Appendix

Study of Training in Multi-Tiered Systems of Support for Behavior: Impacts on Elementary School Students' Outcomes

U.S. Department of Education
Miguel Cardona
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Evaluation and Regional Assistance

Matthew Soldner	Marsha Silverberg
Commissioner	*Associate Commissioner*

Lauren Angelo
Project Officer

July 2022

Study of Training in Multi-Tiered Systems of Support for Behavior: Impacts on Elementary School Students' Outcomes

July 2022

Barbara Condliffe
Pei Zhu
Fred Doolittle
Mark van Dok
Hannah Power
Dakota Denison
MDRC

Anja Kurki
American Institutes for Research

CONTENTS

LIST OF EXHIBITS

INTRODUCTION

Students' early problem behaviors in school can be disruptive and even hinder their learning and long-term success. To prevent and address these problem behaviors, schools across the country report adopting Multi-Tiered Systems of Support for Behavior (MTSS-B). The MTSS-B approach seeks to change the school learning environment by consistently teaching and reinforcing good behavior for all students, and then identifying and providing supplemental supports to students who need it. Given the reported widespread use of MTSS-B but limited evidence of effective programs, this study evaluated a promising, intensive program of MTSS-B training and technical assistance. About 90 elementary schools in six states were randomly assigned either to participate in the program or to continue with their usual strategies for supporting student behavior. Comparing student and teacher experiences in both sets of schools measures the effectiveness of the program. This document provides supporting details on what was evaluated and how the study was carried out including sensitivity analyses for the findings presented in the report.

APPENDIX A. THE STUDY'S MTSS-B TRAINING AND TECHNICAL ASSISTANCE PROGRAM

This appendix describes the version of MTSS-B that participating schools were expected to implement; the training and technical assistance program that the Center for Social Behavior Support (CSBS) delivered to participating schools; the coaching supports provided to participating schools; and the additional implementation supports the study team provided to CSBS, districts, and the schools. It concludes with a brief overview of the cost of the training and technical assistance.

I. Overview of the MTSS-B Practices and Structures Targeted by the Study Program

This section describes the study's intended approach to MTSS-B implementation in schools. The program focused on three primary MTSS-B components, school-wide practices (Tier I), supplemental supports (Tier II), and an infrastructure to support these practices.[1] The program did not provide support for schools to carry out more intensive, often one-on-one supports, sometimes referred to as Tier III. The logic of how the program is intended to work–as fielded in the study–is that implementation of the school-wide practices and supplemental supports aided by the MTSS-B infrastructure will lead to improvements in school climate and classroom practices and these will produce improvements in student behavior, which will lead to improved academic achievement.

School-Wide Practices (Tier I)

The program helps schools develop and adopt three to five positively stated *school-wide behavioral expectations*–for example, be safe, be respectful, be prepared. School teams are to set expectations that reflect the culture of the school. These expectations are intended to be displayed prominently throughout the school, including on posters in classrooms, hallways, and the cafeteria. Using a school-developed lesson plan customized for each grade level, teachers should help all students learn how to meet these consistent expectations in the different locations of the school, explicitly demonstrating appropriate behaviors and helping students practice them. Teachers are asked to deliver the lessons to students at the beginning of the year and reteach them throughout the year when students need reinforcement. A *school-wide acknowledgment system* typically further reinforces the behavioral expectations. All staff are supposed to frequently and positively acknowledge students for displaying the expected behaviors by naming the specific positive behavior that the student is displaying to help reinforce it. For example, at the start of class a teacher might observe a student's responsible behavior and acknowledge it by saying, "I noticed you were on time this morning; that's very responsible." School staff members similarly should acknowledge appropriate behavior in non-classroom settings where school-wide expectations apply. Further, schools were trained to establish systems for reinforcing behavior meeting expectations through incentives such as more choice in activities, special roles in the classroom that students enjoy, or tickets or points that could be exchanged for a tangible positive reinforcement such as extra recess time or more time for free reading. In these systems, it is important that the teachers and staff make sure the children know the specific behavior that earned them the positive reinforcement. Additionally, teachers are to promote the school-wide behavioral expectations in the classroom by using a specific set of *classroom management practices*. (See Exhibit A.1.)

Finally, school staff are expected to handle behavioral violations consistently by putting in place *school-wide discipline policies* that define problem behaviors and how to address them. These policies should clearly indicate what constitutes minor behavioral violations that the teacher should correct in the classroom and major behavioral violations for which a student should be referred to the school office. Minor behavioral violations, like talking out of turn, are opportunities to reteach positive behavior and can be addressed with practices that de-escalate the behavior and minimize the disruption of classroom activity. Major behavioral violations, like physical aggression, present a threat to school and student physical or emotional safety and need attention from an administrator.

Exhibit A.1. Eight Classroom Management Practices Supported by the Study

1. Define and teach classroom rules aligned with the school-wide behavioral expectations.

2. Explain and teach classroom routines.

3. Use the positive behavior game strategy to provide class-wide group contingencies that reinforce positive behavior.

4. Provide specific praise for displays of appropriate behavior.

5. Correct errors by prompting, reteaching, and providing choices to students.

6. Supervise students actively–move, scan, interact.

7. Give students multiple opportunities to respond.

8. Arrange orderly physical environments.

Some features of MTSS-B were not included or were limited, in this study, because schools within a district were randomly assigned to either participate in the training or continue with their usual services. To avoid spillover to the non-participating schools, district-level support for implementation was largely limited to the work of a district-based coach (funded by the project) and the efforts of the district study coordinator to help with logistics in the participating schools. Furthermore, efforts to build support among school staff prior to random assignment were not undertaken because random assignment was used to assign schools to implement MTSS-B or continue with their usual practices. The recruitment team and district leaders made presentations about the MTSS-B program and the research design to principals in schools that were considering joining the study prior to random assignment.[2] However, teachers and other staff were not involved at this stage because their school's status in the study was not yet known. Once random assignment occurred, the initial readiness activities, school-wide training, and kickoff events listed in Exhibit A.3 were intended to build staff buy-in for implementation in the participating schools.

Supplemental Supports (Tier II)

Students who consistently do not meet the school-wide behavioral expectations should receive supplemental supports to promote positive behaviors. For example, teachers and other school staff members should refer students to receive a supplemental supports strategy for repeated disruptive behaviors in the classroom that are inconsistent with a school's expectation for school safety. Teachers and staff should also refer students who exhibit less overt behaviors if they disrupt learning, such as consistently being late for class or forgetting to do homework.

Various types of supplemental supports are available to help students meet behavioral expectations. This study used Check In Check Out (CICO) as the Tier II supplemental supports.[3] CICO is intended for students who are not responding to Tier I and the MTSS-B school-wide expectations, are not in crisis, and find adult attention rewarding. CICO is hypothesized to improve student behavior by providing more structure in support of behavioral expectations, a positive contact with an adult at the start of each day to "set up success", and an increase in the frequency of feedback that is closely tied to student behavior.[4]

To implement CICO with students, a Tier II team should identify staff facilitators who generally have positive relationships with students and are consistently available to briefly meet with students individually at the start and end of each day. Facilitators are generally not classroom teachers because classroom teachers typically are not consistently available for such meetings. During morning "check-in," facilitators are supposed to encourage each enrolled student to meet the behavioral expectations that they are struggling with the most and help them to set behavior goals for the

day. Throughout the day, teachers should provide brief individual feedback to students on their progress toward their goals and document that feedback on a progress card. This card typically lists a small number of expectations with specific behaviors that illustrate success and asks teachers and other staff working with students at recess and lunch to indicate how well they met the expectation using a numbered scale. During the afternoon "check-out," facilitators use the progress card to reflect with students on the feedback they receive and to reinforce positive behaviors by acknowledging accomplishments. In the training, schools were encouraged to include tangible positive reinforcements for students (for example, for collecting and turning in the progress card and for meeting a daily goal of points) and to ask parents of enrolled students to review and sign the progress card each day. The Tier II team is expected to meet weekly to review students' progress and adjust the intervention as needed or end it when students achieve sufficient success.

MTSS-B Infrastructure

An infrastructure typically supports MTSS-B implementation and includes administrative structures and procedures and a data system. For this study, each participating school formed a school team for each tier, was supported by a district-based MTSS-B coach, and utilized a data system for monitoring student behavior and implementation progress, as described in Exhibit A.2.

Exhibit A.2. MTSS-B Infrastructure in the Study

- **School team:** At each participating school, an administrator, such as a principal, should have established a school team for each tier that represented the full school community and included individuals with appropriate expertise in student behavioral issues. While the administrator was intended to attend all the team meetings, they were to designate a different staff member to lead each team. The school-wide (Tier I) and supplemental support (Tier II) teams were intended to develop school implementation plans, meet at least monthly to monitor and adapt the implementation plans based on reviews of data (see "Data System," below), and train and support other school staff members in implementing the practices. The Tier II team was also responsible for identifying students to receive the supplemental support intervention—based on staff member or parent referral and reviews of data—and for using the student progress monitoring data to make decisions about when students should exit the intervention or be referred for more intensive supports.

- **MTSS-B coach:** A district-based MTSS-B coach was supported by the study to spend one day per week in each school to provide additional training and technical assistance. The coach was intended to provide school teams with supplemental content training on MTSS-B practices and procedures as well as logistical support in carrying out the team roles and functions. For example, the coach could help the school-wide team analyze student behavior data to identify expectations that needed to be reinforced with students. The coach was also intended to provide direct support to school staff who, based on reviews of behavior data and/or self-nomination, needed additional assistance in implementing the MTSS-B practices. For example, the coach could observe teachers' use of classroom management practices and provide them with feedback based on data and a classroom demonstration to help improve practice.

(continued)

Exhibit A.2 (continued)

- **Data system:** A data system was provided to participating schools to collect data on student behaviors and to support the school teams' analysis of those data for monitoring the progress of implementation. This study used three components of Positive Behavioral Interventions and Supports (PBIS) Apps as the behavior monitoring data system for participating schools:

  o The School-Wide Information System (SWIS): SWIS tracks disciplinary referrals for each student by location, time, and problem behaviors. Using the data system's analytic reports, the Tier I school team for school-wide practices can monitor student behavior and adapt the implementation plan as needed. For example, the team can use a report that breaks down the places in the school where referrals are most common so staff can provide additional support to those areas. Additionally, both the team for school-wide practices (Tier I) and the team for supplemental supports (Tier II) can use the referral information to identify students in need of supplemental supports.

  o Check In Check Out SWIS (CICO-SWIS): For students enrolled in the Tier II supplemental support intervention, CICO, a staff member inputs data on the students' daily progress into CICO-SWIS so the school team for supplemental supports (Tier II) can monitor progress and decide when students are ready to exit the intervention.

  o PBIS Assessment: PBIS Assessment includes surveys that allow schools and districts to collect and analyze data on each school's implementation fidelity. For this study, schools were supposed to administer the PBIS Self-Assessment Survey (SAS) to all staff members each spring to assess their perceptions of the implementation of the MTSS-B practices and to help teams set priorities and action plans for the next year. Teams and coaches were also supposed to complete the Tiered Fidelity Inventory (TFI) three to four times during the year to monitor implementation fidelity and make adjustments when needed.*

NOTE: *Algozzine et al. (2014).

II. Focus and Structure of the MTSS-B Training and Technical Assistance Program

The study's training and technical assistance program was intended to support each participating school in initiating and sustaining the MTSS-B practices described above. CSBS mostly followed a "train the trainer" model, providing direct training and technical assistance to the school team for each tier, school administrators, and MTSS-B coaches, who then trained and supported school staff in implementing the MTSS-B practices. During Program Year 1, the focus was on training and technical assistance for the school-wide practices (Tier I), and during Program Year 2, the focus was on training and technical assistance for both Tier I and Tier II–the school-wide practices and the supplemental supports.

For each tier, the training and technical assistance included three components: (1) a readiness phase to identify and install the infrastructure (described in Exhibit A.2); (2) in-person and virtual training of administrators, school teams, and MTSS-B coaches to introduce and review MTSS-B practices; and (3) in-person and virtual technical assistance to monitor each school's implementation and provide individualized assistance. Exhibits A.3 and A.4 summarize the key events and their intended purpose as well as the intended participants and duration of each event in each of the two years that the program was implemented. The training and technical assistance materials used for the study have been

archived (see CSBS), and the Midwest Positive Behavioral Interventions and Supports (PBIS) Network regularly updates these materials.[5]

With some minor exceptions noted below, CSBS followed the training and technical assistance plan described in Exhibits A.3 and A.4.

Exhibit A.3. Program Year 1 Training and Technical Assistance for School-Wide Practices (Tier I)

Phase of Training & Technical Assistance	Training/ Technical Assistance Event Name	Goals of the Event	Expected Participants	Expected Duration
Readiness phase (spring prior to Program Year 1)	Introduction to readiness Part 1 (webinar)	Establish infrastructure for school-wide practices (Tier I) by identifying school team and MTSS-B coach and installing data system	Principal, Coach, District Coordinator	2 hours
	SWIS data system training (webinar)		Principal, Coach, District Coordinator	1.5 hours
	Introduction to readiness Part 2 (webinar)		Principal, Coach, District Coordinator	1 hour
	District readiness training (in-person)	Introduce data system and Team Initiated Problem Solving (TIPS) model for structuring school team meetings and using data to make decisions	Principal, Coach, District Coordinator	Approx. 12 hours (2 days)
In-person training for school-wide practices (summer prior to Program Year 1)	Leadership training for school-wide practices (Tier I)	Build capacity to support school's implementation of Tier I practices, data system, TIPS, and fidelity tools	Principal, Coach, Team Leader for School-wide Practices (Tier I Team Leader), District Coordinator	6 hours (offered in one day)
	Team training for school-wide practices (Tier I)	Build fluency in Tier I practices; start preparing school's implementation plan and materials	Principal, Coach, School Team for School-wide Practices (Tier I Team), District Coordinator	11.5 hours (offered across two days)
	Team planning for school-wide practices (Tier I)[a]	Complete school implementation plan for Tier I and materials for school kick-off meeting	Principal, Coach, Tier I Team, District Coordinator	Approx. 14 hours (2 days)
	All-staff kick-off for school-wide practices (Tier I)	Introduce MTSS-B to all school staff	All school staff members	3-4 hours in the morning
	School-level breakout session (same day as after all-staff kick-off)	Review elements of MTSS-B presented in the morning all-staff kick-off and discuss school-level plan for MTSS-B implementation and rollout on first day of school	All school staff members	3-4 hours in the afternoon

(continued)

Exhibit A.3 (continued)

Phase of Training & Technical Assistance	Training/Technical Assistance Event Name	Goals of the Event	Expected Participants	Expected Duration
Virtual training in school-wide practices[b] (throughout Program Year 1)	Expectations and rules	Introduce classroom management practices and how to train, monitor, and support classroom teachers in the application of these practices	Coach, Tier I Team Leader	1.5 hours
	Positive behavior game		Coach, Tier I Team Leader	1.5 hours
	Routines and procedures		Coach, Tier I Team Leader	1.5 hours
	Behavior-specific praise		Coach, Tier I Team Leader	1.5 hours
	Error correction		Coach, Tier I Team Leader	1.5 hours
	Opportunities to respond		Coach, Tier I Team Leader	1.5 hours
	Active supervision		Coach, Tier I Team Leader	1.5 hours
	Physical environment		Coach, Tier I Team Leader	1.5 hours
Virtual and in-person technical assistance in school-wide practices (throughout Program Year 1)	Site visits to each school	Monitor implementation and provide individualized school-specific technical assistance using the Tiered Fidelity Inventory	Principal, Coach, Tier I Team	4.5-6.75 hours for each visit
	Individualized support for MTSS-B coaches[c]	Provide refresher training and as-needed technical assistance to the MTSS-B coach to address implementation challenges and develop responses	Coach	30-60 minutes per support session (2 times a month)

SOURCE: Team summary of CSBS training plan.

NOTES: SWIS is the School-Wide Information System. The district coordinator is a person assigned by the district to work with the study team and foster MTSS-B study implementation in the district. The coach is a district-based MTSS-B coach.

[a]Team planning for school-wide practices (Tier I) was done by the school team in-person with virtual technical assistance and review provided by CSBS

[b]Classroom component training sessions were planned as webinars except for Expectations and rules, which was delivered as part of Tier I in-person training, and the Positive behavior game, which was delivered during a site visit.

[c]Coach support can be provided in the form of face-to-face meetings during site visits, webinars, or phone calls.

Exhibit A.4. Program Year 2 Training and Technical Assistance for School-Wide Practices (Tier I) and Supplemental Supports (Tier II)

Phase of Training & Technical Assistance	Training/Technical Assistance Events	Goals of the Event	Expected Participants	Expected Duration
Readiness phase (spring prior to Program Year 2)	Tier II readiness (on-site or through webinar)	Establish infrastructure for supplemental supports (Tier II) by identifying school team for supplemental supports (Tier II Team) and installing data system for Check In Check Out (CICO-SWIS)	Principal, Coach, Team Leader for Supplemental Supports (Tier II), District Coordinator	Approx. 2 hours
In-person training for school-wide practices (Tier I) & supplemental supports (Tier II) (summer prior to Program Year 2)	Tier I leadership booster training	Review school-wide (Tier I) concepts, adjust plan for school-wide implementation to address challenges, and review data on student behavior	Principal, Coach, Team Leader for School-Wide Practices (Tier I Team Leader), District Coordinator	6 hours
	Tier I team booster training	Improve implementation of school-wide practices (Tier I) identified as needing improvement during school year	Principal, Coach, Team for School-Wide Practices, District Coordinator	12 hours (2 days)
	Leadership training for supplemental supports (Tier II)	Build capacity to support school's implementation of Tier II practices and start preparation of school's implementation plan and materials	Principal, Coach, Team Leader for Supplemental Supports (Tier II Team Leader), District Coordinator	6 hours
	Team training for supplemental supports (Tier II)	Introduce Tier II practices and initial plan for implementation	Principal, Coach, Team for Supplemental Supports (Tier II Team), District Coordinator	10 hours (2 days)
	Tier I and Tier II team planning[a]	Complete the school's implementation plan and materials for school-wide kick-off meeting	Team for School-wide Practices (Tier I) and Team for Supplemental Supports (Tier II), Coach	Approx. 14 hours per team (2 days)

(continued)

Exhibit A.4 (continued)

Phase of Training & Technical Assistance	Training/Technical Assistance Events	Goals of the Event	Expected Participants	Expected Duration
Virtual training and technical assistance (throughout Program Year 2)[b]	Tier II Team facilitation	Introduce strategies to facilitate team meetings and plan for implementation of supplemental supports	Coach, Tiers I and II Team Leaders	1.5 hours
	Webinars	Refresh understanding of school-wide practices and supplemental supports; address implementation issues as needed	Coach, Tiers I and II Team Leaders	1.5 hours (8 webinars across the year)
Virtual and in-person technical assistance in supplemental supports (throughout Program Year 2)	Site visits to each school	Monitor school's implementation and provide individualized school-specific technical assistance using the Tiered Fidelity Inventory	Principal, Coach, Tier I and II teams	5.5-7.5 hours for each visit (four times a year)
	Individualized support for MTSS-B coaches[c]	Provide refresher training and as-needed technical assistance to the MTSS-B coach to address implementation challenges and develop responses	Coach	30-60 minutes (twice a month)

SOURCE: Author's compilation based on CSBS training plan.

NOTES: CICO-SWIS is the School-Wide Information System (SWIS) application for Check In Check Out (CICO).
[a]Team planning for Tier I and Tier II done by the school team in-person with virtual technical assistance and review provided by CSBS.
[b]The content of the technical assistance events/webinars was determined together with MTSS-B coaches to ensure that the ongoing supports provided by CSBS met schools' needs.
[c]Coach support can be in the form of face-to-face meetings during site visits, webinars, or phone calls.

Readiness Phase: Training and Technical Assistance by CSBS to Develop Infrastructure

The readiness phase for each tier took place as intended during the spring preceding each program year and prior to the start of the in-person training. Through the readiness activities in Program Year 1 for Tier I, CSBS trainers worked with all districts and schools to establish all three components of the MTSS-B infrastructure described in Exhibit A.2. Each of the 58 participating schools established a school team to lead implementation of the school-wide practices, installed the PBIS Apps data system, and was assigned an MTSS-B coach.

As part of the Tier II readiness activities in Program Year 2, 57 of the 58 participating schools established a team to lead the implementation of the supplemental support intervention and 56 of the participating schools initiated use of the data system application to support the supplemental support intervention (SWIS-CICO). During Program Year 2, one school withdrew from the MTSS-B training and technical assistance program but continued to allow most data collection activities to support the evaluation. Additionally, another participating school continued in the MTSS-B program but suspended use of the data system.

In-Person and Virtual Training and Technical Assistance by CSBS to Support Schools in MTSS-B Implementation

In Program Year 1, CSBS followed the training and technical assistance plan described in Exhibit A.3. In Program Year 2, CSBS mostly followed the training and technical assistance plan described in Exhibit A.4, but some training events in some districts changed from the intended format and few districts participated in all intended virtual technical assistance webinars. These changes were made to accommodate districts' schedules and preferences. In three districts, the Program Year 2 in-person booster content training for the school-wide practices (Tier I) was either shortened from the intended duration of three days or partially transferred from an in-person format to a webinar format. CSBS was prepared to offer the eight planned training and technical assistance webinars to each district but because of scheduling conflicts and requests for cancellations in some districts, CSBS ultimately hosted between five and nine webinars per district during Program Year 2.

Exhibits A.5 and A.6 show duration and attendance figures for the training and technical assistance events that the study team observed compared with what was expected. Exhibit A.5 shows that observed events generally tended to be shorter than intended. Exhibit A.6 shows that, on average, the majority of participating schools' expected participants attended the observed events in both program years. Attendance at observed events generally dipped in Program Year 2. See Appendix C for findings regarding staff perceptions of the quality of the CSBS training and technical assistance events.

Exhibit A.5. Mean vs. Intended Duration of MTSS-B Training and Technical Assistance (TA) Events (Hours of Attendance)

	Year 1			Year 2		
Type of Training	Mean Duration	Intended Duration	Standard Deviation	Mean Duration	Intended Duration	Standard Deviation
In-person training for Tier I[a]	16.6	17.5	1.1	13.9	18.0	3.2
In-person training for Tier II[b]	NA	NA	NA	14.6	16.0	1.8
Virtual training and TA[c]	4.5	9.0	0.9	6.7	12.0	2.3
In-person TA[d]	4.6	4.5-6.75	1.0	4.4	5.5-7.5	0.9

SOURCE: Authors' calculations based on observed trainings.

NOTES: For in-person Tier I trainings, there were ten in Year 1 and nine in Year 2. For in-person Tier II trainings, there were nine in Year 2. Three districts in Year 2 held at least one of their training events as a webinar instead of in-person training.
  [a] This event includes three days of training. The first day was leadership training and the second and third days were full team training. In district D9, the training days were presented twice to different sets of participating schools.
  [b] This event includes three days of training. The first was intended to be leadership training and the second and third days were full team training.
  [c] Mean duration of Year 1 webinars includes the duration of six webinars. Mean duration of Year 2 virtual training and TA sessions include the duration of up to eight webinars. However, one district had nine webinars in Year 2. If the ninth Year 2 webinar in this district is included in the analysis, the mean Year 2 duration is 6.8 hours and the standard deviation is 2.4.

(continued)

Exhibit A.5 (continued)

[d] In-person technical assistance durations were the average site visit durations of each school for which site visit duration was documented. In Year 1, 97 site visits in 57 schools were documented (site visits to one school were not documented), and in Year 2, 112 site visits in 57 schools were documented (there were 57 participating schools in Year 2).

NA: Not applicable.

Exhibit A.6. Intended School Staff Attendance at MTSS-B Training and Technical Assistance (TA) Events

Staff in Attendance	Year 1	Standard Deviation	Year 2	Standard Deviation
Mean percentage of intended participants in attendance				
In-person training for Tier I	87.0	17.4	75.4	23.5
In-person training for Tier II	NA	NA	79.7	24.7
Virtual training and TA	71.0	30.5	69.6	33.2
In-person TA[a]	83.5	13.0	77.5	13.6
Percentage of schools with administrator in attendance				
In-person training for Tier I	89.7		77.2	35.2
In-person training for Tier II	NA		86.0	28.1

SOURCES: Author's calculations based on training attendance; n = 58 schools for Program Year 1 and n = 57 schools for Program Year 2. Attendance collected at all offered in-person training events and virtual training and TA events. Attendance collected at observed in-person TA events (Program Year 1: n = 96 individual school visits; Program Year 2: n = 116 individual school visits).

NOTES: Weighted attendance scores are based on individuals or groups of attendees who were intended to participate in specific events. The individuals expected to attend an event vary by event. Average weighted attendance scores are the average attendance scores for all events in a training event cluster that occurred. School average scores were averaged across all MTSS-B participating schools to obtain the mean scores in this table.

[a] Statistics for in-person TA events (school site visits) are based on school averages. Observed site visits per school in each year ranged from 0 to 4; statistics are based on 97 observed Program Year 1 visits to 57 schools and 112 observed Program Year 2 visits to 57 schools.

NA: Not applicable.

III. MTSS-B Coaching Support

Each district worked with CSBS during the Program Year 1 readiness phase to identify and interview prospective MTSS-B coaches. Coaches were expected to devote one day a week to each participating school. Across the nine districts, 15 individuals were hired to serve in the coaching role. Over the course of the two program years, five replacement coaches were hired to account for family leave and attrition.

Exhibit A.7 provides an overview of the MTSS-B coach qualifications based on the study team's reviews of the resumes of the 20 individuals hired over the two program years, including the 15 originally hired coaches and the 5 replacement coaches. Most coaches who were hired had at least a master's degree (90 percent), most had worked as a classroom teacher (90 percent), and about half had other prior relevant experience to the MTSS-B coach role (55 percent), including six individuals (30 percent) who had previous experience as a coach.

Exhibit A.7. MTSS-B Coach Characteristics

Characteristics	Number of Coaches	Percentage	Mean Years[a]
Gender			
Female	16	80.0	NA
Male	4	20.0	NA
Highest level of education			
Master's degree or above	18	90.0	NA
Experience			
Classroom or special education teacher	18	90.0	10.7
Other prior relevant experience[b]	11	55.0	10.9
MTSS-B/PBIS/academic/behavior coach	6	30.0	5.2

SOURCE: Based on information from a review of coach resumes.

NOTES: The table includes the information of 20 coaches, of which 15 were originally hired and 5 were replacement coaches.

[a] Only cases with values larger than zero are included in the calculation of the mean years of experience.

[b] Other prior experience includes working experience as a coach (academic, behavioral, or MTSS/PBIS/RTI), school counselor, school psychologist, school social worker, or an administrator.

NA: Not applicable.

The MTSS-B coaches were expected to spend one day per week supporting the school teams and teachers in each of the participating schools on the implementation of MTSS-B practices. Coaches completed activity logs documenting how they spent their time with each school. The logs show that few participating schools received the full expected amount of coaching support.

Exhibit A.8 shows that, on average, coaching activities took place in participating schools in more than 70 percent of instructional weeks in both years. An "instructional week" is defined as Monday to Friday, during which time a school was open to students three days or more. A review of district calendars from each program year shows that, on average, districts had about 35 instructional weeks in Program Year 1 and 38 instructional weeks in Program Year 2. Assuming that a typical school day is six hours, those findings mean that coaches were expected to spend 212 hours in each school in Program Year 1 and 225 hours in each school in Program Year 2. On average, however, MTSS-B coaches spent 151.7 hours per participating school in Program Year 1 and 165.7 hours per participating school in Program Year 2. Coaches provided fewer hours to participating schools than expected as a result of coach attrition, professional development conflicts in their schedules, and illness. It is also possible that some coaches completed activities with schools but did not record those activities in the study's coaching logs.

Exhibit A.8 shows that in both program years, MTSS-B coaches spent the least amount of their time providing direct training to school staff members and the most time providing coaching support to teachers. In both years, coaches also spent a considerable amount of time on "other" activities. These could include both activities that align with the intended coach role (such as supporting program logistics, attending meetings, and working with data) and those that do not (such as working directly with students).

