer, T., Todd, A. W., Horner, R. H., Sugai, G., & Sampson, N. K. (2005). *Tool, version 1.2*. Educational and Community Supports, University of Oregon.

Lloyd, B., Bruhn, A., Sutherland, K. S., & Bradshaw, C. P. (2019). Progress and priorities in research to improve outcomes for students with or at risk for emotional and behavioral disorders. *Behavioral Disorders*, *44*(2), 85–96. https://doi.org/10.1177/0198742918808485

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford University Press.

Musu, L., Zhang, A., Wang, K., Zhang, J., & Oudekerk, B. A. (2019). *Indicators of school crime and safety: 2018* (NCES 2019-047/NCJ 252571). National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. https://nces.ed.gov/pubs2019/2019047.pdf

Olweus, D., Limber, S. P., Flerx, V. C., Mullin, N., Riese, J., & Snyder, M. (2007). *Olweus bullying prevention program: Schoolwide guide*. Hazelden.

Pas, E. T., Cash, A. H., O'Brennan, L., Debnam, K., & Bradshaw, C. P. (2015). Profiles of classroom behavior in high schools: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology*, *53*, 137–148. https://doi.org/10.1016/j.jsp.2014.12.005

Pas, E. T., Johnson, S. R., Debnam, K. J., Hulleman, C., & Bradshaw, C. P. (2019). Examining the relative utility of PBIS implementation fidelity scores in relation to student outcomes. *Remedial and Special Education*, *40*(1), 6–15. https://doi.org/10.1177/0741932518805192

Pas, E. T., Lindstrom Johnson, S., Alfonso, Y. N., & Bradshaw, C. P. (2020). Tracking time and resources associated with systems change and the adoption of evidence-based programs: The "hidden costs" of school-based coaching. *Administration and Policy in Mental Health and Mental Health Services Research*, *47*, 720–734. https://doi.org/10.1007/s10488-020-01039-w

Pas, E. T., Ryoo, J. H., Musci, R., & Bradshaw, C. P. (2019). A state-wide quasi-experimental effectiveness study of the scale-up of school-wide Positive Behavioral Interventions and Supports. *Journal of School Psychology*, *73*(1), 41–55. https://doi.org/10.1016/j.jsp.2019.03.001

Perou, R., Bitsko, R. H., Blumberg, S. J., Pastor, P., Ghandour, R. M., Gfroerer, J. C., . . . Huang, L. N. (2013). Mental health surveillance among children—United States, 2005-2011. *Morbidity and Mortality Weekly Report*, *62*, 1–35.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T. Jr., & du Toit, M. (2008). *Hierarchical Linear and Nonlinear Modeling (HLM 7.03)*. Scientific Software International, Inc.

Reinke, W. M., Herman, K. C., & Newcomer, L. (2016). The Brief Student–Teacher Classroom Interaction Observation: Using dynamic indicators of behaviors in the classroom to predict outcomes and inform practice. *Assessment for Effective Intervention*, *42*(1), 32–42. https://doi.org/10.1177/1534508416641605

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*(3), 315–332.

Rusby, J. C., Taylor, T., & Milchak, C. (2001). *Assessing school settings: Interactions of students and teachers (ASSIST) observation system* [Unpublished manual].

Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children*, *71*, 465–482.

Spybrook, J. (2008). Power, sample size, and design. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 273–311). Information Age.

Stein, B. D., Jaycox, L. H., Kataoka, S. H., Wong, M., Tu, W., Elliott, M. E., & Fink, A. (2003). A mental health intervention for schoolchildren exposed to violence: A randomized controlled trial. *Journal of the American Medical Association*, *290*, 603–611.

Sugai, G., & Horner, R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, *35*, 245–259.

Sugai, G., Lewis-Palmer, T., Todd, A. W., & Horner, R. H. (2001). *School-Wide Evaluation Tool (SET)*. Center for Positive Behavioral Supports, University of Oregon.

U.S. Department of Education. (2018). *Final report of the federal commission on school safety*. https://www2.ed.gov/documents/school-safety/school-safety-report.pdf