Exhibit A.8. MTSS-B Coach Activities in Participating Schools

Coach Activities	Program Year 1 Mean	Program Year 1 Standard Deviation	Program Year 2 Mean	Program Year 2 Standard Deviation
Average percent of instructional weeks with coaching activity in participating schools[a]	79.1	14.8	72.9	18.4
Average total hours of activity per school	151.7	65.7	165.7	76.7
Percentage of hours spent in training activities	10.2	11.1	6.5	7.1
Percentage of hours spent in coaching activities	48.9	20.8	63.6	22.0
Percentage of hours spent in "other" activities	41.0	22.3	30.0	24.2
Total number of schools	58		57	

SOURCE: Author's calculations based on MTSS-B coach activity logs.

NOTES: Coach activity logs documented dates, hours, and activities conducted in each participating school. Values in this table reflect means across 58 participating schools in Program Year 1 and 57 participating schools in Program Year 2.

   [a]Across the nine districts, the average number of instructional weeks was 35.3 in Program Year 1 and 37.5 in Program Year 2. An instructional week is defined as a week in which the school is open for three or more days.

The CSBS training and technical assistance did not provide specific guidance on how coaches should allocate their time, which could explain the patterns observed in the activity logs, including the considerable variance between schools in the types of coaching support they received.

IV. Additional Implementation Support Provided by the Study

To increase the chances that the training and technical assistance program would run smoothly in the participating districts, the study's implementation support team, led by the American Institutes for Research and overseen by MDRC, communicated twice a month with districts and CSBS staff to carefully monitor all program activities. The team:

1. coordinated district-based MTSS-B coach interviews by district personnel and the CSBS staff;

2. reviewed training and technical assistance materials for completeness, alignment with the CSBS training plan, and quality of presentation;

3. coordinated scheduling of training and technical assistance events between district personnel and CSBS staff;

4. monitored delivery of the intended content of group training events and select site visits and webinars;

5. monitored the frequency of coach activities and team meetings;

6. developed processes to collect data on school staff member participation in training, technical assistance, and implementation activities hosted outside school hours to enable the study to compensate the staff for their time; and

7. coordinated meetings with CSBS and district personnel to review data on school participation, program content, and plans for upcoming activities in order to gather feedback from the districts and to problem-solve as needed.

V. Cost of the Study's Training and Technical Assistance Program

The per-school cost of MTSS-B training and technical assistance is driven, in part, by the number of schools participating in a district. During many training events, including the in-person training that takes place over the summer, participating schools are brought together and trained at the same time and place. When more schools participate in these events, the per-school cost goes down. For a district with six schools implementing MTSS-B, CSBS estimated their trainers needed a total of 35 training and technical assistance days in Program Year 1 and 36 in Program Year 2 and a daily rate of $750 per trainer per day. These assumptions translate to a cost of $4,375-$4,500 per school per year in a district with six schools. Coaching costs for this study varied depending on the salaries of specific coaches hired in each district. Based on the study team's records of coach salaries reimbursed by the study, the average cost of a coach to spend one day per week in each school (20 percent full-time equivalent, or FTE) was approximately $20,000 per school per year. Recent work examining MTSS-B implementation in other settings has considered costs beyond training and coaching salaries to include the staff time for training and participation in school team activities and other resources necessary to support implementation.[6]

APPENDIX B. STUDY DESIGN, DATA COLLECTION, AND ANALYTIC APPROACHES

This appendix describes the study design, the recruitment process, and random assignment. It then introduces the data-collection activities and the resulting analytic samples. Lastly, the appendix presents the approaches used for the impact estimation and the exploratory analyses.

I. Study Design

This study used a randomized controlled trial design that randomly assigned schools to participate in the program of training and technical assistance or to continue with their usual strategies for supporting student behavior. This section describes how the team carried out this design through the recruitment and random assignment of schools.

Recruitment and Selection of Sample Schools

The study team recruited 89 elementary schools in nine school districts to participate in the study. To arrive at this sample, the study team developed a set of eligibility guidelines summarized in Exhibit B.1.

Exhibit B.1. Screening Guidelines Used for Recruitment

Screening was conducted at both the district and school level. Here are the guidelines used for the screening process.

District Screening Guidelines

A district had to have at least ten public schools serving students in kindergarten through Grade 5 or Grade 6 that met the following criteria:

- School size: At least 50 students per grade, on average.
- Family income: 25 percent of students qualifying for free and reduced-price lunch or eligible for Title 1 (according to 2011-2012 Common Core of Data or district data).
- Current behavior support strategies: Not implementing MTSS-B Tier I or Tier II programs at the time of recruitment.
- Interest: Interested in implementing the program in nominated schools. The district had to understand fully the random assignment process and the burden of data collection.

School Screening Guidelines

Within an eligible district, a school was considered for nomination if it met the following criteria:

- School size: Serving at least 40 students in each grade (according to 2011-2012 Common Core of Data or district data). Across all nominated schools in a district, there had to be an average of at least 60 students per grade.

(continued)

Exhibit B.1 (continued)

- Family income: There was a preference for schools that had at least 15 percent of students qualifying for free and reduced-price lunch (basing this screening criterion on 2011-2012 Common Core of Data or district data). Across all nominated schools in a district, the percentage of free and reduced-price lunch students had to be greater than 25 percent.*

- Current behavior support practices: Not implementing MTSS-B Tier I and Tier II programs at the time of recruitment.

- Interest: Interested in implementing the program if selected to participate and willing to continue with business as usual if selected as a non-participating school. Schools had to understand the random assignment process and the burden of data collection.

Note: *The study team only accepted schools below 15 percent if they were needed to get the minimum sample size for a district.

Eligibility criteria for site recruitment emphasized the districts' and schools' sizes and demographic profiles. Because of the costs associated with both training and data collection and to avoid losing a district from the impact estimate if a school dropped out, it was important to have a sufficient number of potentially eligible schools within each district. There was a priority for districts and schools serving sizable proportions of children living in low-income households, reflecting the Department of Education's focus on schools serving disadvantaged students.

The recruitment process began with the identification of eligible districts based on the 2011-2012 Common Core of Data, the most recent data available at the time of recruitment. Five hundred and twenty school districts met the school size and demographic criteria. In consultation with the Department of Education and experts in the field knowledgeable about existing MTSS-B implementation, the study team further refined the list of eligible districts to exclude states and districts where it was known that MTSS-B implementation was already widespread (for example, most districts in Maryland and Illinois were excluded because statewide efforts have had a significant uptake).

The study team contacted approximately 350 school districts with an introductory letter about the project and an invitation to participate in an initial screening call. If districts expressed interest in participation, the team conducted screening calls with district officials responsible for student behavior support services to (1) assess existing behavioral support practices in the district and (2) provide information about the research design, which required random assignment of schools and study-related data collection. The screening calls followed a structured protocol. Subsequent follow-up with districts was prioritized to achieve geographic diversity in the sample. Specifically, the team conducted targeted follow-up in East coast and Midwest districts that did not have existing MTSS-B practices in place because the team initially had the most interest from districts in the West where MTSS-B has been less widespread. The study team visited 19 districts that met the district eligibility criteria and expressed interest in participating based on the initial screening calls. In a one- to two-day site visit, the team met with senior district leadership, research and data offices, and school principals to further assess district eligibility and to work with the district to identify eligible schools.

If a district was confirmed to be eligible, the team worked with the district contacts to assess the school eligibility criteria outlined in Exhibit B.1 and relied upon the Common Core of Data and/or district data to confirm school size and demographic criteria. Once a district and the team turned to determining whether a school already had too many elements of MTSS-B in place, the site recruitment team interviewed district and school staff using a rubric that included features such as the presence of any behavioral leadership team, prior training on behavior support programs like MTSS-B (both school-wide and targeted Tier II supports), presence and use of behavioral data systems, use of a rewards

system for student behavior, whether specific behavioral expectations were established and taught to students, consistent use of specific classroom management strategies, presence of a Tier II targeted program of behavioral supports, and referral systems for this Tier II program. On each feature, each school was rated as showing a strong, intermediate, or weak contrast with the MTSS-B program to be fielded in the study. Because MTSS-B has elements in common with other behavioral support approaches, it was deemed acceptable for schools to have *some* elements of MTSS-B in place. However, elements were not all treated as equally important and the team sought schools that had not defined a small number of positively defined expectations, taught them to students, and developed a system to reward meeting expectations and to support students in need of special assistance. It was not possible to field a formal "fidelity" assessment at the point of site recruitment, but when the SET was fielded for the first time at the end of Program Year 1, there was a substantial service contrast in MTSS-B elements between participating and non-participating schools. Only 0.6 percent of non-participating schools met the SET fidelity threshold for school-wide practices and only 15 percent met the threshold for supplemental supports. Finally, schools and districts had to be willing to engage in the study's data-collection activities and accept the results of the school-level random assignment to participate in MTSS-B.

Nine school districts and 89 schools were accepted into the study and signed a memorandum of understanding in the spring of 2015. These nine districts are in six states and have between 7 and 16 participating schools each. One school ceased participation in the study's training and technical assistance program in Program Year 2. This school continued with most data collection activities in that year so that they could be included in the impact analysis for Program Year 2. The team invited all nine districts and 89 schools that participated in the first two years of the study to participate in a follow-up study during the year after the study's MTSS-B training and technical assistance ended. Eight of the nine districts accepted the invitation and modified the memorandum of understanding to include the Follow-Up Year.

The study schools were different from the average elementary school in the nation in many respects (see Exhibit B.2). The study schools were more likely to be in urban areas and served more students. On average, the study schools served a more economically disadvantaged student population, with about 72 percent of students eligible for the free and reduced-price lunch program, compared with 55 percent in the national population. These differences reflect the study's focus on recruiting schools serving disadvantaged students. The study schools also served a higher proportion of Hispanic students and students with limited proficiency, which is likely driven by the demographic characteristics of the regions where the study had the greatest success in recruiting districts with limited prior experience in MTSS-B.

Exhibit B.2. School Background Characteristics for Study and National Samples

Characteristics	Study Sample	All Elementary Schools
Urbanicity (percentage of schools)		
Large or middle-sized city	43.8	22.5 *
Urban fringe and large town	50.6	43.6
Small town and rural area	5.6	33.9 *
Title I status (percentage of schools)	84.3	79.8
Students eligible for free or reduced-price lunch (average percentage of students per school)	71.7	54.9 *

(continued)

Exhibit B.2 (continued)

Characteristics	Study Sample	All Elementary Schools
Students' race/ethnicity (average percentage of students per school)		
White, non-Hispanic	19.4	50.0 *
Black, non-Hispanic	9.9	14.5
Hispanic	60.2	26.0 *
Asian	6.0	4.5
Other	4.4	5.0
Students with Individualized Education Programs (average percentage of students per school)	10.4	12.5 *
Students with limited English proficiency (average percentage of students per school)	38.1	13.2 *
Male students (average percentage of students per school)	51.7	51.6
Mean school enrollment (number of students)	618	477 *

SOURCE: 2014-2015 Common Core Data (CCD) and 2013-2014 Civil Rights Data Collection (CRDC) for school background characteristics.

NOTES: 'Study Sample' refers to the study sample schools (n = 89).

   'All Elementary' refers to all regular public elementary schools serving students in Grades K/1-5/6 (n = 72,487).

   *Indicates a statistically significant difference (p-value < 0.05) between the study sample and the schools in the national population for a given characteristic. A two-tailed t-test was applied to each comparison.

   Joint F test comparing all characteristics between the study sample and all schools in the national population is significant at a 95 percent confidence level (p=0.011).

   Sample sizes for individual characteristics may vary due to missing values.

Random Assignment

In the spring of 2015, the study team randomly assigned two-thirds of the recruited study schools to participate in the program of training and technical assistance, and one-third to continue with business as usual. This two-to-one ratio was selected so that districts could have more of their schools trained in MTSS-B and for the study to get a better sense of how the training and implementation would play out in a variety of schools. The purpose of the random assignment was to create two groups of schools similar to each other before the program's start. That way, all subsequent differences in outcomes between these two groups could be attributed to the program.

The study team conducted random assignments within each district. However, if the selected schools in a district varied a lot in terms of the proportion of students eligible for free or reduced-price lunch, the team stratified the schools in the district based on this proportion and did random assignment within these strata, also called random assignment blocks. This strategy helped to ensure that the participating schools served similar proportions of disadvantaged students as the non-participating schools. By the end of the random assignment process, the team had randomly assigned 89 study schools in 15 blocks, with 58 schools in the participating group and 31 schools in the non-participating group. These two groups were similar to each other in terms of school size and the kinds of students they served during the year before the program began, as shown in Exhibit B.3.

Exhibit B.3. Background Characteristics Comparison of Participating and Non-Participating Schools in the Study

Characteristic	Participating Schools	Non-Participating Schools	Estimated Difference	p-value for Estimated Difference
Title I status (percentage of schools)	81.0	89.1	-8.0	0.121
Students eligible for free or reduced-price lunch (average percentage of students per school)	70.4	74.3	-3.9	0.125
Student race/ethnicity (average percentage of students per school)				
White, non-Hispanic	20.3	17.9	2.3	0.232
Black, non-Hispanic	9.6	11.1	-1.5	0.332
Hispanic	59.2	61.7	-2.6	0.170
Asian	6.4	5.0	1.3	0.133
Other	4.6	4.2	0.4	0.294
Students with Individualized Education Programs (average percentage of students per school)	10.4	10.3	0.0	0.950
Students with limited English proficiency (average percentage of students per school)	37.9	38.0	-0.2	0.937
Male students (average percentage of students per school)	51.7	51.8	-0.1	0.868
School enrollment (number of students)	614	634	-19	0.557

SOURCE: 2014-2015 Common Core Data (CCD) and 2013-2014 Civil Rights Data Collection (CRDC) for school background characteristics.

NOTES: There are 58 participating schools and 31 non-participating schools in the study.

The estimated differences for school-level data are regression-adjusted using ordinary least squares (OLS) regressions, controlling for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The non-participating schools' values in the next column are the difference between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating means and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

An omnibus test was used to determine whether there is a systematic difference between the participating schools and the non-participating schools, with respect to the characteristics included in this table. The p-value for this test is 0.982.

II. Data-Collection Activities

The study team carried out multiple data-collection activities during the two program years and the follow-up year. This section describes the primary data collection activities and the instruments used for these activities. Section III provides an overview of the measures constructed from these data sources.

Exhibit B.4 summarizes the main data sources, including the data obtained from them, their collection times, and the unit of measure for each source. Details of each data source are provided below.

Exhibit B.4. Data Collection Activities

Data Source	Data Obtained	Time Data Collected	Unit of Measure (Respondent)
Data to measure effects on students			
Teacher Survey of Student Behavior (TSSB)	Teacher ratings of student behaviors (Grades 1-5)	Fall of Program Year 1 Spring of Program Year 1 Spring of Program Year 2	Students (reported by classroom teachers) in participating and non-participating schools
District records	Student achievement (Grades 3-5); student background characteristics (Grades 1-5)	Fall of Program Year 1[a] Fall of Program Year 2[a] Fall of Follow-up Year[a] Fall of the school year after Follow-up Year[a]	Participating and non-participating schools; Students in participating and non-participating schools
Data to measure effects on classrooms and schools			
Classroom observations	Ratings of classroom management practices and classroom functioning	Spring of Program Year 2	Classrooms (study observer) in participating and non-participating schools
Staff survey	Staff member[a] perceptions of school climate	Spring of Program Year 2	Staff members[b] in participating and non-participating schools
Data to measure implementation in schools			
Site visits	Measures of implementation of Tier I and Tier II practices	Fall of Program Year 1 Spring of Program Year 2 Spring of Follow-up Year	Participating and non-participating schools (study data collection assessor)
Training and technical assistance observation checklists	Duration and content of training and technical assistance events	Ongoing throughout Program Years 1 and 2[c]	Training event (study observer) in participating schools only
Training participation forms	Training event participation for schools and coaches	Ongoing throughout Program Years 1 and 2[c]	MTSS-B coaches (study observers) in participating schools only
Reviews of program data[d]	Fidelity of program infrastructure core component implementation	Throughout Program Years 1 and 2	MTSS-B coaches (school staff members and coaches) in participating schools only
Staff interviews	Staff perception of training and implementation quality	Spring of Program Year 1 Spring of Program Year 2	Select participating school staff[e]
District coordinator interviews	District coordinator perception of Follow-Up Year implementation quality	Fall and spring of Follow-Up Year	District coordinator

(continued)

Exhibit B.4 (continued)

NOTES: The 2015-2016 school year was Program Year 1, the 2016-2017 school year was Program Year 2, and the 2017-2018 school year was the Follow-Up Year.

[a]This reflects data collection time, not the time of the measurement. Student achievement data for a given school year were only available in the following fall and therefore were collected at that time.

[b]Surveyed staff members included instructional and non-instructional staff in the school who were likely to have regular interactions with students.

[c]Checklists and forms were collected at each in-person group-based training event, including summer training and webinars, and from select technical assistance site visits conducted at least once per participating school per program year.

[d]Data included logs of team meeting minutes; logs of coach activity; and behavior monitoring data from the School-Wide Information Survey (SWIS) and School-Wide Information Survey-Check In Check Out (SWIS-CICO).

[e]The selected staff included Tier I leader and principal in Program Year 1 and Tier I and II team leaders (not including principals) in Program Year 2 from participating schools.

Teacher Survey of Student Behavior

The study asked teachers of students in Grades 1-5 in all study schools to complete a commonly used survey, described below, to rate the behavior of each of their students for whom the study team had parental consent. See Exhibit B.5a for consent rates for the collection of teacher ratings of student behavior for the full sample and by program conditions. Teachers provided these ratings at the beginning of Program Year 1 and then again in the springs of Program Years 1 and 2. Teachers were provided a $5 incentive payment for each child rated in each wave of data collection.

District Records Data Collection

The study team collected district records data for individual students to obtain information on their characteristics in the year before the start of the program, during each program year, and during the follow-up year. One study district provided records data for students with active parental consent only. All other study districts provided records data for all students enrolled in the relevant grade levels in the study schools. For students in Grades 3-5, the study team also collected information on their academic achievement measured by state standardized test scores for reading and math. The team used student demographic information collected from the year before the start of the program as covariates in impact estimations and used students' state reading and math test scores collected for the two program years and the follow-up year as outcomes for student academic achievement.

Classroom Observations

The University of Virginia trained study observers to conduct systematic classroom observations in all study schools in the spring of Program Year 2. All observers were required to demonstrate 80 percent inter-rater reliability with a master coder from the University of Virginia before they were sent into the field. Observers were not informed of the school's research status in advance. They were given a schedule of classrooms/teachers to visit. No incentives were provided to teachers for the classroom observations. The field managers reviewed the protocols within 48 hours of each site visit, and the University of Virginia team provided weekly technical assistance to the assessors throughout the fielding period. Observers used the Assessing School Settings: Interactions of Students and Teachers (ASSIST), discussed further below.[7] The study team used the tallies and global ratings collected from these classroom observations to measure how teachers managed their classrooms and to document teachers' and students' behaviors in the classrooms. The sample of classrooms includes all Grade 1-5 general education and inclusive classrooms in the study schools. The team conducted classroom observations once in each school (in around 20 classrooms per school, on average). The choice of this fielding strategy was based on the existing literature that shows that fielding the ASSIST observation protocol in many classrooms per school can reliably discriminate between schools in their use of behavior support practices, even if the protocol is only fielded once.[8]

Staff Survey

School staff members in all study schools–including teachers, administrators, and other non-instructional staff members–participated in a survey about their receipt of professional development in behavioral support practices; their schools' use of school-wide, classroom, and supplemental systems of support for behavior; and their schools' climate. The following categories of staff were included in the sample:

1. Classroom teacher (e.g., you teach only one group of students who spend most of the day with you)

2. Subject-specific classroom teacher (e.g., art, science, physical education)

3. Principal

4. Other administrator (e.g., Dean or assistant principal)

5. Administrative assistant

6. Special education teacher (e.g., lead teacher for self-contained or co-taught special education class)

7. Other instructional staff (e.g., librarian, specialist, etc.)

8. Student support staff (e.g., school counselor, paraprofessional, school psychologist, nurse, security personnel)

9. Custodial or food service staff (e.g., cafeteria or maintenance staff)

10. Other (Please Specify): ___

For staff in categories 1, 2, and 6, there were specific questions about teaching experience and classroom practices and interaction with students. Staff in all categories completed sections on school-wide behavioral practices and school climate. The survey also included several questions related to individual staff members' background characteristics and experiences. It was conducted in the spring of Program Year 2. Staff completing the survey were provided a $25 gift incentive.

Site Visits

The team visited study schools three times over the study period. The purpose of these visits was to collect data to describe the contrasts in school-wide practices (Tier I) and supplemental supports (Tier II) between the participating and non-participating schools at different stages of the study. Site visitors were not informed of the school's research status in advance of the visits. They were given a list of school staff to meet with including the name of the behavioral support person. The site visit activities included document reviews, observations, and interviews. The study team combined the School-wide Evaluation Tool (SET)[9] and an adapted version of the Individual Student Systems Evaluation Tool (IS-SET)[10] to create the site-visit protocol used to describe the school-wide and supplemental support practices in the participating and non-participating schools. See below for more details on these measures.

Exhibit B.5a presents the response rates for the data sources used in the program impact analysis. This exhibit shows that the low overall response rates and the differential response rates between the participating and non-participating schools for the Teacher Survey of Student Behavior (TSSB) were largely driven by the parental consent rates in all three rounds of collections.[11] Among students with parental consent, the response rates were generally above 80 percent and did not differ by program status. Appendix D provides evidence that this differential overall response pattern was unlikely to bias the impact estimates on student behavior outcomes.

Exhibit B.5a. Response Rates for Data Sources Used in Impact Estimation

Data Collected	Response Rates (%)				
	Full Sample	Participating Schools	Non-Participating Schools	Estimated Difference	p-value of Estimated Difference
Classroom observation	91.8	91.6	90.7	0.8	0.525
Staff survey	61.9	62.9	60.2	2.7	0.158
Teacher Survey of Student Behavior					
Fall 2015					
Consent rate	68.3	71.2	62.5	8.6	0.000 *
Response rate of consented students	83.1	83.4	81.9	1.6	0.549
Overall response rate	57.4	59.8	52.4	7.4	0.018 *
Spring 2016					
Consent rate	71.5	73.4	67.8	5.6	0.001 *
Response rate of consented students	81.6	82.3	79.3	3.0	0.266
Overall response rate	58.9	60.9	54.4	6.6	0.010 *
Spring 2017					
Consent rate	70.8	72.0	67.9	4.0	0.018 *
Response rate of consented students	83.9	85.0	81.4	3.6	0.129
Overall response rate	59.9	61.7	55.5	6.2	0.010 *
Student reading and math achievement from state test					
Program Year 1	99.5	99.5	99.3	0.2	0.257
Program Year 2	99.3	99.3	99.4	-0.1	0.767
Follow-Up Year	99.3	99.4	99.2	0.2	0.344

SOURCE: Authors' calculations based on student- and teacher/classroom-level data collected and compiled by the study team.

NOTES: This table is based on the 58 participating schools and 31 non-participating schools in the study. The estimated differences are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating means and differences.

  *Indicates that the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

In addition to the response rate, the team also checked the response time of teacher ratings for the three rounds of TSSB collections. Exhibit B.5b shows that there was no difference in teachers' submission time of their ratings between the participating and non-participating groups.

Exhibit B.5b. Teacher Ratings of Student Behavior Submission Time Check, by Collection Round

	Submission Time (# of Days Elapsed)[a]			
Data Collection Round	**Participating Schools**	**Non-Participating Schools**	**Estimated Difference**	**p-value of Estimated Difference**
Fall 2015	123.38	126.30	-2.92	0.069
Spring 2016	130.12	130.90	-0.78	0.421
Spring 2017	84.36	86.19	-1.83	0.140

SOURCE: Teacher Survey of Student Behavior (TSSB) collected in fall 2015, spring 2016, and spring 2017.

NOTES: This table is based on the 58 participating schools and 31 non-participating schools in the study. The estimated differences are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating means and differences.

[a]Submission time is measured by the number of days that have elapsed after a fixed time in the year. For the fall 2015 round, the fixed day was August 15, 2015; for the spring 2016 and spring 2017 rounds, the fixed date was January 1st of each year. The average date of submission for participating schools was December 16, 2015 for the fall 2015 round, May 10, 2016 for the spring 2016 round, and March 26, 2017 for the spring 2017 round.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

III. Analytic Approaches

This section presents the analytic approaches the team used to estimate the effects of the program. It starts by describing the impact analysis outcome measures, their construction, and assessments of their reliabilities. It then describes the analytic models and samples used for the impact estimation and approaches used for exploring the relationships among program implementation features, the program's effects on school and classroom outcomes, and its effects on student outcomes.

Constructing Outcomes

This section describes the construction of outcome measures examined in this evaluation. These outcome measures include assessments of student academic achievement, student behavior, classroom functioning, classroom management practices, school climate, and implementation of school-wide practices and supplemental supports.

Student academic achievement. The study measured student academic achievement using students' test scores on state standardized English language arts and math tests. The team standardized the scores from different state tests by converting them to z-scores: within each district and for a given subject, year, and grade, the team subtracted the non-participating group mean from each student's test score and divided the result by the standard deviation of the non-participating group. This standardization allowed the impact analysis to pool data across districts with different state tests.

Teacher ratings of student behavior. The team adopted existing constructs in the Teacher Observation of Classroom Adaptation-Checklist (TOCA-C) to measure student behavior. Prior studies have demonstrated that these are valid and reliable measures of students' behavior for the purpose for which they are used in this study and the interpretations the study is making,[12] and past evaluations of MTSS-B have used this instrument and have shown it to be sensitive to the intervention.[13] The study team re-named some of the TOCA-C's constructs for ease of interpretation in this report, but items used on the survey instrument are consistent with prior studies. The specific student behavior constructs measured include disruptive behavior (aggressive/disruptive behavior in the original), attention to schoolwork

(concentration problems in the original), pro-social behavior, internalizing behavior, and emotional dysregulation (emotional regulation problems in the original).

Teacher ratings of behavior could be affected by teacher biases. In one recent literature review on this by Mason and others[14] the authors discussed the ways in which bias could affect ratings and reviewed the existing literature, finding mixed evidence of the existence of bias due to student ethnicity and stronger evidence of the existence of bias due to differences in the cultural background of students and teachers. This is an important issue but the way in which teacher ratings are used in this study does not raise bias issues present in other uses. First, the study does not compare ratings of individual students of one ethnic or cultural group with those of another group. For example, it does not compare ratings of students who are white with ratings of students of color. The study does compare ratings for groups of students who are initially identified as struggling with behavior, but the comparison is between students identified as part of this subgroup in similar ways across the participating and non-participating schools. Second, the focus is on differences in behavior between two large groups at a point in time rather than comparisons of individual students. The participating and non-participating groups were similar in their students' race/ethnicity composition, as demonstrated in the baseline equivalence tables in Appendix B. To the extent there is bias, it will affect the two groups of schools similarly. Third, the findings presented based on classroom observations by trained observers (which the literature suggests is a more accurate measure) yield similar results to those on teacher ratings: a reduction in disruptive behavior in the classroom. Another consideration is that students' educational experiences are affected by their teacher's perceptions of their behavior–whether or not this perception is biased–so a finding that MTSS-B does lead to teacher perceptions that disruptive behavior has lessened can signal an improved outcome for students, whether or not the measure is an objective measure of behavior.

Classroom management practices and functioning. The team used global ratings and tallies from the Assessing School Settings: Interactions of Students and Teachers (ASSIST) to measure classroom management practices and functioning. The ASSIST is a direct observational coding system developed to evaluate student and teacher behaviors in classroom and non-classroom school settings.[15] It includes tallies of student and teacher behaviors taken over a 15-minute observation period. After direct observation, the viewers assign global ratings to items reflecting the overall classroom environment during the observed period. Analyses of the psychometric properties of the modified tallies have shown them to be a reliable measure of school-level differences in teacher practices.[16] Analyses of the factor structure and internal consistency of many of the scales in the global ratings demonstrate that the ASSIST global ratings are also reliable indicators of classroom management practices and student behaviors.[17] The University of Virginia team slightly modified the ASSIST protocol for the MTSS-B study. These modifications include adding tallies aligned with the intervention, dropping global ratings not related to the MTSS-B study (for example, culturally responsive practices), and making small revisions to other items in the global ratings to improve the instrument.

School climate. School climate is a broad term that encompasses many dimensions of the learning environment. This study focuses on those dimensions of school climate targeted by the program and considers two types of measures: organizational health measures (which touch on academic emphasis, collegial leadership, and teacher affiliation) and school environment measures (which touch on bullying, school safety, teacher-student relationships, and disciplinary environment). Organizational health measures were primarily adapted from the Organizational Health Inventory.[18] School environment measures were adapted from Delaware's school climate survey, fielded in elementary schools throughout the state.[19] The school discipline scale came from the Department of Education School Climate Survey for instructors.[20]

Implementation of school-wide and supplemental support practices. The study team used the total score from the School-wide Evaluation Tool (SET) to measure MTSS-B Tier I (school-wide) implementation fidelity, which has been found to be a valid and reliable measure of Tier I fidelity for the purpose for which they are used in this study and the interpretations the study is making.[21] The SET includes 29 items organized into seven subscales and assesses the implementation of core components of Tier I school-wide practices. The total SET score ranges from 1 to 100. Past randomized controlled trials of MTSS-B have used it as a measure of service contrast between participating and non-participating schools and found it to discriminate accurately between schools that are and are not implementing the core features of MTSS-B.[22]

The team used the Individual Student Systems Evaluation Tool (I-SSET) to assess the implementation of supplemental support practices. In 2012, the instrument was slightly modified, and the internal consistency of the scales were analyzed. Researchers found that the I-SSET scales exhibited low to adequate reliability, with Cronbach's alpha ranging from 0.50 to 0.64.[23] The study team partnered with these earlier researchers to modify the instrument for the MTSS-B study based on their experiences fielding the instrument and content expertise. It includes 41 items organized into four subscales: foundations, school-wide interventions, supplemental support interventions, and intensive individualized interventions. The study assessed schools' use of supplemental support services by averaging the foundations and supplemental support intervention scales.

The study team also reviewed program data from participating schools to describe their implementation of MTSS-B. These data sources included team meeting minutes (consulted for the frequency, attendance, and content of team meetings) and MTSS-B coach logs (consulted for the frequency and focus of coach activities). Additionally, each year the team collected information from the PBIS Apps systems that participating schools used to record behavior monitoring data to track how those systems were used. Finally, the study team conducted structured interviews with school administrators, team leaders, and coaches in the springs of both program years to hear their perceptions of training and implementation quality.

Exhibit B.6 presents the items included in and the reliabilities of each measure discussed in this section. The reliability values were calculated using all available observations.

Exhibit B.6. Items in Final Scales and Estimated Reliabilities

Construct and Associated Items	Number of Items	Reliability (Alpha)
IMPLEMENTATION/SERVICE CONTRAST (SET/ISSET)		
Implementation of school-wide practices	7	0.839
Average of 7 scales assessing implementation of core components of the school-wide practices [0-100%]		
Behavior expectations defined		
Behavior expectations taught		
Ongoing systems for rewarding behavior expectations		
Systems for responding to behavioral violations		
Monitoring & decision making to support implementation		
Management		
District-level support		
Implementation of supplemental supports	2	0.560
Average of 2 scales assessing implementation of targeted practices		
Foundations		
Supplemental support interventions		

(continued)

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
TEACHER RATINGS OF STUDENT BEHAVIOR		
Teachers report on how often each student displayed specific behaviors in the previous three weeks (never, rarely, sometimes, often, very often, almost always)		
Disruptive behavior	10	0.921
Doesn't get along with others		
Breaks rules		
Harms others		
Yells at others		
Fights		
Gets angry when provoked by other children		
Lies		
Teases classmates		
Harms property		
Bullies others		
Attention to schoolwork	7	0.949
Pays attention		
Works hard		
Stays on task		
Is easily distracted (R)		
Learns up to ability		
Completes assignments		
Concentrates		
Pro-social behavior (positive behaviors with peers)	5	0.867
Is liked by classmates		
Has many friends		
Shows empathy & compassion for others' feelings		
Is rejected by classmates (R)		
Is friendly		
Internalizing behavior	5	0.856
Withdrawn		

(continued)

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
Sad		
Nervous		
Fearful		
Worries		
Emotional dysregulation	5	0.747
Easily upset		
Stops and calms down when angry or upset (R)		
Changes moods quickly		
Easily frustrated		
Impulsive		

STAFF SURVEY

	Number of Items	Reliability (Alpha)
Academic focus	5	0.723

Staff report on how often (rarely, sometimes, often, very frequently) students engage in the following activities

Students respect others who get good grades

Students neglect to complete homework (R)

Students seek extra work so they can get good grades

Students try hard to improve on previous work

Students are cooperative during classroom instruction

	Number of Items	Reliability (Alpha)
Principal leadership	6	0.927

Staff report on how often (rarely, sometimes, often, very frequently) the principal does the following:

The principal conducts meaningful evaluations

The principal treats all school staff members as his or her equal

The principal lets school staff know what is expected of them

The principal explores all sides of topics and admits that other opinions exist

The principal discusses classroom issues with teachers

The principal goes out of his or her way to show appreciation to teachers

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
Staff collegiality	6	0.902
Staff report on how often (rarely, sometimes, often, very frequently) staff do or express the following:		
School staff express pride in their school		
There is a feeling of trust and confidence among the staff		
School staff exhibit friendliness to each other		
School staff identify with the school		
School staff show commitment to their students		
School staff in this school like each other		
Teacher-student relationships	4	0.905
Extent to which staff agree (strongly disagree, disagree, agree, strongly agree) with the following:[a]		
Teachers care about their students		
Teachers listen to students when they have problems		
Adults who work here care about the students		
Teachers like their students		
Consistent school discipline	5	0.915
Extent to which staff agree (strongly disagree, disagree, agree, strongly agree) with the following:[a]		
Staff at this school are clearly informed about school policies and procedures		
School rules are applied equally to all students		
Discipline is fair		
This school effectively handles student discipline and behavior problems		
Staff at this school work together to ensure an orderly environment		
School safety	3	0.933
Extent to which staff agree (strongly disagree, disagree, agree, strongly agree) with the following:[a]		
Students are safe in the hallways		
Students feel safe		
Students know they are safe		

Construct and Associated Items	Number of Items	Reliability (Alpha)
CLASSROOM OBSERVATIONS−TEACHER CLASSROOM MANAGEMENT PRACTICES		
Facilitate orderly classroom transitions	4	0.899
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher facilitates instruction or activities for students		
Teacher prepares students for a change in activity		
Changes in activity occur quickly and smoothly		
Students handle transitions well		
Teacher anticipation and responsiveness	6	0.835
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher maintains proximity to students		
Teacher anticipates when students may have problems academically		
Teacher anticipates when students may have problems behaviorally		
Teacher assesses students' understanding of concepts		
Teacher uses verbal reminders or nonverbal cues regarding expected behaviors		
Teacher is responsive to students' behavioral and/or academic needs		
Teacher proactive behavior management	6	0.830
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher gives clear instructions and directives to students		
Teacher is consistent, even handed, and firm		
Teacher clearly explains learning objectives prior to and/or during the lesson through summary or re-orientation statements		
Teacher praises students for specific behaviors or using social skills		
Teacher makes clear to students how they are expected to treat their peers		
Teacher positively acknowledges or praises students who are treating peers kindly (e.g., including others, listening respectfully)		

(continued)

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
Teacher active monitoring	4	0.936
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher positions him/herself so they can see most of the room area		
Teacher scans the room and is aware of what is occurring		
Teacher monitors all students and all areas		
Teacher is able to focus on one or two students while still scanning all other areas		

CLASSROOM OBSERVATIONS−CLASSROOM FUNCTIONING

	Number of Items	Reliability (Alpha)
Student socially disruptive behaviors	10	0.777
How many times (0 (never), 1 time (rarely), 2-3 times (a few times), 4-6 times (sometimes), 6+ times (often)) the following behaviors were observed in the classroom		
Students are irritable or sarcastic toward peers		
Students argue with peers		
Students physically or verbally harass and/or bully others		
Students are irritable or sarcastic toward the teacher		
Students argue with the teacher		
Students engage in verbal aggression toward teachers		
Social conversations occur between students and peers		
Students engage in physical aggression toward teachers		
Students purposefully exclude a peer		
Students say mean comments intended to harm a reputation or friendships of another (e.g., calling a student a slut or a thief)		
Student compliance	6	0.939
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Students comply		
Students consistently follow rules appropriate to settings		

(continued)

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
Students cooperate		
Students are interested, enthusiastic, and involved		
Students are focused and engaged		
Students treat their peers with respect (i.e. listen when peers are talking)		
Student and teacher meaningful participation	10	0.890
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher gets students involved in lesson by asking questions or making comments		
Teacher encourages students to share their ideas and opinions		
Students have opportunities to make choices		
Students have opportunities to take leadership roles in the classroom		
Students are provided opportunities to contribute to discussion		
Students praise and compliment one another		
Students are pro-social toward one another (helping, sharing, and working together cooperatively)		
Students share their personal opinions and experiences		
Students respond to the teacher's questions and/or volunteer when asked		
There is a positive dynamic between teacher and students		
Teacher control of classroom	5	0.740
Observers report on how often (never, seldom, some, a lot, almost continuously) they saw the following in the classroom		
Teacher has good control of or influence on students		
There is evidence of classroom routines–students know what they're supposed to be doing		
Teacher has little/no control of or influence on students		
There are instances of teacher annoyance, irritability, or sarcasm directed at students		
Teacher and students appear comfortable with one another		

(continued)

Exhibit B.6 (continued)

Construct and Associated Items	Number of Items	Reliability (Alpha)
Counts of student problem behaviors in the classroom	5	0.442

Sum of the number of times each of the following behaviors were observed in the classroom during a 15-minute observation period (a single behavior/incident could not be counted as more than one type)

Student(s) did not comply with teacher or staff directive

Student initiates or extends a disruption in the classroom

Student exhibits verbal or relational aggression to a peer

Students exhibits physical aggression to a peer

Student(s) uses profanity

SOURCES: Authors' summary of information from the following data sources: School-wide Evaluation Tool (SET) and Individual Student Systems Evaluation Tool (I-SSET) fielded in fall 2015. Teacher Ratings of Student Behavior collected in fall 2015, spring 2016, and spring 2017; Staff Survey data collected in spring 2017; Classroom Observations data collected with Assessing School Settings: Interactions of Teachers and Students (ASSIST) in spring 2017.

NOTES: The reliability values were calculated using all available observations from the initial round of data collection.
   (R): reverse coded.
    [a]All items in the construct were recoded to take on the following values: strongly disagree = 1; disagree = 2; agree = 3; and strongly agree = 4.

Estimating Program Effects

Given the blocked, school-level random assignment design of the study, conceptually the impact of the MTSS-B program should be the difference in outcomes between the participating schools and the non-participating schools. All impact estimates are based on an intent-to-treat analysis that includes all individuals who have valid outcome measures in the sample schools. In other words, the impact estimates reflect the impact of assignment to the MTSS-B program of training and technical assistance.

The team used different regression models to estimate the program's effects on student, teacher/classroom, and school outcomes. Exhibit B.7 provides a detailed description of each model.

These models share some common features. All estimation models use data from all study districts in a single analysis. They estimate separate program impacts for each district and then average them across the districts, weighting each district's estimate in proportion to the number of participating schools in the district. Therefore, these findings represent the program's impact on the average participating school in the study sample. All models also include random assignment block indicators as fixed effects to account for the blocked random assignment design. Wherever appropriate, grade indicators are also included as fixed effects to account for possible variations in outcome levels across grades.

Exhibit B.7. Impact Estimation Models

This text box provides detailed descriptions of the three prototypical models used for impact estimations for various outcomes in the study.

MODEL 1: THREE-LEVEL MODEL

For student behavior outcomes, the primary impact estimation model is a three-level, hierarchical, linear model that accounts for the clustering of students within teachers, and teachers within schools.

$$Y_{ijk} = \sum_m \alpha_{0m} B_{mijk} + \sum_n \beta_n T_k D_{nijk} + \sum_l \gamma_l X_{lijk} + \mu_k + \omega_{jk} + \varepsilon_{ijk} \qquad (1)$$

Where

Y_{ijk}	=	Outcome measure for student i with teacher j in school k;
B_{mijk}	=	1 if student i with teacher j in school k is in random assignment block m, 0 otherwise;
D_{nijk}	=	1 if student i with teacher j in school k is in district n, 0 otherwise;
T_k	=	1 if school k is a participating school, 0 if it is a non-participating school;
X_{lijk}	=	lth student-level covariate for student i with teacher j in school k;
$\mu_k, \omega_{jk}, \varepsilon_{ijk}$	=	School-, teacher-, and student-level random errors respectively assumed to be independently and identically distributed.

MODEL 2: TWO-LEVEL MODEL

For student academic achievement outcomes, the impact estimation model is a two-level linear hierarchical model with students clustered within schools. For outcomes measured at the staff or classroom level, such as classroom management practice, classroom functioning, or teachers' perception of school climate, the estimation model is a two-level model with staff/class nested within schools.

$$Y_{jk} = \sum_m \alpha_{0m} B_{mjk} + \sum_n \beta_n T_k D_{njk} + \sum_l \gamma_l Z_{ljk} + \mu_k + \omega_{jk} \qquad (2)$$

Where

Y_{jk}	=	Outcome measure for student/teacher/class j in school k;
B_{mjk}	=	1 if student/teacher/class j in school k is in random assignment block m, 0 otherwise;
D_{njk}	=	1 if student/teacher/class j in school k is in district n, 0 otherwise;
T_k	=	1 if school k is a participating school, 0 if it is a non-participating school;
Z_{ljk}	=	lth student/teacher/class-level covariate for student i with student/teacher/class j in school k;
μ_k, ω_{jk}	=	School- and student/teacher/class-level random errors respectively assumed to be independently and identically distributed.

(continued)

Exhibit B.7 (continued)

MODEL 3: ONE-LEVEL MODEL

For outcomes measured at the school level, such as school-level implementation of MTSS-B practices, the estimation model is an ordinary least squares regression model with the school as the unit of analysis.

$$Y_k = \sum_m \alpha_{0m} B_{mk} + \sum_n \beta_n T_k D_{nk} + \mu_k \qquad\qquad (3)$$

Where

Y_k	=	Outcome measure for school k;
B_{mk}	=	1 if school k is in random assignment block m, 0 otherwise;
D_{nk}	=	1 if school k is in district n, 0 otherwise;
T_k	=	1 if school k is a participating school, 0 if it is a non-participating school;
μ_k	=	School-level random errors respectively assumed to be independently and identically distributed.

β_n represents the estimated program effect for district n. The overall program impact is calculated as the weighted average of β_n, using the number of participating schools in each district as weight.

Student subgroups. The same impact models used to estimate the main effects were also used to estimate the impacts on pre-specified subgroups. One key student subgroup of interest is the group of students initially identified as struggling with behavior. It is hypothesized that these students could benefit more from the program than those without such struggles.[24] This could happen through two channels: first, the school-wide practices promoted by the program could be more beneficial to these struggling students; second, the supplemental support practices that target students in need of more support could directly benefit these students. Therefore, it is of interest to investigate whether there are differential impacts for students with different initial behavioral challenges.

To carry out this analysis, the team identified the group of students initially struggling with behavior as students whose baseline disruptive behavior rating was at or above the 85th percentile value in each grade across study schools. In other words, the team defined the 15 percent of students with the highest disruptive behavior ratings in each grade in fall 2015 as the group of students initially struggling with behavior. The team used this definition because it aligned with the literature on tiered systems of support for behavior, which generally suggests that up to 15 percent of a school's population would benefit from supplemental or intensive intervention supports.[25] As shown in Appendix D, the study conducted various sensitivity analyses to demonstrate that the program effect findings for this subgroup were not sensitive to how this group was defined. In addition to the subgroup based on students' initial behavior ratings, the study also explored the program effects on students with different background characteristics such as gender, grade level, special education status, and English learner status.

The team modified Equation (1) for the student subgroup analysis. For each set of subgroup analyses, the team created an indicator ("group") to show if a student belonged to a subgroup and used the following model to estimate separate program effects for each group:

$$Y_{ijk} = \sum_m \alpha_{0m} B_{mijk} + \sum_m \alpha_{1m} B_{mijk} * Group_{ijk} + \sum_n \beta_{0n} T_k D_{nijk} + \sum_n \beta_{1n} T_k D_{nijk} * Group_{ijk} + \sum_l \gamma_l X_{lijk} + \mu_k + \omega_{jk} + \varepsilon_{ijk}$$

$$(1.1)$$

Here the weighted average across β_{0n} is the estimated effect for students not in the subgroup, and the sum of the weighted averages of β_{0n} and β_{1n} is the estimated effect for students in the subgroup.

Missing data. The impact analyses did not include observations with missing outcome values. The team adopted this listwise deletion approach for missing outcome data out of the concern that the imputation methods may not be appropriate for this study. For student behavior measures, there were more than a third of missing cases largely due to low consent rate. Imputing for such a substantial proportion of the sample might increase the risk of bias. What's more, not including observations with missing outcome values in the analysis is a widely accepted method for dealing with missing outcome data.[26] The study team conducted sensitivity checks to see if the program effects on student achievement differ between students with and without teacher ratings and found mostly no difference in impacts between these two groups of students (Exhibit D.11).

For missing covariate values, the team replaced the missing data with zeros and added an indicator for a given covariate's missing status to the model. Research has demonstrated that this approach, known as the "dummy variable imputation method," is unlikely to create estimation bias that is larger than 0.05 standard deviations in an experimental setting.[27]

Covariates. To improve the precision of the impact estimates, the analysis controls for random differences between the participating and non-participating schools' students, staff members, or classes with respect to the following background characteristics:

- For student-level impact estimation, the analysis includes student baseline characteristics such as students' race/ethnicity, gender, free or reduced-price lunch status, English language learner status, and Individualized Education Program status, along with grade indicators and students' baseline achievement and behavior information.

- For class-level impact estimations, the covariates include time of observation (morning or afternoon), the lead teacher's gender, the number of adults and number of students present during the observation period, and indicators for observers and grade levels.

- For staff-level impact estimations, the analysis includes staff demographic information such as age, gender, and experience working in the current school and in any schools.

- All impact estimations include random assignment block indicators.

Using teacher ratings of student behavior collected in fall 2015 as baseline measures. The study team originally planned to collect the teacher rating data as early as possible in the fall 2015 semester. However, due to the lengthy consent process, the actual collection time was delayed and prolonged. Specifically, the collection time ranged from November to February, with the majority of the districts completing the data collection prior to the end of December. Note that the lateness of baseline, or pretest data collection, is not unique to this study. For logistical reasons, in randomized controlled trials (RCT), pretest data are often collected after random assignment. The issue of whether to collect and use late pretest data in RCTs involves a variance-bias tradeoff. Research on this issue found that for RCTs, estimators that include late pretests will typically be preferred to estimators that exclude them or that instead include uncontaminated baseline test score data from other sources. This result holds as long as impacts do not grow very quickly early in the school year. [28] The What Works Clearinghouse also accepts a baseline measure assessed after the start of the intervention to be used to satisfy the baseline equivalence requirement.[29]

The study team conducted the following sensitivity checks to assess the validity of the fall 2015 teacher ratings as baseline measures for student behavior:

- Estimated the program/business as usual (BAU) difference for all student behavior outcomes specified to see if there is an "early impact" of the program;

- Divided the sample into two subgroups of early and late respondents based on the submission time of each rating[30] and estimated program impacts for each subgroup. This allowed the research team to see if the late group obtains significantly larger impacts than the early group;

- Interacted the response time with treatment indicators and added this interaction in the impact estimation model to see how much the estimated program impact grows over time.

Exhibit B.8 presents findings from these checks. In general, results in the table show that there is no evidence of statistically significant early program impacts (Panel A), and the observed early impacts do not grow quickly early on (Panels B and C).[31] In other words, the impacts on student behavior outcomes did not grow fast early on during Program Year 1. Therefore, it is appropriate to use data from the fall 2015 TSSB as a proxy for baseline student behavior measures.

Understanding the Impact Tables. Throughout this report, when a table or figure is presented to report the estimated program impacts, the mean outcome levels for the participating and the non-participating schools are reported to provide context for interpreting the estimated differences. Program impacts are estimated using an impact regression model outlined in Exhibit B.7, and the mean outcome levels are calculated by using the same impact regression model. Specifically, the impact tables and figures report the observed mean outcome levels for schools randomly assigned to the participating group (referred to as the "participating schools mean") and report the regression-adjusted mean outcome levels for students randomly assigned to the non-participating group (referred to as the "non-participating schools" mean), using the observed mean covariate values for the participating group as the basis for the adjustment.

Presenting the observed mean outcome values for the participating schools allows comparison between these actual values and other reference groups. The reported mean outcome levels for the non-participating schools provide an unbiased estimate of how the participating schools would have performed had they not been assigned to receive training and support for MTSS-B. In other words, it represents the "counterfactual".

Statistical significance and multiple hypotheses testing

Statistical significance is a measure of the degree of certainty that one may have that a program's impact is actually non-zero. In this report, statistical significance is indicated in the exhibits by an asterisk (*) when the p-value of the impact estimate is less than 5 percent when using a two-tailed test.

When making judgments about statistical significance, however, it is important to recognize the potential problems associated with conducting multiple hypothesis tests. Specifically, conducting hypothesis tests for estimated impacts on several different outcomes increases the likelihood of concluding that a given impact estimate is statistically significant, when in fact the program has no impact (this is known as a type I error or a false positive). Although it is important to avoid making conclusions based on such errors, efforts to control for this problem may reduce statistical power and unduly increase the likelihood of missing true impacts when they exist (that is, relying on false negative results, or a type II error).

When evaluating the program effects in this report, two sets of safeguards were used to attenuate the risk of drawing inappropriate conclusions about effectiveness on the basis of statistically significant results that may have occurred by chance. The first safeguard is to identify a set of "confirmatory" outcomes and samples before beginning the impact analysis. Primary evidence of the program's effectiveness is based on the estimated impacts on these outcomes for these samples. The confirmatory outcomes for this study are teacher ratings of students' disruptive behavior and their reading and math achievements as measured by state test scores. The confirmatory samples are the overall student sample and the subgroup of students who were initially identified as struggling with behavior. All other outcomes and subgroups are considered "exploratory" and are used either to contextualize the primary impact findings or to generate hypotheses about impacts. These designations were included in the design and analysis plans for the study that were reviewed by outside experts prior to the analysis phase.

Exhibit B.8. Sensitivity Checks on "Early Impacts" of Teacher Ratings of Student Behavior, Overall Sample

Outcome Measures (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impact	Standard Error of Estimated Impact	Estimated Impact in Effect Size	p-value for Estimated Impact	p-value for Early/Late Impact Difference
Panel A. Overall "impact"							
Disruptive behavior	0.50	0.52	-0.02	0.03	-0.03	0.498	
Emotional dysregulation	1.12	1.14	-0.02	0.03	-0.02	0.587	
Internalizing behavior	0.63	0.66	-0.02	0.02	-0.03	0.324	
Attention to schoolwork	3.47	3.47	0.01	0.04	0.01	0.854	
Pro-social behavior	3.95	3.91	0.04	0.04	0.04	0.344	
Panel B. By early/late respondents							
Disruptive behavior			0.05	0.03	0.08		0.106
Early	0.46	0.51	-0.05	0.03	-0.08	0.100	
Late	0.53	0.53	0.00	0.03	0.00	0.979	
Emotional dysregulation			0.06	0.04	0.08		0.132
Early	1.07	1.14	-0.06	0.04	-0.08	0.105	
Late	1.15	1.15	0.00	0.04	0.00	0.981	
Internalizing behavior			0.05	0.04	0.08		0.134
Early	0.59	0.64	-0.06	0.03	-0.08	0.062	
Late	0.67	0.68	-0.01	0.03	-0.01	0.835	
Attention to schoolwork			0.00	0.06	0.00		0.994
Early	3.43	3.42	0.01	0.05	0.01	0.873	
Late	3.44	3.44	0.01	0.05	0.01	0.871	
Pro-social behavior			0.01	0.05	0.02		0.751
Early	3.92	3.90	0.02	0.05	0.02	0.640	
Late	3.90	3.86	0.04	0.04	0.04	0.408	

(continued)

Exhibit B.8 (continued)

Panel C. Impact change per day			Estimated Increase per Day	Standard Error of Estimated Increase	Estimated Increase in Effect Size	p-value for Estimated Increase	
Disruptive behavior			0.002	0.001	0.003	0.015	*
Emotional dysregulation			0.002	0.001	0.002	0.089	
Internalizing behavior			0.003	0.001	0.004	0.012	*
Attention to schoolwork			0.000	0.001	0.000	0.960	
Pro-social behavior			0.000	0.001	0.000	0.770	
Sample size	15,529	7,180					

SOURCE: Teacher Survey of Student Behavior (TSSB) collected in fall 2015.

NOTES: This table is based on the 58 participating schools and 31 non-participating schools in the study. The sample includes students with teacher ratings for all five behavior measures from the fall of 2015. The estimated "early impacts" in Panel A are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating means and differences. The early ratings are defined as teacher ratings submitted by teachers on or before December 13, 2015. The impact change per day was the estimated coefficient for the interaction between the rating submission time and the treatment indicator. The multi-level regression also controlled for the blocking of random assignment.

[a]Submission time is measured by the number of days elapsed since a fixed time in the year. For the fall 2015 round, the fixed day was August 15, 2015.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

As a further safeguard, the study team conducted factor analysis to delineate both the confirmatory outcomes and the exploratory ones to make sure that they each reflect a core domain that is internally cohesive (measuring one underlying construct or dimension), and distinct from each other. This analysis was done prior to the impact analysis phase as well. Because each outcome is a measure of a different domain, it is not necessary to make adjustments to p-values for multiple hypothesis testing, based on standards used in education research.[32]

Samples. The analysis samples used for impact estimation differ by outcome. Exhibit B.9 lists all the samples used in the impact analysis of this study. The team included in the analyses all students, teachers or classes, staff members, and schools for which outcome data were available for a given year. Doing so maximized the sample sizes and the statistical power of the analyses. It also helps to increase the generalizability of the results. Appendix D provides findings based on alternative sample definitions, which generally confirm the findings presented in the report.

Exhibit B.9. Analysis Samples by Outcome

Outcome	Sample	Baseline Comparison
Student achievement	Grade 3-5 students with state test scores in study schools in the springs of 2016, 2017, and 2018	Exhibits B.10a-B.12a: overall student sample Exhibits B.10b-B.12b: students initially identified as struggling with behavior
Student behavior	Grade 1-5 students with teacher ratings of behavior in study schools in the springs of 2016 and 2017	Exhibits B.13a-B.14a: overall student sample Exhibits B.13b-B.14b: students initially identified as struggling with behavior
Classroom management practices	Grade 1-5 general education and inclusion classrooms in study schools observed in the spring of 2017	Exhibit B.15
Classroom functioning	Grade 1-5 general education and inclusion classrooms in study schools observed in the spring of 2017	Exhibit B.15
School climate	Staff members who interact directly with students on a regular basis in study schools and who responded to the staff survey in the spring of 2017	Exhibit B.16
Implementation of school-wide and supplemental support practices	Schools visited by the study team in the fall of 2015 and the springs of 2017 and 2018	Exhibits B.3 and B.17

The team examined the baseline characteristics of the participating and non-participating groups for each of the samples listed above and the corresponding subgroup of students initially identified as struggling with behavior. Across most of the samples, the two groups shared similar characteristics at the start of the study (Exhibit B.10a-B.17). There were a few sporadic differences detected across the samples, but the differences were generally small in magnitude (less than 0.25 in effect size). These characteristics were included as covariates in the impact estimation models for student outcomes.

Exhibit B.10a. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for All Students in Achievement Analyses, Program Year 1

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	8.0	8.0	0.0		0.0	0.01	0.559
Male	50.3	51.6	-1.3		0.7	-0.03	0.084
Race/ethnicity							
Black, non-Hispanic	9.7	10.7	-1.0		1.3	-0.03	0.447
White, non-Hispanic	20.4	17.2	3.2	*	1.6	0.09	0.044
Hispanic	59.7	62.2	-2.5		1.6	-0.05	0.113
Asian	7.0	6.2	0.9		0.8	0.04	0.286
Other	3.1	3.1	0.0		0.3	0.00	0.960
Has free/reduced-price lunch status	56.6	59.8	-3.1		1.8	-0.07	0.075
Has English language learner status	33.2	36.9	-3.7		2.2	-0.08	0.091
Has special education status	9.7	9.7	0.0		0.7	0.00	0.950
Baseline grade levels							
Grade 2 (Grade 3 in Year 1)	33.5	33.3	0.2		0.7	0.00	0.754
Grade 3 (Grade 4 in Year 1)	33.3	34.2	-0.9		0.7	-0.02	0.194
Grade 4 (Grade 5 in Year 1)	33.2	32.5	0.7		0.7	0.01	0.321
State standardized test score at baseline[a]							
Reading (in effect size unit)	0.04	0.01	0.04		0.05	0.04	0.464
Math (in effect size unit)	0.05	0.03	0.02		0.05	0.02	0.689
Sample size	15,470	8,856					

(continued)

Exhibit B.10a (continued)

SOURCE: District records data for the 2014-2015 school year.

NOTES: The sample includes all eligible Grade 3-5 students in 58 participating schools and 31 non-participating schools in the spring of 2016 who have test scores for both Reading and Math state tests (n = 24,326). The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

The values for participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

[a]The state standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015).

Exhibit B.10b. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Students Who Were in Achievement Analyses and Were Initially Identified as Struggling with Behavior, Program Year 1

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	7.9	8.0	0.0		0.1	-0.02	0.831
Male	67.4	65.5	1.9		2.4	0.04	0.435
Race/ethnicity							
Black, non-Hispanic	13.3	14.0	-0.7		2.2	-0.02	0.743
White, non-Hispanic	20.9	19.3	1.6		2.3	0.04	0.485
Hispanic	58.1	59.1	-1.0		2.2	-0.02	0.643
Asian	5.1	4.4	0.7		1.6	0.04	0.690
Other	2.6	3.1	-0.4		0.8	-0.03	0.598
Has free/reduced-price lunch status	57.8	62.4	-4.6	*	2.1	-0.10	0.031
Has English language learner status	31.5	34.6	-3.1		3.2	-0.06	0.334
Has special education status	13.2	18.6	-5.4	*	2.3	-0.15	0.020
Baseline grade levels							
Grade 2 (Grade 3 in Year 1)	37.9	33.7	4.2		3.6	0.09	0.241
Grade 3 (Grade 4 in Year 1)	29.6	33.8	-4.2		4.0	-0.09	0.292
Grade 4 (Grade 5 in Year 1)	32.5	31.8	0.7		4.4	0.01	0.879
State standardized test score at baseline[a]							
Reading (in effect size unit)	-0.29	-0.34	0.05		0.09	0.06	0.555
Math (in effect size unit)	-0.23	-0.37	0.14		0.10	0.15	0.143
Sample size	1,235	644					

(continued)

Exhibit B.10b (continued)

SOURCE: District records data for the 2014-2015 school year.

NOTES: The sample includes all eligible students in Grades 3-5, in 58 participating schools and 31 non-participating schools in the spring of 2016 who have test scores for both Reading and Math state tests. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 1 achievement analysis sample has 1,879 such students. The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

The values for participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

[a] The state standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015).

Exhibit B.11a. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for All Students in Achievement Analyses, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	7.0	7.0	0.0	0.0	-0.02	0.249
Male	50.9	51.7	-0.8	0.8	-0.02	0.317
Race/ethnicity						
Black, non-Hispanic	9.0	11.2	-2.2	1.5	-0.07	0.148
White, non-Hispanic	20.1	17.1	3.0	1.9	0.08	0.124
Hispanic	60.3	62.5	-2.2	1.8	-0.05	0.218
Asian	7.3	5.7	1.7	0.9	0.07	0.075
Other	3.3	3.4	-0.2	0.3	-0.01	0.634
Has free/reduced-price lunch status	57.3	59.2	-1.9	2.1	-0.04	0.357
Has English language learner status	36.6	38.7	-2.1	2.8	-0.04	0.448
Has special education status	8.6	8.9	-0.3	0.7	-0.01	0.641
Baseline grade levels						
Grade 1 (Grade 3 in Year 2)	33.2	32.1	1.1	0.7	0.02	0.109
Grade 2 (Grade 4 in Year 2)	33.4	34.0	-0.6	0.7	-0.01	0.379
Grade 3 (Grade 5 in Year 2)	33.3	33.8	-0.5	0.7	-0.01	0.445
State standardized test score at baseline[a]						
Reading (in effect size unit)	0.06	0.00	0.07	0.06	0.07	0.257
Math (in effect size unit)	0.08	0.02	0.06	0.06	0.06	0.358
Sample size	14,933	8,636				

(continued)

Exhibit B.11a (continued)

NOTES: The sample includes all eligible students in Grades 3-5, in 58 participating schools and 31 non-participating schools in the spring of 2017 who have test scores for both Reading and Math state tests (n = 23,569). The number of observations varies by baseline characteristics due to missing values.

 The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

 The values for participating schools are the weighted average of the observed district means for the participating schools (using number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

 *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

 [a] The state standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015).

Exhibit B.11b. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Students Who Were in Achievement Analyses and Were Initially Identified as Struggling with Behavior, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	7.0	7.0	-0.1		0.1	-0.06	0.497
Male	66.0	66.8	-0.8		2.6	-0.02	0.753
Race/ethnicity							
Black, non-Hispanic	12.1	12.4	-0.3		2.4	-0.01	0.907
White, non-Hispanic	22.1	21.0	1.1		2.8	0.03	0.705
Hispanic	58.1	59.1	-1.0		2.5	-0.02	0.693
Asian	5.4	4.6	0.8		2.0	0.04	0.684
Other	2.4	3.0	-0.6		0.9	-0.04	0.531
Has free/reduced-price lunch status	56.9	64.2	-7.3	*	2.1	-0.16	0.001
Has English language learner status	32.3	34.1	-1.8		3.7	-0.04	0.628
Has special education status	10.8	17.3	-6.4	*	1.9	-0.19	0.001
Baseline grade levels							
Grade 1 (Grade 3 in Year 2)	34.2	31.0	3.1		4.6	0.07	0.498
Grade 2 (Grade 4 in Year 2)	35.6	34.0	1.6		4.1	0.03	0.692
Grade 3 (Grade 5 in Year 2)	30.1	35.1	-5.0		4.1	-0.11	0.222
State standardized test score at baseline[a]							
Reading (in effect size unit)	-0.19	-0.35	0.16		0.14	0.17	0.250
Math (in effect size unit)	-0.14	-0.34	0.20		0.15	0.20	0.196
Sample size	1,052	543					

(continued)

Exhibit B.11b (continued)

SOURCES: District records collected in 2015, 2016, and 2017, and Teacher Survey of Student Behavior fielded in the fall of 2015.

NOTES: The sample includes all eligible students in Grades 3-5, in 58 participating schools and 31 non-participating schools in spring 2017 who have test scores for both Reading and Math state tests. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 2 achievement analysis sample has 1,595 such students. The number of observations varies by baseline characteristics due to missing values.

   The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

   The values for participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

   [a]The state standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015).

Exhibit B.12a. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for All Students in Achievement Analyses, Follow-Up Year

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	6.0	6.0	0.0		0.0	-0.02	0.252
Male	50.7	51.2	-0.5		0.9	-0.01	0.528
Race/ethnicity							
Black, non-Hispanic	6.9	7.5	-0.6		1.0	-0.02	0.555
White, non-Hispanic	16.7	13.1	3.6	*	1.3	0.10	0.005
Hispanic	66.6	71.0	-4.5	*	1.5	-0.10	0.003
Asian	7.5	5.4	2.1	*	0.8	0.09	0.013
Other	2.4	2.3	0.1		0.3	0.01	0.689
Has free/reduced-price lunch status	58.5	59.8	-1.3		1.4	-0.03	0.366
Has English language learner status	42.5	47.1	-4.7	*	2.1	-0.09	0.024
Has special education status	8.2	8.1	0.1		0.7	0.00	0.891
Baseline grade levels							
Grade K (Grade 3 in Follow-up Year)	31.2	31.0	0.2		0.8	0.00	0.813
Grade 1 (Grade 4 in Follow-up Year)	34.4	33.4	1.0		0.8	0.02	0.177
Grade 2 (Grade 5 in Follow-up Year)	34.4	35.3	-0.9		0.7	-0.02	0.219
Sample size	12,643	7,325					

(continued)

Exhibit B.12a (continued)

NOTES: The sample includes all eligible students in Grades 3-5, in 52 participating schools and 28 non-participating schools in the spring of 2018 who have test scores for both Reading and Math state tests (n =19,968). The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

The values for participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

Exhibit B.12b. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Students Who Were in Achievement Analyses and Were Initially Identified as Struggling with Behavior, Follow-Up Year

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Age at baseline (year)	6.1	6.1	-0.1		0.1	-0.09	0.264
Male	64.5	69.5	-5.0		3.0	-0.11	0.099
Race/ethnicity							
Black, non-Hispanic	8.5	9.5	-1.1		2.1	-0.04	0.610
White, non-Hispanic	15.3	13.9	1.4		2.8	0.04	0.615
Hispanic	69.8	73.1	-3.3		2.7	-0.07	0.229
Asian	4.3	2.2	2.1		1.6	0.11	0.188
Other	2.1	1.0	1.1		0.8	0.08	0.195
Has free/reduced-price lunch status	61.7	70.6	-8.9	*	2.3	-0.20	0.000
Has English language learner status	40.1	42.8	-2.7		4.0	-0.05	0.496
Has special education status	10.9	17.0	-6.1	*	2.5	-0.19	0.014
Baseline grade levels							
Grade K (Grade 3 in Follow-up Year)	29.0	26.0	2.9		3.6	0.07	0.416
Grade 1 (Grade 4 in Follow-up Year)	35.7	35.9	-0.2		4.3	0.00	0.958
Grade 2 (Grade 5 in Follow-up Year)	35.2	36.7	-1.6		4.1	-0.03	0.698
Sample size	921	402					

(continued)

Exhibit B.12b (continued)

SOURCE: District records data for the 2014-2015 school year.

NOTES: The sample includes all eligible students in Grades 3-5, in 52 participating schools and 28 non-participating schools in the spring of 2018 who have test scores for both Reading and Math state tests. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Follow-up Year achievement analysis sample has 1,323 such students. The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with students nested within schools). The models control for indicators of random assignment blocks.

The values for participating schools are the weighted average of the observed district means for the participating schools (using the number of participating schools in each district as weight). The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

Exhibit B.13a. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Analysis of Student Behavior, Overall Student Sample, Program Year 1

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics						
Age (year)	7.0	7.0	0.0	0.1	-0.01	0.837
Male	50.6	50.9	-0.2	0.8	0.00	0.770
Race/ethnicity						
Black, non-Hispanic	8.7	9.8	-1.1	1.7	-0.04	0.511
White, non-Hispanic	20.9	18.5	2.4	2.1	0.06	0.246
Hispanic	59.5	62.0	-2.5	1.8	-0.05	0.184
Asian	7.7	6.7	1.0	0.9	0.04	0.276
Other	3.2	3.0	0.2	0.3	0.01	0.562
Has free/reduced-price lunch status	56.0	58.5	-2.5	2.0	-0.05	0.215
Has English language learner status	35.6	37.2	-1.6	2.6	-0.03	0.546
Has special education status	9.2	9.7	-0.5	1.0	-0.02	0.644
Baseline grade level (2014-2015)						
Grade K (Grade 1 in Year 1)	20.1	20.5	-0.4	2.1	-0.01	0.846
Grade 1 (Grade 2 in Year 1)	20.0	19.9	0.1	2.8	0.00	0.970
Grade 2 (Grade 3 in Year 1)	20.5	19.7	0.7	2.3	0.02	0.752
Grade 3 (Grade 4 in Year 1)	20.0	19.4	0.6	2.1	0.01	0.785
Grade 4 (Grade 5 in Year 1)	19.5	20.6	-1.1	2.5	-0.03	0.653

(continued)

Exhibit B.13a (continued)

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Teacher ratings of student behavior at start of Year 1 (0 to 5 scale)						
Attention to schoolwork (scale point)	3.50	3.48	0.01	0.04	0.01	0.787
Pro-social behavior (scale point)	3.96	3.92	0.04	0.04	0.04	0.300
Disruptive behavior (scale point)	0.49	0.51	-0.02	0.03	-0.03	0.530
Emotional dysregulation (scale point)	1.11	1.12	-0.02	0.03	-0.02	0.621
Internalizing behavior (scale point)	0.62	0.66	-0.03	0.03	-0.05	0.221
Standardized score in baseline year state tests (2014-2015)[a]						
Reading (in effect size unit)	0.09	0.04	0.05	0.06	0.05	0.405
Math (in effect size unit)	0.09	0.02	0.07	0.07	0.07	0.300
Sample size	16,991	8,400				

SOURCES: Teacher Ratings of Student Behavior Survey from the fall of 2015 and spring of 2016, and district records data collected for the 2014-2015 and 2015-2016 school years.

NOTES: The sample includes all eligible Grade 1-5 students in the study schools in the spring of 2016 who have teacher ratings for all five behavior measures (n = 25,391). The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the difference between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

[a] The state standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015).

Exhibit B.13b. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Analysis of Student Behavior, Students Initially Identified as Struggling with Behavior, Program Year 1

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics						
Age at baseline (year)	6.9	7.0	-0.1	0.1	-0.05	0.440
Male	65.7	66.5	-0.8	2.1	-0.02	0.694
Race/ethnicity						
Black, non-Hispanic	13.4	13.3	0.1	1.8	0.00	0.960
White, non-Hispanic	21.0	19.1	2.0	2.2	0.05	0.365
Hispanic	57.2	60.4	-3.2	2.1	-0.07	0.124
Asian	5.3	4.3	1.0	1.1	0.05	0.372
Other	3.0	2.8	0.2	0.7	0.01	0.740
Has free/reduced-price lunch status	58.2	62.2	-4.0 *	2.0	-0.09	0.049
Has English language learner status	32.1	35.3	-3.3	2.9	-0.07	0.255
Has special education status	13.6	16.3	-2.7	2.1	-0.07	0.190
Baseline grade level (2014-2015)						
Grade K (Grade 1 in Year 1)	20.7	19.9	0.9	2.8	0.02	0.760
Grade 1 (Grade 2 in Year 1)	20.8	20.4	0.4	2.7	0.01	0.876
Grade 2 (Grade 3 in Year 1)	22.1	18.6	3.5	2.8	0.09	0.217
Grade 3 (Grade 4 in Year 1)	17.4	20.6	-3.2	2.6	-0.08	0.215
Grade 4 (Grade 5 in Year 1)	19.1	20.6	-1.5	2.7	-0.04	0.581

(continued)

Exhibit B.13b (continued)

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Teacher ratings of student behavior at start of year 1 (0- to 5-point scale)						
Attention to schoolwork (scale point)	2.31	2.31	0.01	0.05	0.01	0.890
Pro-social behavior (scale point)	2.80	2.76	0.04	0.04	0.05	0.364
Disruptive behavior (scale point)	1.77	1.78	-0.01	0.03	-0.01	0.767
Emotional dysregulation (scale point)	2.29	2.24	0.05	0.04	0.06	0.230
Internalizing behavior (scale point)	1.15	1.14	0.00	0.04	0.01	0.908
Standardized score in baseline year state tests (2014-2015)[a]						
Reading (in effect size unit)	-0.26	-0.36	0.10	0.09	0.11	0.290
Math (in effect size unit)	-0.26	-0.42	0.20 *	0.10	0.21	0.039
Sample size	2,040	939				

SOURCES: Teacher Ratings of Student Behavior Survey from the fall of 2015 and spring of 2016, and district records data collected for the 2014-2015 and 2015-2016 school years.

NOTES: The sample includes all eligible students in Grades 1-5 in the study schools, in the spring of 2016 who have teacher ratings for all five behavior measures. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 1 behavior analysis sample has 2,979 such students. The number of observations varies by baseline characteristics due to missing values.

   The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the ob-served district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

   [a] The state standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015).

Exhibit B.14a. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Analysis of Student Behavior, Overall Student Sample, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics						
Age at baseline (year)	6.0	6.0	0.0	0.1	0.02	0.657
Male	51.0	50.4	0.7	0.8	0.01	0.402
Race/ethnicity						
Black, non-Hispanic	8.0	9.4	-1.4	1.5	-0.05	0.350
White, non-Hispanic	20.6	18.4	2.2	2.0	0.06	0.285
Hispanic	60.4	62.2	-1.8	1.9	-0.04	0.338
Asian	7.6	6.6	1.0	0.9	0.04	0.262
Other	3.4	3.3	0.1	0.3	0.01	0.697
Has free/reduced-price lunch status	56.7	58.1	-1.4	1.9	-0.03	0.476
Has English language learner status	37.3	38.3	-1.0	2.5	-0.02	0.696
Has special education status	9.1	8.9	0.2	1.1	0.01	0.844
Baseline grade level (2014-2015)						
Pre-K (Grade 1 in Year 2)	18.9	19.3	-0.5	2.1	-0.01	0.829
Grade K (Grade 2 in Year 2)	19.8	20.5	-0.6	2.1	-0.02	0.758
Grade 1 (Grade 3 in Year 2)	20.1	19.0	1.1	2.1	0.03	0.601
Grade 2 (Grade 4 in Year 2)	20.7	22.9	-2.2	2.1	-0.05	0.304
Grade 3 (Grade 5 in Year 2)	20.4	18.0	2.4	2.3	0.06	0.288

(continued)

Exhibit B.14a (continued)

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Teacher ratings of student behavior at start of Year 1 (0 to 5 scale)						
Attention to schoolwork (scale point)	3.50	3.48	0.01	0.04	0.01	0.745
Pro-social behavior (scale point)	3.99	3.95	0.04	0.05	0.04	0.398
Disruptive behavior (scale point)	0.48	0.50	-0.02	0.03	-0.02	0.598
Emotional dysregulation (scale point)	1.10	1.12	-0.02	0.03	-0.02	0.564
Internalizing behavior (scale point)	0.62	0.66	-0.04	0.03	-0.05	0.213
Standardized score in baseline year state tests (2014-2015)[a]						
Reading (in effect size unit)	0.10	0.06	0.04	0.07	0.04	0.558
Math (in effect size unit)	0.11	0.09	0.02	0.08	0.02	0.825
Sample size	16,560	8,282				

SOURCES: Teacher Survey of Student Behavior from fall of 2015, spring of 2016, and spring of 2017. District records data collected for 2014-2015 and 2016-2017 school years.

NOTES: The sample includes all eligible Grade 1-5 students in the study schools in spring 2017 who have teacher ratings for all five behavior measures (n = 24,842). The number of observations varies by baseline characteristics due to missing values.

   The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the difference between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates that the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

   [a] The state standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015).

Exhibit B.14b. Background Characteristics Comparison of Students in Participating and Non-Participating Schools, for Analysis of Student Behavior, Students Initially Identified as Struggling with Behavior, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference		Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics							
Age at baseline (year)	6.5	6.6	-0.1		0.1	-0.08	0.312
Male	66.2	66.6	-0.4		2.5	-0.01	0.888
Race/ethnicity							
Black, non-Hispanic	11.0	10.6	0.4		2.1	0.01	0.858
White, non-Hispanic	21.5	21.6	-0.1		2.4	0.00	0.953
Hispanic	59.5	61.3	-1.7		2.3	-0.04	0.459
Asian	4.7	3.8	0.9		1.6	0.05	0.580
Other	3.3	2.6	0.7		0.8	0.04	0.433
Has free/reduced-price lunch status	57.8	64.0	-6.2	*	2.2	-0.14	0.004
Has English language learner status	33.2	33.6	-0.3		3.2	-0.01	0.917
Has special education status	12.8	15.7	-2.9		2.2	-0.08	0.182
Baseline grade level (2014-2015)							
Grade K (Grade 2 in Year 2)	24.3	22.4	2.0		3.3	0.05	0.550
Grade 1 (Grade 3 in Year 2)	24.8	23.2	1.6		3.2	0.04	0.609
Grade 2 (Grade 4 in Year 2)	27.6	25.4	2.2		3.3	0.05	0.505
Grade 3 (Grade 5 in Year 2)	23.2	29.0	-5.9		3.1	-0.13	0.063

(continued)

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Teacher ratings of student behavior at start of year 1 (0- to 5-point scale)						
Attention to schoolwork (scale point)	2.36	2.32	0.04	0.06	0.04	0.525
Pro-social behavior (scale point)	2.83	2.80	0.03	0.05	0.04	0.601
Disruptive behavior (scale point)	1.75	1.80	-0.05	0.03	-0.08	0.150
Emotional dysregulation (scale point)	2.26	2.26	0.01	0.05	0.01	0.867
Internalizing behavior (scale point)	1.16	1.18	-0.02	0.05	-0.03	0.688
Standardized score in baseline year state tests (2014-2015)[a]						
Reading (in effect size unit)	-0.21	-0.29	0.08	0.17	0.08	0.640
Math (in effect size unit)	-0.16	-0.26	0.10	0.18	0.10	0.580
Sample size	1,275	572				

SOURCES: Teacher Ratings of Student Behavior Survey from the fall of 2015, spring of 2016, and spring of 2017. District records data collected for the 2014-2015 and 2016-2017 school years.

NOTES: The sample includes all eligible students in Grades 2-5 in the study schools, in spring 2017 who have teacher ratings for all five behavior measures. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 2 behavior analysis sample for this subgroup does not include Grade 1 students because they were in kindergarten in the fall of 2015 and were not rated by their teachers at that time. The subgroup sample includes 1,847 students initially identified as struggling with behavior. The number of observations varies by baseline characteristics due to missing values.

   The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

   [a] The state standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015).

Exhibit B.15. Comparison of Classroom Features in Participating and Non-Participating Schools, for Classroom Observation Sample, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Observations that took place in the afternoon	41.6	44.9	-3.4	2.5	-0.07	0.191
Female teachers	87.3	87.3	-0.1	1.7	0.00	0.977
Class size (number of students)	20.2	20.3	-0.2	0.3	-0.03	0.572
Number of adults in the classroom	1.2	1.2	0.0	0.0	-0.05	0.385
Grade 1 class	21.0	20.3	0.7	2.1	0.02	0.729
Grade 2 class	19.6	19.8	-0.2	2.1	0.00	0.933
Grade 3 class	20.7	19.3	1.4	2.1	0.04	0.496
Grade 4 class	18.2	19.4	-1.2	2.0	-0.03	0.541
Grade 5 class	17.8	17.6	0.2	2.0	0.01	0.911
Mixed grade class	2.7	3.6	-1.0	1.0	-0.05	0.361
Number of classrooms	1,152	639				

SOURCE: Classroom observation data collected with Assessing School Settings: Interactions of Students and Teachers (ASSIST) in spring 2017.

NOTES: The analysis sample used in this table is defined as all non-special education classes in Grades 1-5 in sample schools (n = 1,791). The number of observations varies due to missingness in the classroom feature measures.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with classes within schools). The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the difference between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates that the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

[a] State standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015).

Exhibit B.16. Comparison of Staff Background Characteristics in Participating and Non-Participating Schools, for Staff Who Responded to Staff Survey, Program Year 2

Characteristic (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Male	10.3	9.8	0.5	1.0	0.02	0.619
Staff type						
Administrative	7.7	7.9	-0.2	0.9	-0.01	0.860
Instructional	57.7	57.1	0.6	1.6	0.01	0.717
Other	34.0	34.6	-0.5	1.6	-0.01	0.733
Highest degree						
High school	13.3	13.7	-0.4	1.1	-0.01	0.690
College (bachelor's or associate's)	45.2	43.8	1.5	1.6	0.03	0.357
Graduate	39.2	40.0	-0.7	1.6	-0.02	0.632
Part time	10.7	11.1	-0.4	1.1	-0.01	0.748
Has a certificate	70.1	69.3	0.8	1.5	0.02	0.586
Experience (fewer than 3 years in district)	46.3	47.0	-0.8	1.9	-0.02	0.680
Sample size	2,782	1,498				

SOURCE: MTSS-B Staff Survey fielded in the spring of 2017.

NOTES: The analysis sample used in this table is defined as all staff members responded to the staff survey (n = 4,280).

   The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data (with staff members within schools). The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates that the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

Exhibit B.17. Background Characteristics Comparison of Participating and Non-Participating Schools, in Follow-Up Year

Characteristic	Participating Schools	Non-Participating Schools	Estimated Difference	p-value of Estimated Difference
Title I status (percentage of schools)	88.5	95.5	-7.1	0.101
Students eligible for free or reduced-price lunch (average percentage of students per school)	73.3	77.4	-4.2	0.108
Student race/ethnicity (average percentage of students per school)				
White, non-Hispanic	17.1	15.1	1.9	0.276
Black, non-Hispanic	7.7	9.2	-1.4	0.351
Hispanic	64.6	67.5	-2.9	0.165
Asian	6.7	4.9	1.8	0.069
Other	3.9	3.3	0.6	0.153
Students with Individualized Education Programs (average percentage of students per school)	10.1	10.3	-0.3	0.664
Students with limited English proficiency (average percentage of students per school)	42.1	42.2	-0.2	0.943
Male students (average percentage of students per school)	51.5	51.9	-0.4	0.498
School enrollment (number of students)	623	649	-27	0.457
Sample size	52	28		

(continued)

Exhibit B.17 (continued)

NOTES: There are 52 participating schools and 28 non-participating schools that remained in the study for the follow-up year.

The estimated differences for school-level data are regression-adjusted using ordinary least squares (OLS) regressions, controlling for indicators of random assignment blocks. The values for the MTSS-B participating schools are the weighted average of the observed district means for schools randomly assigned to the MTSS-B participating schools (using number of MTSS-B participating schools in each district as weight). The non-participating schools' values in the next column are the regression-adjusted means using the observed distribution of the MTSS-B participating schools across blocks as the basis of the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates that the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

An omnibus test was used to determine whether there is a systematic difference between the participating schools and the non-participating schools, with respect to the characteristics included in this table. The p-value for this test is 0.942.

Exploring Relationships Among Contrasts in Implementation, the Program's Effects on School and Classroom Functioning, and its Effects on Students

The study team made use of the study's multisite random assignment design to explore the association between the intended outcomes and mediational factors. The purpose of the analysis was to understand the circumstances under which the program was more or less successful in affecting student outcomes.

The study team examined the mediational effects of a range of factors, including implementation fidelity, school climate, classroom management practices, and classroom functioning. The analysis focused on the correlation between these mediators and two key student outcomes of interest: teacher ratings of students' disruptive behavior and students' reading achievement. The analysis used data from the spring of Program Year 2 (the 2016-2017 school year) because that was when the team collected data on all the mediational factors. The team conducted separate analyses for the overall sample and the sample of students initially identified as struggling with behavior.

The analytic approach had two steps. The team first estimated separate program impacts on the specified student outcomes and mediational factors for each random assignment block, using one of the models described in Exhibit B.7. The team then correlated the estimated random-assignment-block-level impacts on the mediational factors with those estimated impacts on student outcomes using a variance-known, random-effect meta-analysis model. Recent research presents the specification, identification, and estimation of these models in detail.[33]

Given the small number of random assignment blocks (15) in this study, the analysis focused on the bivariate association between each of the mediational factors and the student outcomes. This approach allowed the team to assess the relationship for one factor at a time, independent of other factors. However, the statistical power of this analysis was still limited, and one needs to interpret the findings with caution. For example, a non-significant finding could be the result of small sample sizes and limited statistical power, rather than an indication that there is no association between the factor and the impacts.

APPENDIX C. SUPPLEMENTAL INFORMATION ON FINDINGS IN THE REPORT

This appendix provides additional details for the study findings. It starts with supplemental information on the findings presented in the report. Next, it presents additional information that a systematic review might need to assess the impact findings for student outcomes. It then provides additional findings on the implementation of the program and MTSS-B practices. Lastly, it compares this study with other studies of similar programs.

I. Additional Details of Report Findings

This section provides supplementary information on the program impact findings presented in the report. It starts with supplemental information on the estimated program impacts on student, classroom, and school outcomes. It then presents details of the analysis that provide contextual information for the findings featured in the report. Such analyses include the implementation of MTSS-B practices in schools and classrooms, and exploratory correlational analyses that assess the relationship between school and classroom outcomes and student outcomes.

Program Impacts on Student Behavior and Academic Achievement Outcomes

This part of the appendix presents details of the program's impact findings on student behavior and academic achievement for students, overall, and for those initially identified as struggling with behavior.

Exhibit 3 in the report shows that the program did not affect student behaviors for Grade 1-5 students enrolled in study schools in Program Year 1 and Program Year 2, overall. Exhibit C.1 presents these findings with their corresponding standard errors and p-values.

Exhibit C.1. Estimated Impacts on Teacher Ratings of Student Behavior, Overall Sample, by Program Year

Measures (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Program Year 1						
Disruptive behavior	0.58	0.61	-0.03	0.02	-0.04	0.100
Emotional dysregulation	1.15	1.17	-0.02	0.02	-0.02	0.410
Internalizing behavior	0.68	0.68	0.00	0.02	0.00	0.966
Attention to schoolwork	3.45	3.46	-0.01	0.02	-0.01	0.599
Pro-social behavior	3.90	3.89	0.01	0.03	0.01	0.801
Number of schools	58	31				
Program Year 2						
Disruptive behavior	0.56	0.58	-0.02	0.02	-0.02	0.343
Emotional dysregulation	1.14	1.16	-0.02	0.03	-0.02	0.450
Internalizing behavior	0.68	0.71	-0.03	0.02	-0.04	0.220
Attention to schoolwork	3.45	3.43	0.01	0.03	0.01	0.627
Pro-social behavior	3.93	3.88	0.04	0.03	0.04	0.183
Number of schools	58	31				

(continued)

Exhibit C.1 (continued)

SOURCES: Teacher Survey of Student Behavior data, collected in fall 2015, spring 2016 (Program Year 1, n = 25,391), and spring 2017 (Program Year 2, n = 24,842). Student records data from the 2014-2015 school year.

NOTES: The student sample used in this table is defined as students with all five behavior measures for that year.

The impacts are estimated using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit 4 in the report shows that the program had no effect on students' reading and math achievements for Grade 3-5 students, on average, in both program years as well as the follow-up year. Exhibit C.2 provides the details of these findings.

Exhibit C.2. Estimated Impacts on Student Academic Achievement, Overall Sample, by Program Year

Measures	Participating Schools	Non-Participating Schools	Estimated Impact	Standard Error of Estimated Impact	p-value of Estimated Impact	Number of Observations
Program Year 1						
Standardized reading scores	0.03	0.02	0.01	0.03	0.581	24,326
Standardized math scores	0.02	0.02	0.00	0.03	0.985	24,326
Program Year 2						
Standardized reading scores	0.04	0.04	0.00	0.03	0.891	23,569
Standardized math scores	0.04	0.04	0.00	0.04	0.981	23,569
Follow-Up Year						
Standardized reading scores	0.00	0.00	0.01	0.03	0.835	19,968
Standardized math scores	0.02	0.00	0.02	0.04	0.569	19,968

(continued)

66

Exhibit C.2 (continued)

SOURCES: District records data for spring 2016, spring 2017, and spring 2018. Teacher Survey of Student Behavior data from fall 2015.

NOTES: For Program Years 1 and 2, the analysis includes students from 58 participating schools and 31 non-participating schools. For the Follow-Up Year, the analysis includes students from 52 participating schools and 28 non-participating schools. The analysis sample used in this table is defined as students with both Reading and Math state test scores for that year.

The impacts are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

All test scores are standardized within school districts and grade levels using the means and standard deviations of the full non-participating group of students in each school-by-grade cell. The values in the estimated impact column are therefore in effect size unit.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit 5 in the report presents the program effects for both program years on the behavior of students who were initially identified as struggling with behavior. Exhibit C.3 shows the program affected disruptive behavior but did not affect other measured behaviors for these students. For comparison, this exhibit also includes the estimated program effects on the behavior of students who were not initially identified as struggling with behavior and provides statistical tests for the differences in the estimated impacts for these two groups of students. To provide context for the actions of the students initially identified as struggling with behavior, Exhibit C.4 provides information on the average behavior ratings for them at the start of the program (in the fall of 2015).

Exhibit 6 in the report shows that the program positively affected the reading achievements of students initially identified as struggling with behavior during the two years of the program, but the effect was not sustained in the follow-up year. Exhibit C.5 provides the details of the estimated program effects on academic achievement for these students initially identified as struggling with behavior and their counterparts who were not identified as struggling with behavior. It also provides comparisons between the estimated impacts for these two groups of students.

Program Effects on Additional Student Outcomes[34]

For context, the study team also examined program effects on other student behavior outcomes using data collected in the Teacher Survey of Student Behavior at the end of each program year. First, the study assessed the program's effects on student behavior incidences, in other words, the occurrences of disciplinary actions experienced by a student such as office disciplinary referrals. Teacher reports of office disciplinary referrals have been shown in other studies to be a reliable measure of actual referrals to the office and to be correlated with disruptive student behaviors.[35] Exhibit C.6 shows that the program did not affect overall student receipt of referrals or suspensions in either program year. Exhibit C.7 shows the same result for students initially identified as struggling with behavior.

Exhibit C.3. Estimated Impacts on Teacher Ratings of Student Behavior by Whether Students Were Initially Identified as Struggling with Behavior, by Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							0.04	0.026 †	
Students initially identified as struggling with behavior	1.64	1.69	-0.06 *	0.02	-0.07	0.010			
Students not initially identified	0.38	0.39	-0.01	0.01	-0.02	0.413			
Emotional dysregulation							0.01	0.632	
Students initially identified as struggling with behavior	2.20	2.22	-0.02	0.03	-0.02	0.477			
Students not initially identified	0.95	0.96	-0.01	0.02	-0.01	0.672			
Internalizing behavior							0.00	0.904	
Students initially identified as struggling with behavior	1.09	1.10	0.00	0.03	0.00	0.944			
Students not initially identified	0.59	0.59	0.00	0.02	-0.01	0.804			

(continued)

Exhibit C.3 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Attention to schoolwork							-0.03	0.348	
Students initially identified as struggling with behavior	2.43	2.41	0.02	0.03	0.01	0.627			
Students not initially identified	3.67	3.68	-0.01	0.02	-0.01	0.544			
Pro-social behavior							-0.02	0.368	
Students initially identified as struggling with behavior	2.92	2.87	0.04	0.03	0.04	0.212			
Students not initially identified	4.10	4.08	0.02	0.02	0.02	0.483			
Students initially identified as struggling with behavior (n)									2,979
Students not initially identified (n)									17,208
Program Year 2									
Disruptive behavior							0.07	0.022 †	
Students initially identified as struggling with behavior	1.17	1.25	-0.08 *	0.03	-0.11	0.015			
Students not initially identified	0.41	0.43	-0.01	0.02	-0.02	0.551			

(continued)

Exhibit C.3 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Emotional dysregulation							0.06	0.081	
Students initially identified as struggling with behavior	1.74	1.81	-0.06	0.04	-0.07	0.143			
Students not initially identified	0.99	0.99	0.00	0.03	0.00	0.976			
Internalizing behavior							0.04	0.197	
Students initially identified as struggling with behavior	0.86	0.93	-0.07	0.04	-0.09	0.090			
Students not initially identified	0.63	0.65	-0.03	0.03	-0.04	0.383			
Attention to schoolwork							-0.03	0.499	
Students initially identified as struggling with behavior	2.70	2.68	0.02	0.05	0.02	0.662			
Students not initially identified	3.64	3.65	-0.01	0.04	-0.01	0.781			

(continued)

Exhibit C.3 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Pro-social behavior							-0.05	0.170	
Students initially identified as struggling with behavior	3.30	3.23	0.07	0.05	0.07	0.156			
Students not initially identified	4.08	4.06	0.01	0.04	0.02	0.692			
Students initially identified as struggling with behavior (n)									1,847
Students not initially identified (n)									11,183

SOURCES: Teacher Survey of Student Behavior data, collected in fall 2015, spring 2016, and spring 2017. Student records data from the 2014-2015 school year.

NOTES: The analysis includes students from 58 participating schools and 31 non-participating schools. The sample for each program year includes students with all five behavior measures for that year. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools.

   The impacts are estimated using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

   The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

   The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating group members in the analysis sample.

   A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

   An F-test was applied to the differences in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit C.4. Teacher Ratings of Student Behavior at the Start of the Study (Fall 2015) for Students Who Were in Program Years 1 and 2 Samples and Were Initially Identified as Struggling with Behavior

Sample and Baseline Behavior Ratings (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Program Year 1						
Disruptive behavior	1.77	1.78	-0.01	0.03	-0.01	0.767
Emotional dysregulation	2.29	2.24	0.05	0.04	0.06	0.230
Internalizing behavior	1.15	1.14	0.00	0.04	0.01	0.908
Pro-social behavior	2.80	2.76	0.04	0.04	0.05	0.364
Attention to schoolwork	2.31	2.31	0.01	0.05	0.01	0.890
Program Year 2						
Disruptive behavior	1.75	1.80	-0.05	0.03	-0.08	0.150
Emotional dysregulation	2.26	2.26	0.01	0.05	0.01	0.867
Internalizing behavior	1.16	1.18	-0.02	0.05	-0.03	0.688
Pro-social behavior	2.83	2.80	0.03	0.05	0.04	0.601
Attention to schoolwork	2.36	2.32	0.04	0.06	0.04	0.525

SOURCES: Teacher Survey of Student Behavior from the fall of 2015, spring of 2016, and spring of 2017, and district records data collected for the 2014-2015 and 2015-2016 school years.

NOTES: The analysis includes students from 58 participating schools and 31 non-participating schools. The student sample used in this table is defined as students with all five behavior measures for that program year. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 1 sample has 2,979 students, and the Program Year 2 sample has 1,847 students. The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the difference between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed test.

Exhibit C.5. Estimated Impacts on Student Academic Achievement by Whether Students Were Initially Identified as Struggling with Behavior, by Program Year

Measures	Participating Schools	Non-Participating Schools	Estimated Impact	Standard Error of Estimated Impact	p-value of Estimated Impact	Estimated Subgroup Difference	p-value of Estimated Subgroup Difference	Number of Observations
Program Year 1								
Standardized reading scores						0.13	0.006 †	
Students initially identified as struggling with behavior	-0.16	-0.27	0.11 *	0.05	0.023			
Students not initially identified	0.14	0.15	-0.02	0.03	0.579			
Standardized math scores						0.07	0.107	
Students initially identified as struggling with behavior	-0.18	-0.23	0.05	0.05	0.296			
Students not initially identified	0.13	0.15	-0.02	0.03	0.576			
Students initially identified as struggling with behavior (n)								1,879
Students not initially identified (n)								10,993
Program Year 2								
Standardized reading scores						0.14	0.004 †	
Students initially identified as struggling with behavior	-0.14	-0.25	0.11 *	0.05	0.050			
Students not initially identified	0.16	0.20	-0.04	0.03	0.256			
Standardized math scores						0.09	0.058	
Students initially identified as struggling with behavior	-0.17	-0.24	0.07	0.06	0.209			
Students not initially identified	0.17	0.19	-0.02	0.04	0.680			
Students initially identified as struggling with behavior (n)								1,595
Students not initially identified (n)								9,701

(continued)

Exhibit C.5 (continued)

Measures	Participating Schools	Non-Participating Schools	Estimated Impact	Standard Error of Estimated Impact	p-value of Estimated Impact	Estimated Subgroup Difference	p-value of Estimated Subgroup Difference	Number of Observations
Follow-Up Year								
Standardized reading scores						0.14	0.022 †	
Students initially identified as struggling with behavior	-0.23	-0.32	0.09	0.07	0.173			
Students not initially identified	0.15	0.20	-0.05	0.04	0.253			
Standardized math scores						0.02	0.762	
Students initially identified as struggling with behavior	-0.22	-0.21	-0.02	0.07	0.817			
Students not initially identified	0.19	0.22	-0.03	0.05	0.479			
Students initially identified as struggling with behavior (n)								1,323
Students not initially identified (n)								8,165

SOURCES: District records data for spring 2016 and spring 2017. Teacher Survey of Student Behavior data from fall 2015.

NOTES: For Program Years 1 and 2, the analysis includes students from 58 participating schools and 31 non-participating schools. For the Follow-Up Year, the analysis includes students from 52 participating schools and 28 non-participating schools. The analysis sample used in this table is defined as students with both Reading and Math state test scores for that year. A student is initially identified as struggling with behavior if his/her baseline disruptive behavior rating is in the highest 15 percent among students in the same grade across all sample schools.

    The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

    The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using the number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

    All test scores are standardized within school districts and grade levels using the means and standard deviations of the full non-participating schools' students in each school-by-grade cell. The values in the estimated impact column are therefore in effect size units.

    A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

    An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit C.6. Estimated Impacts on Other Behavior-Related Outcomes as Reported by Teachers, Overall Sample, by Program Year

Measures (%)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Program Year 1						
Disciplinary referral	14.2	14.0	0.2	0.82	0.01	0.791
Currently receiving SPED	10.2	10.4	-0.2	0.55	-0.01	0.665
Referred to SPED Assessment	5.7	6.1	-0.4	0.55	-0.02	0.486
Suspensions						
In-school suspension	3.3	3.4	-0.1	0.48	-0.01	0.817
Out-of-school suspension	2.0	1.8	0.3	0.32	0.02	0.425
Either	4.1	3.9	0.2	0.50	0.01	0.732
Receipt of targeted support						
Academic support	28.7	29.4	-0.6	1.52	-0.01	0.675
Behavioral support	4.5	4.2	0.3	0.54	0.01	0.582
Psychological support/counseling	7.3	6.0	1.3	0.73	0.06	0.071
Program Year 2						
Disciplinary referral	13.2	12.4	0.8	0.81	0.02	0.317
Currently receiving SPED	10.0	9.8	0.2	0.59	0.01	0.689
Referred to SPED Assessment	5.7	6.6	-0.9	0.56	-0.04	0.112
Suspensions						
In-school suspension	2.9	3.0	-0.1	0.43	-0.01	0.777
Out-of-school suspension	1.6	1.5	0.1	0.28	0.01	0.752
Either	3.5	3.4	0.1	0.43	0.00	0.856

(continued)

Exhibit C.6 (continued)

Measures (%)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Receipt of targeted support						
Academic support	30.4	33.8	-3.4	1.79	-0.07	0.054
Behavioral support	5.9	4.4	1.5 *	0.53	0.07	0.004
Psychological support/counseling	7.2	7.8	-0.7	0.70	-0.03	0.334

SOURCES: Teacher Survey of Student Behavior data, collected in spring 2016 (Program Year 1, n = 25,391), and spring 2017 (Program Year 2, n = 24,842). Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The analysis sample used in this table is defined as students with all five behavior measures for that year. It includes students from 58 participating schools and 31 non-participating schools. Sample sizes vary across measures due to missing data.

The estimated impacts are estimated using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviations of all non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.7. Estimated Impacts on Other Behavior-Related Outcomes as Reported by Teachers for Students Initially Identified as Struggling with Behavior, by Program Year

Measures (%)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Program Year 1						
Disciplinary referral	47.8	48.8	-1.0	2.77	-0.02	0.712
Currently receiving SPED	16.3	16.8	-0.5	1.30	-0.01	0.705
Referred to SPED Assessment	11.3	11.3	0.0	1.56	0.00	0.997
Suspensions						
In-school suspension	13.2	15.4	-2.1	1.95	-0.06	0.271
Out-of-school suspension	8.1	7.9	0.2	1.42	0.01	0.880
Either	15.5	16.9	-1.3	1.97	-0.03	0.504
Receipt of targeted support						
Academic support	36.3	34.3	2.0	2.57	0.04	0.441
Behavioral support	13.9	12.6	1.3	1.61	0.04	0.415
Psychological support/counseling	19.6	16.4	3.2	2.44	0.09	0.188
Program Year 2						
Disciplinary referral	36.4	37.4	-1.0	2.82	-0.02	0.721
Currently receiving SPED	17.4	16.5	0.9	1.61	0.02	0.598
Referred to SPED Assessment	8.7	10.0	-1.2	1.61	-0.04	0.444
Suspensions						
In-school suspension	10.0	10.5	-0.4	2.17	-0.01	0.839
Out-of-school suspension	5.9	6.1	-0.2	1.58	-0.01	0.893
Either	12.3	12.4	0.0	2.16	0.00	0.984

(continued)

Exhibit C.7 (continued)

Measures (%)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Receipt of targeted support						
Academic support	37.0	37.8	-0.8	3.05	-0.02	0.803
Behavioral support	19.0	12.8	6.2 *	2.47	0.19	0.012
Psychological support/counseling	17.1	18.9	-1.8	2.58	-0.05	0.491

SOURCES: Teacher Survey of Student Behavior data, collected in spring 2016 (Program Year 1, n = 2,979), and spring 2017 (Program Year 2, n = 1,847). Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The analysis sample used in this table is defined as students with all five behavior measures for that year. It includes students from 58 participating schools and 31 non-participating schools. Sample sizes vary across measures due to missing data.

The estimated impacts are estimated using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviations of all non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Second, the study estimated the program's effects on the identification of students for special education services related to behavior. The program could potentially reduce such identifications if it had positive effects on student behaviors. The study assessed the program's effects on teachers' reports of students that were referred to be assessed for special education services. It is worth noting that this referral rate can be influenced by many factors other than student behavior. Such factors include, but are not limited to, other learning-related disabilities, policy regarding special education referrals, and school policies regarding when and how teachers make referrals to special education. The study also assessed the receipt of special education services. Exhibit C.6 shows that the program had no effect on teachers' reports of students' referrals for special education assessment or receipt of special education services in either program year. Exhibit C.7 shows the same pattern for the students initially identified as struggling with behavior.

Additionally, the study team analyzed the program's effects on the proportion of students receiving supplemental support services. All elementary schools need to have mechanisms for providing additional support services to students identified as struggling with behavior, academics, and/or mental health. The program's training and technical assistance for supplemental supports in Program Year 2 intended to provide students who needed it with access to a supplemental support intervention–the Check In Check Out (CICO) program. As a result, the study explored whether the program increased the proportion of students receiving behavior support services beyond the levels of elementary schools not participating in the program. Thus, the study team asked teachers in participating and non-participating schools if each of their students had received any additional supports or services related to academics, behavior, or psychological counseling.

Exhibits C.6 and C.7 report the program's effects on these student outcomes for the overall student sample and for students initially identified as struggling with behavior, respectively. Exhibit C.6 shows that for students overall, the program did not produce an effect on students' receipt of supplemental supports for behavior in the first year but did in the program's second year after that aspect of the program was introduced to school staff. Exhibit C.7 shows this same pattern of findings for students initially identified as struggling with behavior.

Program Effects on School Climate

This section includes supplemental information for the program's effects on school climate. In particular, Exhibit 8 in the report shows that the program changed staff perceptions of certain aspects of school climate. Exhibit C.8 provides details of these findings.

Program Effects on Classroom Outcomes

Exhibits 9 and 10 in the report show that the program had positive effects on how teachers managed their classrooms and how classrooms functioned. Exhibits C.9 and C.10 present detailed findings of the program's effects on these classroom outcomes.

Differences Between Participating and Non-Participating Schools in Receipt of Professional Development and Use of MTSS-B Practices

This section summarizes the findings on the differences between participating and non-participating schools in school staff's receipt of professional development, implementation of school-wide support practices, and implementation of supplemental supports.

Differences in receipt of professional development. The staff survey, conducted in the spring of Program Year 2, asked the teachers to report their participation in various professional development activities related to student behavior for Program Year 1 and the current school year (Program Year 2). It also asked the teachers to report the number of hours spent in professional development activities in Program Year 2. Analyses based on these data demonstrated that the program produced positive and significant differences in teacher-reported participation in professional development activities in Program Years 1 and 2. However, staff in non-participating schools also reported participating in some professional development and coaching related to student behavior (Exhibit C.11). In the Follow-Up Year when the study's training and technical assistance program had ended, such differences, especially the difference for the training on school-wide support practices, were limited according to the school site visit data collected in the Follow-Up Year (Exhibit C.12).

Exhibit C.8. Estimated Impacts on Staff Perceptions of School Climate, Program Year 2

Measures (1- to 4-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Organizational health						
Academic focus	2.78	2.71	0.07 *	0.04	0.14	0.044
Principal leadership	3.17	3.10	0.07	0.08	0.09	0.386
Staff collegiality	3.30	3.17	0.14 *	0.06	0.21	0.025
School environment						
Bullying	2.16	2.20	-0.04	0.05	-0.05	0.473
School safety	3.35	3.33	0.02	0.03	0.04	0.485
Teacher-student relationships	3.58	3.51	0.07 *	0.03	0.15	0.016
Consistent school discipline	3.12	3.02	0.10	0.06	0.14	0.114
Number of schools	58	31				

SOURCE: MTSS-B Staff Survey data collected in spring 2017 (n = 4,280).

NOTES: The analysis sample used for the survey is defined as all respondents who consented to and answered at least one question in the survey. The number of respondents varies due to missingness in the constructs. A respondent is counted as non-missing if they responded to at least 80 percent of the questions that constitute the scale.

  The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with staff nested within schools. The models control for the blocking of random assignment and for baseline differences between staff members in the participating schools and non-participating schools with respect to the following variables: staff members' age, gender, years of experience in the district, and years of experience in the school. All missing covariate values are imputed with zero and missing indicators for all covariates are also included in the model.

  The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

  The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

  A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Exhibit C.9. Estimated Impacts on Classroom Management Practices, Program Year 2

Measures of Teacher Practices (1- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Teacher active monitoring of students	4.56	4.45	0.11 *	0.04	0.14	0.017
Teacher proactive behavior management	3.85	3.52	0.33 *	0.07	0.36	0.000
Teacher anticipation and responsiveness to student needs	4.22	4.05	0.17 *	0.06	0.22	0.004
Facilitate orderly classroom transitions	4.17	4.03	0.14 *	0.06	0.12	0.029
Sample size (classrooms)	1,152	639				

SOURCE: Classroom Observations data, collected with Assessing School Settings: Interactions of Students and Teachers (ASSIST) in the spring of 2017.

NOTES: The analysis includes students from 57 participating schools and 31 non-participating schools. The classroom sample used in this table is defined as all non-special education classes in Grades 1-5 in sample schools.

The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with classes nested within schools. The models control for the blocking of random assignment and for baseline differences between the classrooms in the participating and non-participating schools with respect to the following variables: time of day of the observation (am/pm), observer indicators, grade level indicators, class size, gender of teacher, and number of adults in the class. All missing values are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for classrooms from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.10. Estimated Impacts on Classroom Functioning, Program Year 2

Measures (1- to 5-point scale unless otherwise noted)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts
Counts of student disruptive behaviors in the classroom (tally)	6.96	8.88	-1.92 *	0.54	-0.21	0.001
Student compliance	4.43	4.30	0.14 *	0.04	0.19	0.003
Student engagement with classroom activities	3.63	3.48	0.16 *	0.06	0.16	0.015
Teacher control of the classroom	4.67	4.56	0.11 *	0.04	0.19	0.009
Student disruptive behaviors	1.19	1.24	-0.04 *	0.02	-0.16	0.020
Sample size (classrooms)	1,152	639				

SOURCE: Classroom Observations data, collected with Assessing School Settings: Interactions of Students and Teachers (ASSIST) in spring 2017.

NOTES: The analysis includes students from 57 participating schools and 31 non-participating schools. The classroom sample used in this table is defined as all non-special education classes in Grades 1-5 in sample schools.

The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with classes nested within schools. The models control for the blocking of random assignment and for baseline differences between the classrooms in the participating and non-participating schools with respect to the following variables: time of day of the observation (am/pm), observer indicators, grade level indicators, class size, gender of teacher, and number of adults in the class. All missing values are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for classrooms from the MTSS-B participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.11. Estimated Differences in Professional Development for Behavior Support Practices Between Participating and Non-Participating Schools, Program Years 1 and 2

Measures	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Professional development for behavior support practices						
Percentage participated in any PD (Program Year 1)	81.96	58.70	23.26 *	2.33	0.47	0.000
Percentage participated in any PD (Program Year 2)	72.23	39.96	32.28 *	3.41	0.66	0.000
Estimated hours in PD (Program Year 2)	21.70	12.19	9.52 *	1.47	0.44	0.000
Percentage of teachers receiving training from a coach (Program Year 2)	58.85	34.54	24.31 *	2.82	0.51	0.000

SOURCE: MTSS-B Staff Survey fielded in spring 2017.

NOTES: The analysis sample used for the survey is defined as all instructional staff members who consented to and answered at least one question in the survey (n = 2,436). The number of respondents varies due to missingness in the constructs. A respondent is counted as non-missing if they responded to at least 80 percent of the questions that constitute the scale.

The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with teachers nested within schools. The model controls for the blocking of random assignment and for baseline differences between staff members in the participating and non-participating schools with respect to the following variables: age, gender, years of experience in the school, and years of experience in the district. All missing covariate values are imputed with zero and missing indicators for all covariates are included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for staff members from the participating schools (using the number of participating schools in each district as weight). The non-participating schools' values in the next column are the regression-adjusted means using the observed mean covariate values for the participating schools as the basis for the adjustment. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated difference effect sizes are calculated as a proportion of the standard deviation of the full non-participating schools' members in the analysis sample.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Exhibit C.12. Differences in Training and Coaching Receipt for Remaining Study Schools Between Participating and Non-Participating Schools, Follow-Up Year

Measures (%)	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Impacts	Estimated Difference in Effect Size	p-value of Estimated Difference	Number of Observations
School-wide practices (Tier I) training							
Schools receiving any Tier I training	78.4	63.1	15.3	10.1	0.3	0.137	
Schools receiving at least a full day of Tier I training	21.6	17.9	3.7	9.6	0.1	0.700	
Schools in which "all staff" received Tier I training	51.0	38.0	13.0	11.2	0.3	0.252	
Supplemental support (Tier II) training							
Schools receiving any Tier II training	62.7	31.0	31.7 *	10.7	0.7	0.005	
Schools receiving at least a full day of Tier II training	13.7	9.8	3.9	7.0	0.1	0.579	
Schools in which "all staff" received Tier II training	33.3	10.8	22.5 *	10.2	0.7	0.031	
MTSS-B coaching							
Schools receiving support from behavior support coach	60.8	35.8	24.9 *	9.5	0.5	0.011	
Schools with no behavior support coach	39.2	64.2	-24.9 *	9.5	-0.5	0.011	
Schools receiving support from coach less than once a month	17.6	18.6	-1.0	9.6	0.0	0.918	
Schools receiving support from coach more than once a month but less than once a week	27.5	9.4	18.0 *	7.6	0.6	0.020	
Schools receiving support from coach once a week or more	7.8	7.6	0.3	6.1	0.0	0.967	
Other (e.g., varies or as-needed support)	9.8	0.5	9.3	6.2	NA	0.135	

(continued)

84

Exhibit C.12 (continued)

SOURCE: MTSS-B site visits (supplemental questions to SET/I-SSET) in spring 2018 for a total of 79 schools (51 participating schools, 28 non-participating schools).

NOTES: The values in the column labeled "Participating Schools" are the weighted average of the observed district means from the participating schools (using the number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

    The estimated difference effect sizes are calculated as a proportion of the standard deviation of the full non-participating schools in the analysis sample.

    A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

    NA: Not applicable because the unadjusted standard deviation for this variable across non-participating schools is zero.

Differences in implementing school-wide behavior support practices and supplemental supports aligned with MTSS-B. The program produced a positive and significant contrast on overall implementation fidelity of the *school-wide practices* in Program Years 1 and 2 (first row of the top panel in Exhibit C.13 and C.14), which continued into the Follow-Up Year after the training and technical assistance ended (Exhibit C.15). As shown in the remaining rows of the top panel, except for "systems for responding to behavioral violations", the program also produced a positive and significant difference in Program Years 1 and 2 on each of the core components of school-wide practices that contribute to the implementation fidelity score: expectations defined and taught, system for rewarding behavior meeting expectations, monitoring and management and district level support.[36] These differences continued into the Follow-Up Year, when a difference in systems for responding to behavioral violations did appear.

Participating schools were trained in the *supplemental support practices* in Program Year 2. As expected, the program did not produce a statistically significant difference on the implementation of supplemental supports in Program Year 1 (first row of lower panel of Exhibit C.13) but it did produce a positive and statistically significant difference on implementation of supplemental supports in Program Year 2 (Exhibit C.14). The study's MTSS-B training and technical assistance to participating schools finished at the end of Program Year 2, but the program continued to produce a positive and significant difference between participating and non-participating schools on the implementation of supplemental supports in the Follow-Up Year (Exhibit C.15).

While these results show that the program produced large differences between the participating and non-participating schools in the implementation of school-wide behavior support practices and supplemental supports aligned with MTSS-B, they also demonstrate that the non-participating schools were implementing elements of MTSS-B at noticeable levels in each program year and the year afterward.

Relationship Between Effects on School, Classroom, and Student Outcomes

This section presents findings on the relationship between the effects on implementation fidelity, school and classroom outcomes, and student outcomes. This analysis focused on two key student outcomes: students' disruptive behavior ratings and students' reading achievements because they are outcomes for which there were impacts–at least for students initially identified as struggling with behavior. For each student outcome, the study examined the separate relationship between the program's effect on fidelity, school climate, classroom management practices, classroom functioning, and student outcomes. It also looked at the relationship between the effects on student behavior ratings and that on students' reading achievements. As described in Appendix B, this kind of analysis assessed the relationship between impacts on each school or classroom outcome and the student impacts separately. However, a key limitation of this analytic approach is that the small number of random assignment blocks in the study limited the statistical power of the estimation. The non-significant findings could be the result of limited sample size, no association between the factors, or both.

Given the logic of how the MTSS-B approach is expected to work, one would expect that places where implementation of the program made more of a difference in schools' behavioral support practices, or where the program affected school climate or classroom practices and functioning more, would also have had more favorable impacts on student behavior and achievement. However, the correlational analysis by and large failed to detect such relationships for the overall student sample (see Exhibit C.16). There were suggestions that the program impact on teachers' proactive management practices might be related to the program impact on disruptive behavior (p-value = 0.082), and the impact on collegiality among school staff might be associated with the impact on reading achievements (p-value = 0.058), but there was no sign of any other significant relationship between the mediational factors and the student outcomes examined here.

Exhibit C.13. Implementation of School-Wide Behavior Support Practices and Supplemental Supports in Participating and Non-Participating Schools, Program Year 1

Measures	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
School-wide practices (Tier I)[a]						
% of schools implementing Tier I with fidelity	69.0	0.6	68.3 *	7.9	3.80	0.000
Mean total Tier I fidelity score	83.2	56.9	26.4 *	2.1	1.40	0.000
Expectations defined	88.8	60.5	28.3 *	3.8	1.02	0.000
Behavior expectations taught	86.2	54.1	32.1 *	3.5	1.28	0.000
Ongoing system for rewarding expectations	79.3	52.1	27.2 *	5.6	0.82	0.000
Systems for responding to behavioral violations	75.9	73.2	2.7	3.3	0.13	0.420
Monitoring & decision making	93.5	70.3	23.3 *	3.9	0.69	0.000
Management	81.5	48.2	33.3 *	4.4	1.03	0.000
District-level support	76.7	31.7	45.0 *	4.9	1.35	0.000
Supplemental supports (Tier II)[b]						
% of schools implementing Tier II with fidelity	22.4	14.9	7.5	9.1	0.20	0.414
Mean total Tier II fidelity score	57.8	51.0	6.8	4.5	0.30	0.138
Foundations	64.5	61.4	3.0	3.5	0.19	0.388
Supplemental support intervention	51.1	40.5	10.6	7.7	0.27	0.172

SOURCES: Author's calculations based on:

[a]MTSS-B School-wide Evaluation Tool (SET) fielded in the fall of 2015.

[b]Individual Student Systems Evaluation Tool (I-SSET) fielded in the fall of 2015.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The estimated differences are regression-adjusted using Ordinary Least Squares (OLS) models that control for the blocking of random assignment. The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the program group (using number of program schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.14. Implementation of School-Wide Behavior Support Practices and Supplemental Supports in Participating and Non-Participating Schools, Program Year 2

Measures	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
School-wide practices (Tier I)[a]						
% of schools implementing Tier I with fidelity	79.3	23.1	56.2 *	7.8	1.32	0.000
Mean total Tier I fidelity score	86.0	54.7	31.2 *	2.1	1.45	0.000
Expectations defined	91.8	54.3	37.5 *	3.4	1.25	0.000
Behavior expectations taught	91.2	42.6	48.6 *	3.0	1.93	0.000
Ongoing system for rewarding expectations	87.4	50.9	36.5 *	4.2	1.09	0.000
Systems for responding to behavioral violations	80.0	78.8	1.2	3.3	0.06	0.712
Monitoring & decision making	94.2	66.0	28.2 *	3.8	0.90	0.000
Management	82.1	45.6	36.5 *	4.6	1.09	0.000
District-level support	75.0	45.0	30.0 *	4.6	0.86	0.000
Supplemental supports (Tier II)[b]						
% of schools implementing Tier II with fidelity	74.1	25.3	48.9 *	8.6	1.15	0.000
Mean total Tier II fidelity score	81.6	58.9	22.7 *	3.2	0.91	0.000
Foundations	75.9	65.1	10.8 *	3.1	0.50	0.001
Supplemental support intervention	87.2	52.6	34.6 *	5.3	0.87	0.000

SOURCES: Author's calculations based on

[a]MTSS-B School-wide Evaluation Tool (SET) fielded in the spring of 2017.

[b]Individual Student Systems Evaluation Tool (I-SSET) fielded in the spring of 2017.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The estimated differences are regression-adjusted using Ordinary Least Squares (OLS) models that control for the blocking of random assignment. The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the program group (using number of program schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.15. Implementation of School-Wide Behavior Support Practices and Supplemental Supports in Participating and Non-Participating Schools, Follow-Up Year

Measures	Participating Schools	Non-Participating Schools	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
School-wide practices (Tier I)[a]						
% of schools implementing Tier I with fidelity	76.5	12.2	64.3 *	9.5	1.80	0.000
Mean total Tier I fidelity score	84.6	58.2	26.4 *	3.1	1.10	0.000
Expectations defined	92.2	60.4	31.8 *	4.3	0.97	0.000
Behavior expectations taught	89.8	50.0	39.8 *	3.8	1.38	0.000
Ongoing system for rewarding expectations	87.3	56.1	31.2 *	5.0	0.91	0.000
Systems for responding to behavioral violations	85.7	73.9	11.8 *	4.3	0.68	0.008
Monitoring & decision making	92.4	72.4	20.0 *	3.8	0.61	0.000
Management	71.4	49.1	22.4 *	4.8	0.66	0.000
District-level support	73.5	45.4	28.1 *	8.0	0.64	0.001
Supplemental supports (Tier II)[b]						
% of schools implementing Tier II with fidelity	76.5	34.4	42.0 *	10.2	0.86	0.000
Mean total Tier II fidelity score	83.4	63.2	20.2 *	4.1	0.73	0.000
Foundations	76.4	66.6	9.8 *	3.6	0.48	0.007
Supplemental support intervention	90.4	59.8	30.6 *	6.1	0.75	0.000

SOURCES: Author's calculations based on:

[a]MTSS-B School-wide Evaluation Tool (SET) fielded in the spring of 2018.

[b]Individual Student Systems Evaluation Tool (I-SSET) fielded in the spring of 2018.

NOTES: The sample includes 51 participating schools and 28 non-participating schools. The estimated differences are regression-adjusted using Ordinary Least Squares (OLS) models that control for the blocking of random assignment. The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the program group (using number of program schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.16. Relationship Between Program Impacts on Selected Student Outcomes and Program Impacts on School or Classroom Outcomes, for Overall Sample, Program Year 2

Explanatory Variable	Teacher Rating of Student Behavior, Disruptive Behavior		Student Academic Achievement, Standardized Reading Test Score	
	Estimate	p-value	Estimate	p-value
Implementation fidelity				
Tier I: SET total score	-0.02	0.607	0.09	0.107
Tier II: I-SSET total score	-0.06	0.276	0.07	0.219
School climate				
Academic focus	-0.05	0.653	0.17	0.126
Principal leadership	-0.05	0.579	0.03	0.714
Staff collegiality	-0.09	0.272	0.15	0.058
School safety	-0.17	0.190	0.19	0.157
Teacher-student relationships	-0.09	0.465	0.22	0.120
Consistent school discipline	-0.13	0.156	0.12	0.218
Classroom management practices				
Teacher proactive behavioral management	-0.17	0.082	0.07	0.426
Teacher active monitoring of students	-0.09	0.354	0.12	0.211
Teacher anticipation and responsiveness to student needs	-0.12	0.347	0.09	0.439
Facilitate orderly classroom transitions	-0.08	0.678	0.11	0.579
Classroom functioning				
Counts of student disruptive behaviors in classroom (tally)	0.08	0.389	-0.07	0.502
Student compliance	-0.11	0.313	0.11	0.282
Student engagement with classroom activities	-0.11	0.391	0.13	0.336
Teacher control of classroom	-0.15	0.150	0.06	0.531
Student disruptive behaviors	0.11	0.222	-0.03	0.671
Student behavior rating				
Disruptive behavior			-0.04	0.886
Emotional dysregulation			0.00	0.994
Internalizing behavior			0.05	0.804
Attention to schoolwork			0.44	0.255
Pro-social behavior			0.24	0.313

(continued)

Exhibit C.16 (continued)

SOURCES: District Records for the 2016-2017 school year; Teacher Ratings of Student Behavior collected in spring 2017; Classroom Observations data collected with Assessing School Settings: Interactions of Teachers and Students (ASSIST) in spring 2017; Staff Survey data collected in spring 2017; School-wide Evaluation Tool (SET) and Individual Student Systems Evaluation Tool (I-SSET) fielded in spring 2017.

NOTES: The analysis sample is all eligible students in the study schools who have non-missing values for the outcomes and the explanatory variables used in a given model.

The analysis uses a three-step procedure to estimate the relationship between a given outcome and a given explanatory variable. The first step uses a hierarchical model to estimate the block-level program impacts on the outcome. This model controls for the blocking of random assignment and for baseline covariates appropriate to the outcome. The second step uses a hierarchical model to estimate the block-level program impacts on the explanatory variable. This model controls for the blocking of random assignment and for baseline covariates appropriate to the variable. The last step regresses the block-level impacts for the outcome on those for the explanatory variable using a precision-weighted random-effect bivariate model. The coefficient for the explanatory variable impact reflects the amount of outcome change that is associated with one unit of change in the explanatory variable.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Findings for the group of students initially identified as struggling with behavior showed more support for the logic of how the program is expected to work (see Exhibit C.17). For example, the analysis found several significant correlations between impacts on school climate, classroom practices, and classroom functioning and impacts on student disruptive behavior for this student subgroup. Even though the correlations between impacts on school and classroom outcomes and students' reading achievement were not statistically significant at the 5 percent level, there were suggestions that better implementation of the supplemental support practices and larger impacts on students' attention to schoolwork might be related to impacts on students' reading achievements (p-values were 0.083 and 0.061, respectively).

Exhibit C.17. Relationship Between Program Impacts on Selected Student Outcomes and Program Impacts on School or Classroom Outcomes, for Students Initially Identified as Struggling with Behavior, Program Year 2

Explanatory Variable	Teacher Rating of Student Behavior, Disruptive Behavior		Student Academic Achievement, Standardized Reading Test Score	
	Estimate	p-value	Estimate	p-value
Implementation fidelity				
SET	-0.20	0.104	0.03	0.711
I-SSET	-0.41 *	0.012	0.16	0.083
School climate				
Academic focus	-0.61 *	0.042	0.11	0.509
Principal leadership	-0.17	0.403	0.00	0.978
Staff collegiality	-0.38 *	0.025	0.13	0.188
School safety	-0.47 *	0.042	0.01	0.951
Teacher-student relationships	-0.21	0.414	0.05	0.741
Consistent school discipline	-0.52 *	0.024	-0.01	0.948

(continued)

Exhibit C.17 (continued)

Explanatory Variable	Teacher Rating of Student Behavior, Disruptive Behavior		Student Academic Achievement, Standardized Reading Test Score	
	Estimate	p-value	Estimate	p-value
Classroom management practices				
Teacher proactive behavior management	-0.57 *	0.033	0.22	0.154
Teacher active monitoring of students	-0.50	0.068	0.13	0.421
Teacher anticipation and responsiveness to student needs	-0.58	0.076	0.04	0.812
Facilitate orderly classroom transitions	-0.15	0.760	0.11	0.670
Student behavior in class				
Student compliance	-0.43	0.115	0.09	0.545
Counts of student problem behaviors in the classroom	0.25	0.387	-0.03	0.837
Student behavior rating				
Disruptive behavior			-0.18	0.200
Emotional dysregulation			-0.28	0.140
Internalizing behavior			-0.23	0.246
Attention to schoolwork			0.39	0.061
Pro-social behavior (positive behaviors with peers)			0.19	0.196

SOURCES: District Records for the 2016-2017 school year; Teacher Ratings of Student Behavior collected in spring 2017; Classroom Observations data collected with Assessing School Settings: Interactions of Teachers and Students (ASSIST) in spring 2017; Staff Survey data collected in spring 2017; School-wide Evaluation Tool (SET) and Individual Student Systems Evaluation Tool (I-SSET) fielded in spring 2017.

NOTES: The analysis sample is all eligible students in the study schools who were initially identified as struggling with behavior and have non-missing values for the outcomes and the explanatory variables used in a given model. These students are defined as those students with the highest 15 percent disruptive behavior rating in the fall of 2015.

The analysis uses a three-step procedure to estimate the relationship between a given outcome and a given explanatory variable. The first step uses a hierarchical model to estimate the block-level program impacts on the outcome. This model controls for the blocking of random assignment and for baseline covariates appropriate to the outcome. The second step uses a hierarchical model to estimate the block-level program impacts on the explanatory variable. This model controls for the blocking of random assignment and for baseline covariates appropriate to the variable. The last step regresses the block-level impacts for the outcome on those for the explanatory variable using a precision weighted random-effect bivariate model. The coefficient for the explanatory variable impact reflects the amount of outcome change that is associated with one unit of change in the explanatory variable.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

II. Additional Information for Systematic Review

This section provides supplemental information about the estimation of the program's effects on student, school, and classroom outcomes that a systematic review might need to assess the quality of the study. It first presents the summary statistics and the estimated effects for the impact findings presented in the report, and then provides the realized minimum detectable effects based on the analysis samples used in the study.

Exhibits C.18 and C.19 present the means and standard deviations for all the student outcome measures used in the report for two samples, the overall student sample and the sample of students initially identified as struggling with behavior, respectively. These tables also report the estimated effects, corresponding standard errors and p-values, and sample sizes for each outcome. Exhibit C.18 also reports the estimated intra-class correlations (ICCs) for the student outcomes. Exhibit C.20 shows the same summary statistics for all school and classroom outcomes presented in the report.

A common way to convey a study's statistical power is through the minimum detectable effect (MDE) or the minimum detectable effect size (MDES). Formally, the MDE is the smallest true program impact that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 5 percent for a two-tailed test). The MDES is the MDE scaled as an effect size–in other words, it is the MDE divided by the standard deviation of the outcome of interest. Exhibit C.21 reports the realized values of the minimum detectable effects and the corresponding minimum detectable effect sizes based on the actual data and analytical approaches used in this study.

III. Supplemental Findings on Program Implementation

This section of the appendix presents detailed information that provides context for the program effect findings. It reports the extent of implementation of core components of the MTSS-B model and summarizes feedback from school staff about the quality of the training and technical assistance as well as factors that facilitated or hindered the implementation of MTSS-B practices in schools and classrooms.

Detailed Information of Core Components of MTSS-B Implementation

To describe the extent of implementation of the core components of the MTSS-B model in schools, the study team used detailed subscale data from the site visit protocols (described in Appendix B) as well as information from classroom observations and staff surveys.

The core components of school-wide practices to be implemented by all staff for all students include the establishment of school-wide positive behavior expectations, teaching and reinforcement of behavioral expectations, the use of classroom management practices to support adherence to behavioral expectations in the classroom, and systems for consistently responding to behavioral infractions. Exhibit C.22 shows that most *school-wide support practices* were implemented in most of the participating schools by the end of Program Year 2, but some components were implemented more widely than others. Specifically, systems for responding to behavioral violations were only implemented with fidelity in 61 percent of the participating schools by the end of Program Year 2.

Exhibit C.23 provides information on the study team's observation of whether there was evidence in classrooms that some of the *classroom management practices* targeted by the program were being implemented. The study team could only expect to observe evidence of a subset of the eight practices emphasized by the program since a teacher would not be expected to use all of the practices during a 15-minute observation period. The extent of implementation varied among the observed practices. Most classrooms were observed to have posted three to five positively stated behavioral expectations; in a typical participating school, 82 percent of classrooms were found to be partially or fully implementing this practice. While most classrooms were observed to have a system for reinforcing behaviors, in a typical participating school, relatively few classrooms (40 percent) were observed to be partially or fully using that system during the observation period.

Exhibit C.18. Supplemental Information for Estimated Program Effects on Student Outcomes, by Year and Outcome, Overall Sample

| Outcome | Participating Schools | | Non-Participating Schools | | Estimated Effect | Standard Error of Estimated Effect | p-value of Estimated Effect | Intra-Class Correlations (ICCs) | | Sample Size | |
	Adjusted Mean	Unadjusted Standard Deviation	Adjusted Mean	Unadjusted Standard Deviation				School-level	Teacher-level	Number of Students	Number of Schools
Student achievement											
Program Year 1											
Reading	0.03	1.01	0.02	1.00	0.01	0.03	0.581	0.032	NA	24,326	89
Math	0.02	1.01	0.02	1.00	0.00	0.03	0.985	0.035	NA	24,326	89
Program Year 2											
Reading	0.04	1.00	0.04	1.00	0.00	0.03	0.891	0.034	NA	23,569	89
Math	0.04	0.98	0.04	1.00	0.00	0.04	0.981	0.047	NA	23,569	89
Follow-up Year											
Reading	0.00	1.03	0.00	0.99	0.01	0.03	0.835	0.040	NA	19,968	80
Math	0.02	1.02	0.00	0.99	0.02	0.04	0.569	0.049	NA	19,968	80
Student behavior											
Program Year 1											
Disruptive behavior	0.58	0.72	0.61	0.75	-0.03	0.02	0.100	0.013	0.215	25,391	89
Emotional dysregulation	1.15	0.89	1.17	0.91	-0.02	0.02	0.410	0.007	0.306	25,391	89
Internalizing behavior	0.68	0.72	0.68	0.72	0.00	0.02	0.966	0.000	0.387	25,391	89
Pro-social behavior	3.90	0.96	3.89	0.96	0.01	0.03	0.801	0.007	0.310	25,391	89
Attention to schoolwork	3.45	1.21	3.46	1.20	-0.01	0.02	0.599	0.003	0.168	25,391	89
Program Year 2											
Disruptive behavior	0.56	0.69	0.58	0.73	-0.02	0.02	0.343	0.006	0.221	24,842	89
Emotional dysregulation	1.14	0.88	1.16	0.87	-0.02	0.03	0.450	0.003	0.314	24,842	89
Internalizing behavior	0.68	0.71	0.71	0.71	-0.03	0.02	0.220	0.003	0.371	24,842	89
Pro-social behavior	3.93	0.94	3.88	0.93	0.04	0.03	0.183	0.007	0.322	24,842	89
Attention to schoolwork	3.45	1.20	3.43	1.18	0.01	0.03	0.627	0.002	0.180	24,842	89

SOURCES: District records from school years 2014-2015, 2015-2016, 2016-2017, and 2017-2018; Teacher Ratings of Student Behavior collected in fall 2015, spring 2016, and spring 2017.

NOTES: Means for the participating and non-participating schools were adjusted using the regression models described in Appendix B. Unadjusted standard deviations for the participating and non-participating schools were the standard deviations across all students in the respective group. Rounding may cause slight discrepancies in calculating means and differences. Intra-Class Correlations (ICCs) were estimated using a multi-level model that controls for random assignment blocks.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.19. Supplemental Information for Estimated Program Effects on Student Outcomes, by Year and Outcome, for Students Initially Identified as Struggling with Behavior

Outcome	Participating Schools		Non-Participating Schools		Estimated Effect	Standard Error of Estimated Effect	p-value of Estimated Effect	Sample Size	
	Adjusted Mean	Unadjusted Standard Deviation	Adjusted Mean	Unadjusted Standard Deviation				Number of Students	Number of Schools
Student achievement									
Program Year 1									
Reading	-0.16	0.99	-0.27	0.89	0.11 *	0.05	0.023	1,879	89
Math	-0.18	0.97	-0.23	0.94	0.05	0.05	0.296	1,879	89
Program Year 2									
Reading	-0.14	0.95	-0.25	0.97	0.11 *	0.05	0.050	1,595	89
Math	-0.17	0.97	-0.24	1.10	0.07	0.06	0.209	1,595	89
Follow-Up Year									
Reading	-0.23	1.00	-0.32	0.95	0.09	0.07	0.173	1,323	80
Math	-0.22	1.03	-0.21	1.01	-0.02	0.07	0.817	1,323	80
Student behavior									
Program Year 1									
Disruptive behavior	1.64	0.86	1.69	0.86	-0.06 *	0.02	0.010	2,979	89
Emotional dysregulation	2.20	0.96	2.22	0.98	-0.02	0.03	0.477	2,979	89
Internalizing behavior	1.09	0.78	1.10	0.77	0.00	0.03	0.944	2,979	89
Pro-social behavior	2.92	0.95	2.87	0.90	0.04	0.03	0.212	2,979	89
Attention to schoolwork	2.43	1.09	2.41	1.06	0.02	0.03	0.627	2,979	89

(continued)

Exhibit C.19 (continued)

Outcome	Participating Schools		Non-Participating Schools		Estimated Effect	Standard Error of Estimated Effect	p-value of Estimated Effect	Sample Size	
	Adjusted Mean	Unadjusted Standard Deviation	Adjusted Mean	Unadjusted Standard Deviation				Number of Students	Number of Schools
Program Year 2									
Disruptive behavior	1.17	0.92	1.25	0.98	-0.08 *	0.03	0.015	1,847	89
Emotional dysregulation	1.74	1.02	1.81	1.05	-0.06	0.04	0.143	1,847	89
Internalizing behavior	0.86	0.76	0.93	0.74	-0.07	0.04	0.090	1,847	89
Pro-social behavior	3.30	1.06	3.23	1.00	0.07	0.05	0.156	1,847	89
Attention to schoolwork	2.70	1.16	2.68	1.11	0.02	0.05	0.662	1,847	89

SOURCES: District Records from the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 school years; Teacher Ratings of Student Behavior collected in fall 2015, spring 2016, and spring 2017.

NOTES: For Program Years 1 and 2, the analysis includes students from 58 participating schools and 31 non-participating schools. For the Follow-Up Year, the analysis includes students from 52 participating schools and 28 non-participating schools. A student is initially identified as struggling with behavior if his/her baseline disruptive behavior rating is in the highest 15 percent among students in the same grade across all sample schools

   Means for the participating and non-participating schools were adjusted using regression models described in Appendix B. Unadjusted standard deviations for the participating and non-participating schools were the standard deviations across all students in the respective group. Rounding may cause slight discrepancies in calculating means and differences.

   A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.20. Supplemental Information for Estimated Program Effects on School and Classroom Outcomes, by Outcome, Program Year 2

| Outcome | Participating Schools | | Non-Participating Schools | | Estimated Effect | Standard Error of Estimated Effect | p-value of Estimated Effect | Unit of Observation | Sample Size | |
	Adjusted Mean	Unadjusted Standard Deviation	Adjusted Mean	Unadjusted Standard Deviation					Number of Observations	Number of Schools
School climate										
Academic focus	2.78	0.52	2.71	0.53	0.07 *	0.04	0.044	Staff member	3,986	89
Principal leadership	3.17	0.76	3.10	0.76	0.07	0.08	0.386	Staff member	4,097	89
Staff collegiality	3.58	0.47	3.51	0.48	0.07 *	0.03	0.016	Staff member	4,192	89
Teacher-student relationships	3.30	0.61	3.17	0.64	0.14 *	0.06	0.025	Staff member	4,210	89
Consistent school discipline	3.12	0.71	3.02	0.72	0.10	0.06	0.114	Staff member	4,121	89
School safety	3.35	0.56	3.33	0.56	0.02	0.03	0.485	Staff member	4,165	89
Classroom management practices										
Facilitate orderly classroom transitions	4.17	1.05	4.03	1.15	0.14 *	0.06	0.029	Classroom	1,791	88
Teacher anticipation and responsiveness to student needs	4.22	0.76	4.05	0.79	0.17 *	0.06	0.004	Classroom	1,791	88
Teacher proactive behavior management	3.85	0.86	3.52	0.91	0.33 *	0.07	0.000	Classroom	1,791	88
Teacher active monitoring of students	4.56	0.68	4.45	0.76	0.11 *	0.04	0.017	Classroom	1,791	88
Classroom functioning										
Student disruptive behaviors	1.19	0.24	1.24	0.28	-0.04 *	0.02	0.020	Classroom	1,791	88
Student compliance	4.43	0.66	4.30	0.72	0.14 *	0.04	0.003	Classroom	1,791	88
Student engagement with classroom activities	3.63	0.89	3.48	0.96	0.16 *	0.06	0.015	Classroom	1,791	88
Teacher control of classroom	4.67	0.49	4.56	0.57	0.11 *	0.04	0.009	Classroom	1,791	88
Counts of student disruptive behaviors in the classroom	6.96	7.18	8.88	9.00	-1.92 *	0.54	0.001	Classroom	1,791	88

(continued)

Exhibit C.20 (continued)

SOURCES: Classroom Observations data collected with Assessing School Settings: Interactions of Teachers and Students (ASSIST) in spring 2017; Staff Survey data collected in spring 2017.

NOTES: For school climate analysis, the sample includes staff members from 58 participating schools and 31 non-participating schools. For classroom outcome analysis, the sample includes classrooms from 57 participating schools and 31 non-participating schools.
Means for the participating and non-participating schools were adjusted using regression models described in Appendix B. Unadjusted standard deviations for the participating and non-participating schools were the standard deviations across all students in the respective group. Rounding may cause slight discrepancies in calculating means and differences.
   A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit C.21. Realized Minimum Detectable Effects and Minimum Detectable Effect Sizes for Student, School, and Classroom Outcomes, by Year and Outcome, for Overall Sample and Students Initially Identified as Struggling with Behavior

Outcome	Unit of Measure	Overall Sample		Students Initially Identified as Struggling with Behavior	
		Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size	Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size
Student achievement					
Program Year 1					
Reading	Standardized score	0.07	0.07	0.14	0.14
Math	Standardized score	0.08	0.09	0.14	0.14
Program Year 2					
Reading	Standardized score	0.08	0.08	0.15	0.15
Math	Standardized score	0.11	0.11	0.17	0.17
Follow-Up Year					
Reading	Standardized score	0.10	0.10	0.18	0.19
Math	Standardized score	0.11	0.11	0.19	0.19
Student behavior					
Program Year 1					
Disruptive behavior	0- to 5-point scale	0.05	0.06	0.06	0.08
Emotional dysregulation	0- to 5-point scale	0.06	0.07	0.08	0.09
Internalizing behavior	0- to 5-point scale	0.05	0.07	0.07	0.10
Pro-social behavior	0- to 5-point scale	0.07	0.07	0.09	0.10
Attention to schoolwork	0- to 5-point scale	0.06	0.05	0.09	0.08

(continued)

Exhibit C.21 (continued)

Outcome	Unit of Measure	Overall Sample		Students Initially Identified as Struggling with Behavior	
		Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size	Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size
Program Year 2					
Disruptive behavior	0- to 5-point scale	0.05	0.07	0.09	0.13
Emotional dysregulation	0- to 5-point scale	0.08	0.09	0.12	0.14
Internalizing behavior	0- to 5-point scale	0.07	0.09	0.11	0.16
Pro-social behavior	0- to 5-point scale	0.09	0.09	0.14	0.15
Attention to schoolwork	0- to 5-point scale	0.08	0.07	0.15	0.13
School climate					
Program Year 2					
Academic focus	1- to 4-point scale	0.10	0.18	NA	NA
Principal leadership	1- to 4-point scale	0.21	0.28	NA	NA
Staff collegiality	1- to 4-point scale	0.17	0.26	NA	NA
Teacher-student relationships	1- to 4-point scale	0.08	0.17	NA	NA
Consistent school discipline	1- to 4-point scale	0.17	0.24	NA	NA
School safety	1- to 4-point scale	0.10	0.17	NA	NA
Classroom management practices					
Program Year 2					
Facilitate orderly classroom transitions	1- to 5-point scale	0.18	0.16	NA	NA
Teacher anticipation and responsiveness to student needs	1- to 5-point scale	0.16	0.21	NA	NA
Teacher proactive behavior management	1- to 5-point scale	0.19	0.20	NA	NA
Teacher active monitoring of students	1- to 5-point scale	0.12	0.16	NA	NA

(continued)

Exhibit C.21 (continued)

Outcome	Unit of Measure	Overall Sample		Students Initially Identified as Struggling with Behavior	
		Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size	Realized Minimal Detectable Effects	Realized Minimal Detectable Effect Size
Classroom functioning					
Program Year 2					
Student disruptive behaviors	1- to 5-point scale	1.50	0.17	NA	NA
Student compliance	1- to 5-point scale	0.12	0.17	NA	NA
Student engagement with classroom activities	1- to 5-point scale	0.17	0.18	NA	NA
Teacher control of classroom	1- to 5-point scale	0.11	0.20	NA	NA
Counts of student disruptive behaviors in the classroom	Number of occurrences	0.05	0.18	NA	NA

SOURCES: District Records from the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 school years; Teacher Ratings of Student Behavior collected in fall 2015, spring 2016, and spring 2017; Classroom Observations data collected with Assessing School Settings: Interactions of Teachers and Students (ASSIST) in spring 2017; Staff Survey data collected in spring 2017; School-wide Evaluation Tool (SET) and Individual Student Systems Evaluation Tool (I-SSET) fielded in spring 2017.

NOTES: For each outcome, the minimum detectable effect (MDE) is calculated as the standard error of the associated impact estimate multiplied by 2.8. The minimum detectable effect size (MDES) is calculated as the associated MDE divided by the standard deviation of the outcome across all observations in the non-participating schools.
 NA: Not Applicable.

Exhibit C.22. Implementation of Core Components of School-Wide Practices in Participating Schools

Measures	Program Year 1		Program Year 2	
	Mean	Standard Deviation	Mean	Standard Deviation
Behavior expectations defined (% scoring at or above 80% on SET subscale)[a]	60.3	49.3	71.9	45.3
Behavior expectations taught (% scoring at or above 80% on SET subscale)[a]	77.6	42.1	89.5	31.0
System for rewarding behavioral expectations (% scoring at or above 80% on SET subscale)[a]	70.7	45.9	87.7	33.1
System for responding to behavioral violations (% scoring at or above 80% on SET subscale)[a]	58.6	49.7	61.4	49.1
Mean percentage of 8 MTSS-B classroom practices for which teachers reported placing moderate or major emphasis in their classrooms[b]	NA	NA	90.8	6.2
Mean percentage of 8 MTSS-B classroom practices for which teachers reported placing major emphasis in their classrooms[b]	NA	NA	64.2	11.8

SOURCES: Author's calculations based on fall 2015 and spring 2017 school site visits and MTSS-B spring 2017 staff survey.

NOTES: The sample includes 58 participating schools for Program Year 1 and 57 participating schools for Program Year 2. Except where noted, the "mean" column reports the unadjusted mean values across all participating schools in the sample, and the "standard deviation" column reports the unadjusted standard deviation across the same schools.

[a]Fall 2015 and spring 2017 site visits using the School-wide Evaluation Tool (SET) protocol; Year 1 n = 58 schools, Year 2 n = 57 schools.

[b]MTSS-B spring 2017 staff survey, n = 1,498 teachers.

NA: Not Applicable.

Exhibit C.23. Percentage of Classrooms Observed to be Partially or Fully Implementing MTSS-B Classroom Practices, for Participating Schools

Measures (%)	Partially or Fully		Fully	
	Mean	Standard Deviation	Mean	Standard Deviation
3-5 positively stated behavioral expectations are posted in the classroom	82.1	17.1	63.1	25.6
A behavioral matrix defining the expectations is clearly visible in the classroom	64.9	29.9	54.8	32.3
Teacher verbally references or points to the behavioral expectation signage during observation	43.9	26.8	32.9	24.7
Observed evidence that the teacher has a reinforcement system to reward positive behaviors	62.7	22.0	52.4	22.7
Classroom has a system in place where students are "gaining" points or status based on their behavior	55.1	18.7	45.9	19.0
Teacher uses the reinforcement system during observation	39.9	18.8	35.0	17.6

SOURCE: MTSS-B spring 2017 classroom observations (n = 1,791 classrooms, 57 participating schools).

NOTE: The "mean" column reports the unadjusted mean values across all participating schools in the sample, and the "standard deviation" column reports the unadjusted standard deviation across the same schools.

The core components of the supplemental supports intervention used in this study, CICO, include providing additional support for students to develop behavioral skills and increased opportunity for feedback aligned with the school-wide expectations. Exhibit C.24 shows that among those students who were enrolled in CICO, the intervention was implemented as intended.

Exhibit C.24. Implementation of Core Tier II Practices for Participating Schools, for Program Year 2

Measures	Mean	Standard Deviation
Schools implementing a supplemental support intervention with fidelity (percentage of schools)[a]	86.0	NA
Average percentage of teachers who report providing feedback to enrolled students at least daily[b]	75.4	16.7

SOURCES: Author's calculations based on:

[a]MTSS-B fall 2015 and spring 2017 site visits using the Targeted Intervention subscale on the modified Individual Student Systems Evaluation Tool (I-SSET) protocol (Debnam, Pas, and Bradshaw, 2012), n = 57 schools.

[b]Spring 2017 staff survey, n = 836 teachers in 57 schools; this question was only asked to teachers who had already reported using a targeted intervention.

NOTES: Except where noted, the "mean" column reports the unadjusted mean values across all participating schools in the sample, and the "standard deviation" column reports the unadjusted standard deviation across the same schools.

NA: Not applicable because this is a dichotomous variable.

In addition to the school-wide and supplemental support components, the program also called for the installment of specific infrastructure to support the implementation of MTSS-B. This component included a team to organize, lead, and support implementation; an MTSS-B coach to support the teams' staff in implementation; and a behavior data monitoring system that allowed the teams to monitor and adapt implementation. Exhibit C.25 shows that even though most schools were able to hold team meetings and collect fidelity data at the intended frequency for both Tier I and Tier II, some schools struggled to operate teams as intended and many were challenged by a lack of stability in team members. Exhibit C.26 shows that on average, schools received somewhat less coaching support than intended. The study intended for coaches to spend one day per week in each participating school, and on average, schools received coaching support for 73 percent of instructional weeks in Program Year 1 and 79 percent of instructional weeks in Program Year 2. By Program Year 2, coaches were spending the majority of their time in schools "coaching" (observing and providing feedback to teachers) as opposed to other activities and about 60 percent of teachers in the typical participating school reported receipt of coaching supports in Program Year 2. Exhibit C.27 shows that for the school-wide practices, most schools had practices in place to use their data systems to track student behavior. While most schools had a system to monitor students enrolled in the supplemental supports intervention (SWIS-CICO), schools enrolled relatively few students in the intervention.

Staff Perceptions of MTSS-B Training and Technical Assistance and Implementation of the MTSS-B Practices

The study collected staff feedback on the useful and challenging aspects of the training and technical assistance provided to schools and teachers through phone interviews with school staff in the spring of each program year. Exhibits C.28-C.30 show staff perceptions of the in-person training (summer content training), virtual training and technical assistance (webinars), and in-person technical assistance (site visits).

The study also collected staff views on useful strategies for and challenges to implementing the school-wide support practices (Exhibit C.31); as well as challenges associated with the implementation of the supplemental support practices (Exhibit C.32); and team functioning (Exhibit C.33).

Exhibit C.25. Implementation of Core Components of Team Functioning for Participating Schools

Teams	Program Year 1 Percentage of Schools	Program Year 2 Percentage of Schools
Team for school-wide practices (Tier I)		
Schools implementing Tier I management practices with fidelity[a]	70.7	70.2
Schools with team for school-wide practices meeting at least 6 times per year[b]	93.1	87.7
Schools collected Tier I fidelity data at least 3 times per year[c]	87.9	93.0
Schools with stable team[b,d]	24.1	40.4
Team for supplemental supports (Tier II)		
Schools implementing Tier II foundations practices with fidelity[a]	NA	47.4
Schools with team for supplemental supports meetings at least 6 times per year[b]	NA	91.2
Schools collected Tier II fidelity data at least 3 times per year[c]	NA	93.0
Schools with stable team[b,d]	NA	19.3

SOURCES: Author's calculations based on:

[a]Fall 2015 and spring 2017 site visits using the Schoolwide Evaluation Tool (SET) and modified Individual Student Systems Evaluation Tool (I-SSET) protocols.

[b]Team meeting minutes logs that teams completed as part of the program at each meeting.

[c]PBIS Apps Tiered Fidelity Inventory entered by teams when completed.

[d]Teams defined as "stable" when three key team member roles were fulfilled by the same individual for at least 80 percent of the team meetings.

NOTE: The sample includes 58 participating schools for Program Year 1 and 57 participating schools for Program Year 2. The "standard deviation" column reports the unadjusted standard deviation across all participating schools in the sample.

NA: Not Applicable.

Exhibit C.26. Implementation of Core Components of MTSS-B Coach Role

Measures	Program Year 1		Program Year 2	
	Mean	Standard Deviation	Mean	Standard Deviation
Mean percent of instructional weeks schools received coaching[a]	79.1	14.8	72.9	18.4
Mean percent of coach time in participating schools spent observing or providing feedback to teachers[a]	48.9	20.8	63.6	22.0
Mean percent of a school's teachers who report having received coaching support[b]	NA	NA	59.1	16.6

SOURCES: Author's calculations based on:
[a]MTSS-B coaches' logs from 2015-2016 and 2016-2017 school years.
[b]Spring 2017 staff survey (n = 1,500 teachers).

NOTES: The "mean" column reports the unadjusted mean values across all participating schools in the sample, and the "standard deviation" column reports the unadjusted standard deviation across the same schools. Year 1 n = 58 schools, Year 2 n = 57 schools.
NA: Not Applicable.

Exhibit C.27. Implementation of Core Components of Data System

Measures	Program Year 1		Program Year 2	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
School-wide practices (Tier I)				
Schools implementing Tier I data use practices with fidelity (percentage of schools)[a]	82.8	NA	93.0	NA
Schools using Tier I SWIS System (percentage of schools)[b]	100.0	NA	98.2	NA
Average number of SWIS reports generated per school[b]	425	323	433	255
Supplemental Supports (Tier II)				
Schools using Tier II SWIS-CICO (percentage of schools)[c]	NA	NA	98.2	NA
Mean percent of a school's population (K-5) enrolled in CICO[c]	NA	NA	5.3	3.1

SOURCES: Author's calculations based on:

[a]Fall 2015 and spring 2017 site visits using the Schoolwide Evaluation Tool (SET).

[b]PBIS Apps School-wide Information System (SWIS) usage data from 2015-2016 and 2016-2017 school years.

[c]SWIS Check In Check Out (CICO) data from 2016-2017 school year.

NOTES: Except where noted, the "mean" column reports the unadjusted mean values across all participating schools in the sample, and the "standard deviation" column reports the unadjusted standard deviation across the same schools.

Tier I analyses are based on information from 58 participating schools in Program Year 1 and 57 participating schools in Program Year 2; Tier II analyses are based on Program Year 2 information from 57 participating schools among which 56 schools used the SWIS-CICO system.

NA: Not Applicable.

Exhibit C.28. Perceptions of In-Person Trainings

Helpful aspects[a]

■ Top three categories

	Year 1 (n = 99)	Year 2 (n = 86)
Content from trainers	63.6	68.6
Action planning with team	25.3	26.7
Opportunity to build team commitment	11.1	8.1
Other	10.1	7.0
Tangible resources to be used for implementation	8.1	NA[b]
Collaboration with other schools	8.1	14.0
Not helpful	4.0	7.0

Percentage of respondents

Recommendations for improvement[c]

■ Top three categories

	Year 1 (n = 99)	Year 2 (n = 86)
More convenient scheduling	30.3	8.1
Prepare more for training participants before event	23.2	NA[b]
Less density of information	23.2	8.1
More differentiation to meet individual school needs	21.2	17.4
Slower pace	17.2	15.1
Shorter duration	16.2	4.7
No improvement needed	10.1	36.0
Different or expanded target audience	7.1	NA[b]
Other	4.0	10.5

Percentage of respondents

SOURCE: Authors' calculations based on spring 2016 phone interviews with 99 respondents and spring 2017 phone interviews with 86 respondents.

NOTES: Data mostly come from the following questions specified below asked of Tier I team leaders and principals in Year 1 and of Tier I and Tier II team leaders in Year 2. However, if a participant's response to other interview questions applied to this table, those responses were also coded. Items may sum to more than 100 percent because respondents may state more than one answer choice.

[a]Do you find the readiness and summer training helpful in preparing you to implement MTSS-B? If so, what was helpful about the on-site readiness and summer trainings by CSBS?

[b]Not applicable: This theme was not mentioned by enough respondents to be coded.

[c]What could be improved about the on-site readiness and summer training from CSBS?

Exhibit C.29. Perceptions of Virtual Training & Technical Assistance (Webinars)

Helpful aspects of webinars[a]

■ Top three categories

	Year 1 (n = 87)	Year 2 (n = 76)
Content from trainers	50.6	42.1
Tangible resources to be used for implementation	37.9	14.5
Collaboration with other schools	21.8	57.9
Not helpful	13.8	5.3
Reminders to implement practices	13.8	17.1
Other	8.0	9.2

Percentage of respondents
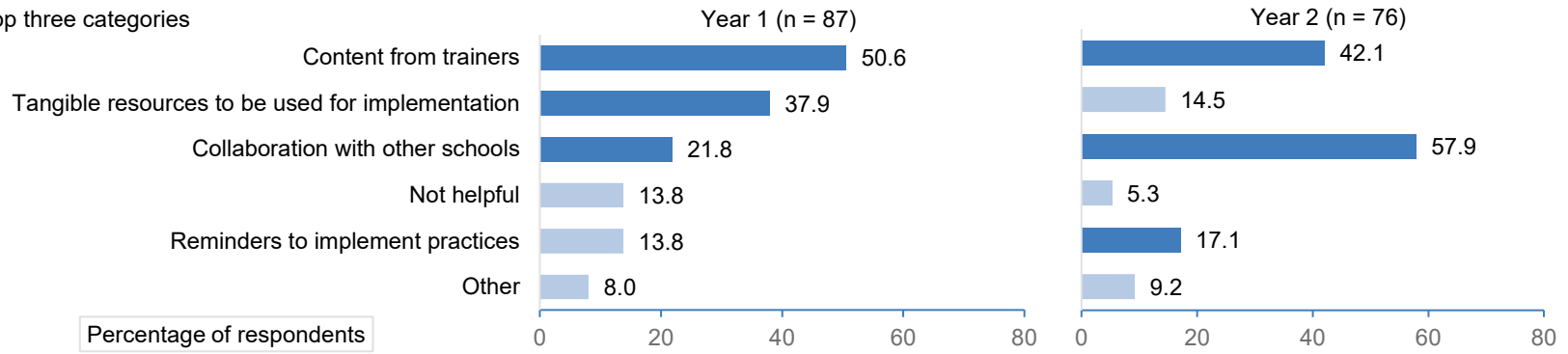
Recommendations for improvement[b]

■ Top three categories

	Year 1 (n = 87)	Year 2 (n = 76)
Alternate medium for training delivery	27.6	22.4
No need for improvement	23.0	32.9
More convenient scheduling	17.2	14.5
More differentiation to meet individual school needs	14.9	25.0
Different or expanded target audience	11.5	NA[c]
Shorter duration	11.5	10.5
More tangible resources for implementation	8.0	5.3
Other	5.7	10.5

Percentage of respondents
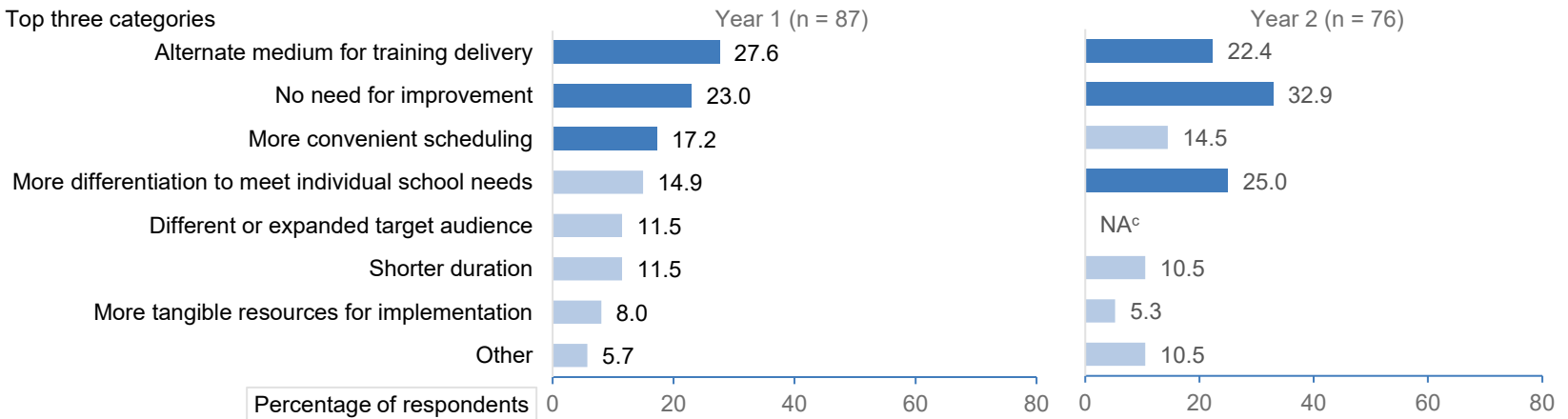
SOURCE: Authors' calculations based on spring 2016 phone interviews with 87 respondents and spring 2017 phone interviews with 76 respondents.

NOTES: Data mostly come from the following questions specified below asked of Tier I team leaders and principals in Year 1 and of Tier I and Tier II team leaders in Year 2. However, if a participant's response to other interview questions applied to this table, those responses were also coded.

Items may sum to more than 100 percent because respondents may state more than one answer choice.

[a] Do you find these webinars helpful? If so, what was helpful about the webinar?

[b] What could be improved about the webinars?

[c] Not applicable: This theme was not mentioned by enough respondents to be coded.

Exhibit C.30. Perceptions of In-Person Technical Assistance (School Site Visits)

Helpful aspects[a]

■ Top three categories

	Year 1 (n = 102)	Year 2 (n = 98)
Feedback on fidelity	72.5	69.4
Content from trainers	22.5	10.2
Positive relationships with trainers	16.7	17.3
Other	4.9	9.2
Not helpful	2.9	5.1
Motivation to staff to implement practices	NA[b]	9.2

Percentage of respondents

Recommendations for improvement[c]

■ Top three categories

	Year 1 (n = 102)	Year 2 (n = 98)
No improvement needed	32.4	50.0
More convenient scheduling	13.7	10.2
More actionable or useful feedback on fidelity	12.7	18.4
More direct training of all staff during visit	10.8	7.1
Other	9.8	6.1
More frequent or longer visits	7.8	8.2
Improved relationship with trainers	5.9	5.1

Percentage of respondents

(continued)

Exhibit C.30 (continued)

SOURCE: Authors' calculations based on spring 2016 phone interviews with 102 respondents and spring 2017 phone interviews with 98 respondents.

NOTES: Data mostly come from the following questions asked of Tier I team leaders and principals in Year 1 and of Tier I and Tier II team leaders in Year 2 specified below. However, if a participant's response to other interview questions applied to this table, those responses were also coded.

   Items may sum to more than 100 percent because respondents may state more than one answer choice.

   [a] Do you find these school-year visits from CSBS helpful? If so, what was helpful about the visits?

   [b] Not applicable: This theme was not mentioned by enough respondents to be coded.

   [c] What could be improved about the school-year visits from CSBS?

Exhibit C.31. Principals' and MTSS-B Team Leaders' Perceptions of Strategies and Challenges for Implementation of School-Wide Practices (Tier I)

Strategies[a]

■ Top three categories

Year 1 (n = 92)

Strategy	Percentage
Positive feedback to staff from team	41.3
Peer influence	27.2
Sharing student data with staff	26.1
Reminders to staff to implement	19.6
Perception by staff that MTSS-B has a positive effect	12.0
Other	9.8
Teacher feedback on implementation encouraged	8.7
School leaders holding staff accountable	8.7
Kick-off meeting to introduce program to staff	8.7
Coach support	6.5

Year 2 (n = 45)

Strategy	Percentage
Positive feedback to staff from team	37.8
Peer influence	20.0
Sharing student data with staff	33.3
Reminders to staff to implement	11.1
Perception by staff that MTSS-B has a positive effect	37.8
Other	4.4
Teacher feedback on implementation encouraged	6.7
School leaders holding staff accountable	13.3
Kick-off meeting to introduce program to staff	20.0
Coach support	15.6

Percentage of respondents

Challenges[b]

■ Top three categories

Year 1 (n = 86)

Challenge	Percentage
MTSS-B philosophy and staff beliefs don't fully align	45.3
Incomplete understanding of program	29.1
Overwhelmed by time and demands of MTSS-B	24.4
Overwhelmed by other initiatives and priorities	22.1
No challenges identified	14.0
Limited school leadership support	8.1
Other	10.5
Perception by staff that MTSS-B is not effective	5.8

Year 2 (n = 50)

Challenge	Percentage
MTSS-B philosophy and staff beliefs don't fully align	40.0
Incomplete understanding of program	22.0
Overwhelmed by time and demands of MTSS-B	8.0
Overwhelmed by other initiatives and priorities	16.0
No challenges identified	18.0
Limited school leadership support	8.0
Other	6.0
Perception by staff that MTSS-B is not effective	20.0

Percentage of respondents

SOURCE: Authors' calculations based on spring 2016 and spring 2017 phone interviews.

NOTE: Data mostly come from the following questions specified below asked of Tier I team leader and principals in Year 1 (n = 86-92 respondents) and Tier I team leaders in Year 2 (n = 45-50 respondents). However, if a participant's response to other interview questions applied to this table, those responses were also coded. Items may sum to more than 100 percent because respondents may state more than one theme.
  [a]What strategies have you and your team used to develop staff commitment and buy-in?
  [b]Has your school encountered any challenges in developing commitment and buy-in? If yes, please describe the challenges you have faced.

Exhibit C.32. Principals' and MTSS-B Team Leaders' Perceptions of Challenges in Implementation of Supplemental Supports (Tier II)

Challenges

■ Top three categories   Year 2 (n = 53)

Challenge	Percentage
Inconsistent participation by students	66.7
Limited staff commitment or participation	63.0
Incomplete staff understanding of program	44.4
Insufficient staff time to be CICO Facilitators	35.2
Limited staff capacity for data entry	27.8
Misidentification of students for intervention	16.7
Other	7.4

Percentage of respondents
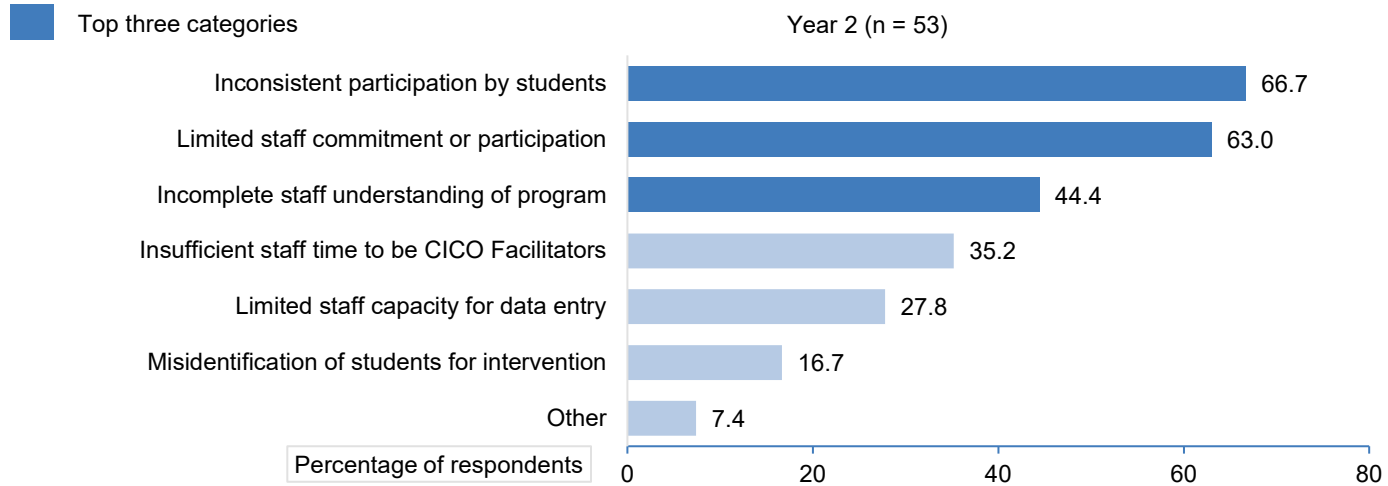
SOURCE: Authors' calculations based on spring 2017 phone interviews.

NOTE: Data mostly comes from the following questions specified below asked of Tier II team leaders in Year 2 (n = 53). However, if a participant's response to other interview questions applied to this table, those responses were also coded. Items may sum to more than 100 percent because respondents may state more than one theme.

1. Has your school encountered any challenges in the implementation of CICO with students who are enrolled in the program?

Exhibit C.33. Team Leader Perceptions of Challenges with Team Functioning

Team for school-wide praties (Tier I)

Top three categories (blue legend)

Year 1 (n = 42)

Category	Percentage
Establishing or maintaining team roles	35.7
Completing all agenda items in alloted time	35.7
Scheduling regular meetings	28.6
Preparing or analyzing data for meeting	23.8
Following the meeting protocol	21.4
No challenges identified	19.0
Creating or maintaining productive team culture	19.0
Dedicating time to team leader responsibilities	11.9
Navigating the web portal for note-taking	9.5
Other	7.1

Year 2 (n = 52)

Category	Percentage
Establishing or maintaining team roles	36.5
Completing all agenda items in alloted time	17.3
Scheduling regular meetings	17.3
Preparing or analyzing data for meeting	7.7
Following the meeting protocol	25.0
No challenges identified	51.9
Creating or maintaining productive team culture	13.5
Dedicating time to team leader responsibilities	13.5
Navigating the web portal for note-taking	21.2
Other	9.6

Percentage of respondents

Team for supplemental supports (Tier II)

Top three categories (blue legend)

Year 2 (n = 50)

Category	Percentage
Establishing or maintaining team roles	42.0
Completing all agenda items in alloted time	6.0
Scheduling regular meetings	24.0
Preparing or analyzing data for meeting	16.0
Following the meeting protocol	2.0
No challenges identified	68.0
Creating or maintaining productive team culture	8.0
Dedicating time to team leader responsibilities	24.0
Navigating the web portal for note-taking	2.0
Other	8.0

Percentage of respondents

SOURCE: Spring 2016 and spring 2017 phone interviews.

NOTE: Data mostly comes from the questions specified below answered by 42 Tier I team leaders in Year 1, 52 Tier I team leaders in Year 2, and 50 Tier II team leaders in Year 2. However, if a participant's response to other interview questions applied to this table, those responses were also coded. Items may sum to more than 100 percent because respondents may state more than one theme.

1. Have you faced any challenges so far in running SLT or MTSS-B (or Tier II) team meetings? If yes, please describe the challenges you have faced.

IV. Findings From This Study vs. Other Rigorous Studies of MTSS-B Training

This section compares the present study to other evaluations of school-wide positive behavioral interventions and supports. A 2020 meta-analysis by Lee and Gage identified 32 studies, of 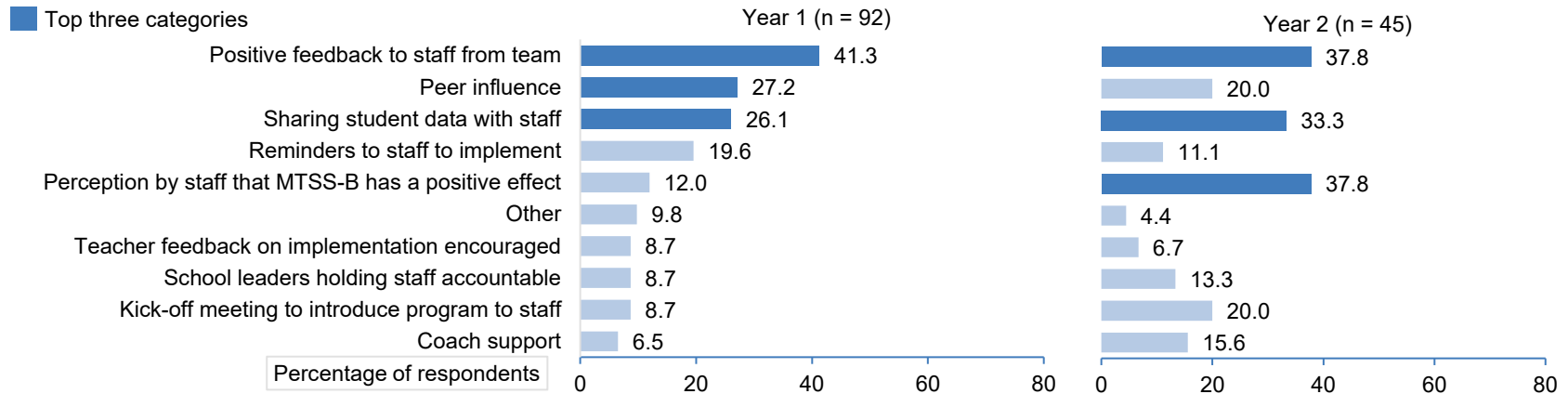which 6 used a random assignment design and 26 were quasi-experimental. Focusing on those most comparable to the present study—the six random assignment studies—Exhibit C.34 lists the type of program, setting and sample, and key impacts reported. Since the analysis differed across the studies, it is not feasible to report comparable effect sizes, but it is notable that there is a mixed pattern of findings across the domains of school-level outcomes, teacher practices, and student outcomes. Lee and Gage conducted a quantitative meta-analysis across all 32 random assignment and quasi-experimental studies and found statistically and educationally significant effects across all three domains. They characterized the effects on organizational and behavioral outcomes as large (greater than 0.2) and those on academic outcomes as medium (0.05 to 0.2). Narrowing the focus to the 14 studies that met What Works Clearinghouse (WWC) standards and examined behavioral outcomes, Lee and Gage reported that most of the studies that met WWC standards and examined behavioral outcomes found positive results, while the results for organizational and student academic outcomes were a mixture of significant and null findings.[37] Narrowing the focus still further to the research most comparable to the present study—that reported by Bradshaw and co-authors and by Waasdorp, Bradshaw, and Leaf, which used a similar random assignment design and measures—the exhibit shows effects sizes on disruptive behavior for the full sample in the Bradshaw and co-authors study to be similar to those for the subgroup of initially struggling students in the current study.[38] However, Bradshaw and co-authors differed from the current study by finding impacts on a number of other student behaviors. In addition, the current study reported findings on student achievement outcomes while Bradshaw and co-authors did not.[39]

Exhibit C.34. Comparison of Findings of this Study with Prior Random Assignment Research on School-Wide Positive Behavioral Interventions and Supports

Study	Program Studied	Context			Key Impacts Reported		
		Sites	Student Sample	School Level Outcomes	Teacher Practices	Student Behavioral and Academic Outcomes	
1. Algozzine et al. (2012)	Tiers I, II, and III, implemented for 1 year with School-wide Evaluation Tool (SET) fidelity measure of 94%	6 elementary schools in a large school district in North Carolina. 4 schools assigned to School-Wide Positive Behavioral Interventions and Supports (SWPBIS) and 2 as comparison schools	768 students in K-3, of which 59% Black, 28% Hispanic, 64% Free and Reduced-Price Lunch (FRL), and 11% Special Education (SPED)	Not measured	Not measured	Reduced office disciplinary referrals, improvements in early reading skills, increased rate of meeting 3rd grade reading standards	
2. Bradshaw et al. (2009, 2010)	Tier I implemented for 4 years with SET fidelity of ~95% in Year 4	37 elementary schools in 5 districts (48% suburban, 41% urban, and 11% rural) in Maryland. 21 schools assigned to SWPBIS and 16 as comparison schools	~18,000 students, of which 60% White, 40% FRL, and 14% SPED. Percentage of Hispanic students not reported	Improvements in overall organizational health and on resource influence, staff affiliation, and academic emphasis. No impacts on institutional integrity and collegial leadership	Higher SET and Effective Behavior Supports (EBS) scores related to implementation of SWPBIS practices	Reductions in student suspensions. No impacts on state 3rd and 5th grade reading and math tests	

(continued)

Exhibit C.34 (continued)

Study	Program Studied	Context		Key Impacts Reported		
		Sites	Student Sample	School Level Outcomes	Teacher Practices	Student Behavioral and Academic Outcomes
3. Bradshaw et al. (2012); Wassdorp et al. (2012)	Tier I implemented for 4 years with all schools reaching SET fidelity threshold of 80% by Year 4	37 elementary schools in 5 districts (48% suburban, 41% urban, and 11% rural) in Maryland. 21 schools assigned to SWPBIS and 16 as comparison schools	12,344 students, of which 46% Black, 45% White, 4% Hispanic, 49% FRL, and 13% SPED	Not reported	Not reported	Reduced rates of student office disciplinary referrals, bullying, peer rejection, disruptive behavior (-0.12), and higher rates of concentration (0.08), socio-emotional functioning (-0.11), and pro-social behavior (0.17) in Year 4. No impact on suspensions
4. Horner et al. (2009)	Tier I implemented for 3 years with SET fidelity of 80% or above	63 elementary schools in Hawaii and Illinois. 33 assigned to SWPBIS and 30 to comparison schools	~30,000 students, of which 61% non-White, 51% FRL, and 9% with Individualized Education Programs (IEPs)	Improvements in perceived safety of the school setting	Not measured	Improvement in proportion of 3rd graders meeting or exceeding state reading standard
5. Current study	Tier I implemented for 3 years with SET average fidelity scores of 83%, 86% and 85%, and Tier II implemented for 2 years with ISSET average fidelity scores of 82% and 83%	89 elementary schools in 9 districts in 6 states. 58 schools assigned to MTSS-B training and 31 schools assigned to comparison schools	~20,000 to 25,000 depending on the measure being analyzed, of which 9% Black, 60% Hispanic, 19% White, 57% FRL, and 9% SPED	Improvements in teacher-student relationships, academic focus, and staff collegiality	Improvements in classroom management practices and classroom functioning	No impacts on behavior or achievement for students overall. Reductions in disruptive behavior (-0.07, -0.11) and improvements in reading achievement (0.11, 0.11) in Years 1 and 2 for students initially identified as struggling with behavior

APPENDIX D. SUPPLEMENTAL SENSITIVITY ANALYSES

This appendix presents findings from additional checks that assessed whether the study's primary findings were sensitive to decisions about which students were included in the analyses. It also provides the program effects findings on additional student subgroups.

I. Sensitivity of Model Specification

The team examined whether the impact findings presented in the report were sensitive to model specifications, particularly covariate selections. Exhibit D.1 presents findings from these checks for the estimated program effects on student behavior for the overall sample and the subgroup of students initially identified as struggling with behavior issues. Exhibit D.2 shows results from similar checks for the impact estimates on student academic achievement. Exhibits D.3 and D.4 provide findings from model specification checks for impact estimates on school climate and classroom management practices and functioning, respectively. In general, the magnitudes of the impact estimates were not sensitive to different model or sample specifications.

II. Sensitivity of Student Subgroup Definition

A key student subgroup in this study is students who were initially identified as struggling with behavior. As noted earlier, the team defines this group of students as all students whose baseline disruptive scale was at or above the 85th percentile value in a given grade across all schools. In other words, these students had the highest 15 percent baseline disruptive ratings. Doing so allowed for a more homogenous group of students (in terms of their disruptive behavior) within the subgroup. This definition used the disruptive behavior rating as the defining variable because the program focused on reducing such behavior. The team chose the 15 percent cut-off point because it aligned with the literature on tiered systems of support for behavior, which generally suggested that up to 15 percent of a schools' student population would benefit from targeted or intensive intervention supports.[40] Exhibit 5 and 6 in the report presented the findings of the program's effects on this subgroup of students.

The team conducted two types of analyses to assess whether the impact findings for this group of students were sensitive to the group definition. First, the team used a different defining variable to identify students who were initially struggling with behavior. Instead of using teachers' initial ratings on disruptive behavior, the team defined this group by whether they received any office disciplinary referrals (ODRs) as reported by the teachers in the fall of Program Year 1. This alternative definition is likely to be policy-/practitioner-relevant because the identified group has a clear meaning to schools and because some schools implementing MTSS-B do use ODRs to identify students for supplemental supports. Baseline data from the TSSB show that this group constitutes about 8-9 percent of the analysis samples. The correlation between receipt of ODR and the disruptive behavior rating in the fall of 2015 is around 0.5, suggesting that the struggling student subgroup defined by the disruptive behavior ratings and the one defined by receipt of ODR overlapped with each other but also included distinct groups of students.

Exhibit D.5 presents impact findings based on this definition. The top panel of the exhibit shows that, in both program years, the program produced significantly reduced disruptive behavior and emotional dysregulation for the students who were initially identified as struggling with behavior based on this alternative definition. In Program Year 1, the program also significantly increased the pro-social behavior for this group of students. The bottom panel shows the estimated impacts on student achievements for the student subgroups defined by ODR for both program years and the follow-up year. The results suggest that the program did not affect the achievements of the students who received any ODRs as reported by teachers in the fall of Program Year 1.

Exhibit D.1. Sensitivity Analyses for Impacts on Student Behavior, by Sample and Program Year

Year and Outcome (0- to 5-point scale)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
OVERALL SAMPLE									
	Model 1 - As in the report			**Model 2 - No covariate**			**Model 3 - Demographic covariates**		
Program Year 1									
Disruptive behavior	-0.03	0.02	0.100	-0.04	0.03	0.175	-0.03	0.03	0.207
Emotional dysregulation	-0.02	0.02	0.410	-0.04	0.03	0.303	-0.03	0.03	0.337
Internalizing behavior	0.00	0.02	0.966	-0.02	0.03	0.386	-0.02	0.03	0.458
Pro-social behavior	0.01	0.03	0.801	0.04	0.04	0.304	0.03	0.04	0.353
Attention to schoolwork	-0.01	0.02	0.599	0.02	0.04	0.596	0.00	0.03	0.933
Program Year 2									
Disruptive behavior	-0.02	0.02	0.343	-0.02	0.02	0.348	-0.02	0.02	0.334
Emotional dysregulation	-0.02	0.03	0.450	-0.03	0.03	0.413	-0.03	0.03	0.385
Internalizing behavior	-0.03	0.02	0.220	-0.04	0.02	0.154	-0.03	0.02	0.163
Pro-social behavior	0.04	0.03	0.183	0.05	0.04	0.182	0.05	0.03	0.156
Attention to schoolwork	0.01	0.03	0.627	0.03	0.04	0.433	0.02	0.03	0.495

(continued)

Exhibit D.1 (continued)

Year and Outcome (0- to 5-point scale)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
STUDENTS INITIALLY IDENTIFIED AS STRUGGLING WITH BEHAVIOR									
	Model 1 - As in the report			**Model 2 - No covariate**			**Model 3 - Demographic covariates**		
Program Year 1									
Disruptive behavior	-0.06 *	0.02	0.010	-0.06 *	0.03	0.021	-0.05	0.03	0.057
Emotional dysregulation	-0.02	0.03	0.477	-0.01	0.04	0.770	0.00	0.04	0.964
Internalizing behavior	0.00	0.03	0.944	-0.01	0.03	0.799	0.00	0.03	0.955
Pro-social behavior	0.04	0.03	0.212	0.05	0.05	0.232	0.04	0.04	0.314
Attention to schoolwork	0.02	0.03	0.627	0.03	0.05	0.508	0.00	0.05	0.964
Program Year 2									
Disruptive behavior	-0.08 *	0.03	0.015	-0.11 *	0.04	0.003	-0.09 *	0.03	0.006
Emotional dysregulation	-0.06	0.04	0.143	-0.08	0.05	0.084	-0.07	0.05	0.130
Internalizing behavior	-0.07	0.04	0.090	-0.07	0.04	0.080	-0.06	0.04	0.117
Pro-social behavior	0.07	0.05	0.156	0.09	0.05	0.109	0.08	0.05	0.143
Attention to schoolwork	0.02	0.05	0.662	0.06	0.06	0.376	0.03	0.06	0.618

(continued)

Exhibit D.1 (continued)

SOURCES: Teacher Survey of Student Behavior data, collected in fall 2015, spring 2016 (Program Year 1, n = 25,391), and spring 2017 (Program Year 2, n = 24,842). Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The overall student sample used in this table is defined as students with all five behavior measures for that year. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 1 sample of these students has 2,979 observations and the Program Year 2 sample of these students has 1,847 observations.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. All models control for the blocking of random assignment. Model 1 is the primary impact estimation model used in Exhibit C.1, serving as benchmark here. Model 2 contains no covariates other than the random assignment block indicators. Model 3 controls for student background characteristics such as grade, age, gender, race, free-/reduced-price lunch status, English language learner status, and Individualized Education Plan status. All missing covariate values are imputed with zero and missing indicators for all covariates are also included in the model.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Exhibit D.2. Sensitivity Analyses for Impacts on Student Academic Achievement, by Sample and Program Year

Year and Outcome (standardized score)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts

OVERALL SAMPLE

	Model 1 - As in the report			Model 2 - No Covariate			Model 3 - Demographics as Covariates		
Program Year 1									
Standardized reading score	0.01	0.03	0.669	0.05	0.04	0.150	0.01	0.03	0.617
Standardized math score	0.00	0.03	0.952	0.04	0.04	0.367	0.00	0.03	0.998
Program Year 2									
Standardized reading score	0.00	0.03	0.867	0.04	0.04	0.302	0.01	0.03	0.706
Standardized math score	0.00	0.04	0.976	0.05	0.05	0.322	0.02	0.04	0.669
Follow-Up Year									
Standardized reading score	0.01	0.03	0.829	0.04	0.04	0.268	0.01	0.04	0.738
Standardized math score	0.02	0.04	0.576	0.06	0.04	0.181	0.02	0.04	0.570

STUDENTS INITIALLY IDENTIFIED AS STRUGGLING WITH BEHAVIOR

	Model 1 - As in the report			Model 2 - No Covariate			Model 3 - Demographics as Covariates		
Program Year 1									
Standardized reading score	0.11 *	0.05	0.023	0.18 *	0.06	0.004	0.12 *	0.05	0.033
Standardized math score	0.05	0.05	0.296	0.13 *	0.06	0.040	0.07	0.06	0.232

(continued)

Exhibit D.2 (continued)

Year and Outcome (standardized score)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
Program Year 2									
Standardized reading score	0.11 *	0.06	0.050	0.20 *	0.07	0.005	0.11	0.07	0.074
Standardized math score	0.07	0.06	0.209	0.15 *	0.07	0.036	0.08	0.07	0.239
Follow-Up Year									
Standardized reading score	0.09	0.07	0.173	0.20 *	0.08	0.010	0.09	0.07	0.202
Standardized math score	-0.02	0.07	0.817	0.07	0.08	0.341	-0.02	0.07	0.781

SOURCES: Teacher Survey of Student Behavior data, collected in fall 2015. Student records data from the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 school years.

NOTES: For Program Years 1 and 2, the analysis includes students from 58 participating schools and 31 non-participating schools with 24,326 and 23,569 students, respectively. For the Follow-Up Year, the analysis includes students from 52 participating schools and 28 non-participating schools with 19,968 students. The analysis sample used in this table is defined as students with both reading and math state test scores for that year. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 1 sample has 1,879 such students, the Program Year 2 sample has 1,579 such students, and the Follow-Up Year sample has 1,323 such students.

  The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. All models control for the blocking of random assignment. Model 1 is the primary impact estimation model used in Exhibit C.2, serving as benchmark here. Model 2 contains no covariates other than the random assignment block indicators. Model 3 controls for student background characteristics such as grade, age, gender, race, free-/reduced-price lunch status, English language learner status, and Individualized Education Plan status. All missing covariate values are imputed with zero and missing indicators for all covariates are also included in the model.

  A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Exhibit D.3. Sensitivity Analyses for Impacts on Staff Perceptions of School Climate

Measures	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
	Model 1 - As in the report			Model 2 - No covariates			Model 3 - Partial covariates		
Organizational health									
Academic focus	0.07 *	0.04	0.043	0.07 *	0.04	0.048	0.07 *	0.04	0.044
Principal leadership	0.07	0.08	0.376	0.06	0.08	0.401	0.07	0.08	0.386
Staff collegiality	0.14 *	0.06	0.025	0.14 *	0.06	0.025	0.14 *	0.06	0.025
School environment									
Bullying	-0.04	0.05	0.465	-0.03	0.05	0.477	-0.04	0.05	0.473
School safety	0.03	0.03	0.468	0.03	0.03	0.458	0.02	0.03	0.485
Teacher-student relationships	0.07 *	0.03	0.017	0.07 *	0.03	0.015	0.07 *	0.03	0.016
Consistent school discipline	0.10	0.06	0.111	0.10	0.06	0.114	0.10	0.06	0.114
Number of schools	89			89			89		

SOURCE: MTSS-B Staff Survey data collected in spring 2017.

NOTES: The analysis sample used for the survey is defined as all respondents who consented to and answered at least one question in the survey (n = 4,280). The number of respondents varies due to missingness in the constructs. A respondent is counted as non-missing if they responded to at least 80 percent of the questions that constitute the scale.

 The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with staff nested with schools. All models control for the blocking of random assignment. Model 2 also controls for baseline differences between staff members in the participating and non-participating schools with respect to age and gender as reported in the Staff Survey. Model 3 controls for staff member characteristics such as age, gender, experience, education attainment, and certification status as reported in the Staff Survey. All missing covariate values are imputed with zero and missing indicators for all covariates are also included in the model.

 A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than or equal to 5 percent.

Exhibit D.4. Sensitivity Analyses for Impacts on Classroom Outcomes

Outcome (1- to 5-point scale unless otherwise noted)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
	Model 1 - As in the report			**Model 2 - No covariate**			**Model 3 - No mixed grade class**			**Model 4 - No outliers[a]**		
Classroom management practices												
Facilitate orderly classroom transitions	0.14 *	0.06	0.029	0.15	0.11	0.181	0.14 *	0.06	0.032	0.13 *	0.07	0.045
Teacher anticipation and responsiveness to student needs	0.17 *	0.06	0.004	0.09	0.07	0.191	0.18 *	0.06	0.004	0.17 *	0.06	0.006
Teacher proactive behavior management	0.33 *	0.07	0.000	0.34 *	0.08	0.000	0.33 *	0.07	0.000	0.32 *	0.07	0.000
Teacher active monitoring of students	0.11 *	0.04	0.017	0.10	0.06	0.079	0.11 *	0.05	0.016	0.09 *	0.04	0.041

(continued)

Exhibit D.4 (continued)

Outcome (1- to 5-point scale unless otherwise noted)	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Standard Error of Estimated Impacts	p-value of Estimated Impacts
	Model 1 - As in the report			**Model 2 - No covariate**			**Model 3 - No mixed grade class**			**Model 4 - No outliers[a]**		
Classroom functioning												
Student disruptive behavior	-0.04 *	0.02	0.020	-0.04	0.02	0.061	-0.05 *	0.02	0.014	-0.03	0.02	0.079
Student compliance	0.14 *	0.04	0.003	0.12 *	0.05	0.024	0.14 *	0.04	0.002	0.11 *	0.04	0.011
Student engagement with classroom activities	0.16 *	0.06	0.015	0.12	0.08	0.145	0.15 *	0.06	0.018	0.15 *	0.06	0.017
Teacher control of classroom	0.11 *	0.04	0.009	0.11 *	0.04	0.009	0.11 *	0.04	0.010	0.10 *	0.04	0.018
Counts of student disruptive behavior in the classroom (tally)	-1.92 *	0.54	0.001	-2.03 *	0.63	0.002	-2.09 *	0.54	0.000	-1.26 *	0.43	0.005
Number of classrooms	1,791			1,791			1,737			1,761		

(continued)

Exhibit D.4 (continued)

SOURCE: Classroom observations data, collected with Assessing School Settings: Interactions of Students and Teachers (ASSIST) in spring 2017.

NOTES: The analysis includes students from 57 participating schools (1,152 classrooms) and 31 non-participating schools (639 classrooms). The classroom sample used in this table is defined as all non-special education classes in Grades 1-5 in sample schools.

    The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with classes nested within schools. The models control for the blocking of random assignment. Model 1 is the primary impact estimation model used in Exhibit C.1, serving as benchmark here. Model 2 contains no covariates other than the random assignment block indicators. Model 3 has the same model specification as Model 1 but excludes mixed-grade classes (sample size = 1,123 classrooms in participating schools and 614 classrooms in non-participating schools). Model 4 has the same model specification as Model 1 but excludes classrooms identified as outliers (sample size = 1,139 classrooms in participating schools and 622 classrooms in non-participating schools). All missing values are imputed with zero and missing indicators for all covariates are also included in the model.

    A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

    [a] A classroom is considered an outlier if its count of student problem behaviors in the classroom is more than three standard deviations above or below the average count across all classrooms. By this definition, 30 classrooms in the sample are identified as outliers.

Exhibit D.5. Estimated Impacts on Student Outcomes by Whether Students Were Initially Identified as Struggling with Behavior Defined by Office Disciplinary Referral Status in Fall 2015, by Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
STUDENT BEHAVIOR OUTCOME (0- TO 5-POINT SCALE)									
Program Year 1									
Disruptive behavior							0.07	0.006 †	
Students initially identified as struggling with behavior	1.55	1.63	-0.08 *	0.03	-0.11	0.001			
Students not initially identified	0.47	0.49	-0.02	0.01	-0.02	0.231			
Emotional dysregulation							0.10	0.001 †	
Students initially identified as struggling with behavior	2.11	2.22	-0.11 *	0.03	-0.12	0.002			
Students not initially identified	1.04	1.04	0.00	0.02	-0.01	0.812			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Internalizing behavior							0.06	0.032 †	
Students initially identified as struggling with behavior	0.92	0.98	-0.06	0.03	-0.08	0.064			
Students not initially identified	0.64	0.63	0.00	0.02	0.00	0.909			
Attention to schoolwork							-0.09	0.022 †	
Students initially identified as struggling with behavior	2.42	2.34	0.08	0.04	0.06	0.058			
Students not initially identified	3.60	3.61	-0.01	0.02	-0.01	0.616			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Pro-social behavior							-0.06	0.067	
Students initially identified as struggling with behavior	3.01	2.93	0.08 *	0.04	0.08	0.040			
Students not initially identified	4.02	4.00	0.02	0.02	0.02	0.434			
Students initially identified as struggling with behavior (n)									1,821
Students not initially identified (n)									18,365
Program Year 2									
Disruptive behavior							0.12	0.001 †	
Students initially identified as struggling with behavior	1.22	1.36	-0.14 *	0.04	-0.19	0.001			
Students not initially identified	0.46	0.48	-0.01	0.02	-0.02	0.515			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Emotional dysregulation							0.11	0.018 †	
Students initially identified as struggling with behavior	1.80	1.91	-0.11 *	0.05	-0.13	0.039			
Students not initially identified	1.04	1.04	0.00	0.03	0.00	0.986			
Internalizing behavior							0.05	0.190	
Students initially identified as struggling with behavior	0.84	0.92	-0.08	0.05	-0.12	0.083			
Students not initially identified	0.64	0.67	-0.03	0.03	-0.04	0.329			
Attention to schoolwork							-0.14	0.027 †	
Students initially identified as struggling with behavior	2.65	2.51	0.13	0.07	0.11	0.052			
Students not initially identified	3.59	3.59	-0.01	0.04	-0.01	0.817			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Pro-social behavior							-0.04	0.389	
Students initially identified as struggling with behavior	3.25	3.18	0.07	0.06	0.07	0.249			
Students not initially identified	4.03	4.00	0.02	0.04	0.03	0.508			
Students initially identified as struggling with behavior (n)									1,043
Students not initially identified (n)									11,984

STUDENT ACHIEVEMENT OUTCOMES (STANDARDIZED SCORE)

Program Year 1

Reading							0.07	0.203	
Students initially identified as struggling with behavior	-0.20	-0.26	0.06	0.06	0.06	0.263			
Students not initially identified	0.12	0.12	-0.01	0.03	-0.01	0.849			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Math							0.03	0.548	
Students initially identified as struggling with behavior	-0.20	-0.23	0.02	0.06	0.02	0.684			
Students not initially identified	0.12	0.12	-0.01	0.03	-0.01	0.814			
Students initially identified as struggling with behavior (n)									1,245
Students not initially identified (n)									11,626
Program Year 2									
Reading							0.11	0.090	
Students initially identified as struggling with behavior	-0.20	-0.29	0.09	0.07	0.09	0.202			
Students not initially identified	0.15	0.17	-0.02	0.03	-0.02	0.495			

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Math							0.16	0.014 †	
Students initially identified as struggling with behavior	-0.15	-0.29	0.14	0.07	0.14	0.050			
Students not initially identified	0.15	0.17	-0.01	0.04	-0.01	0.729			
Students initially identified as struggling with behavior (n)									880
Students not initially identified (n)									10,414
Follow-Up Year									
Reading							0.05	0.510	
Students initially identified as struggling with behavior	-0.29	-0.30	0.02	0.08	0.02	0.834			
Students not initially identified	0.12	0.16	-0.03	0.04	-0.03	0.419			

(continued)

Exhibit D.5 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Math							0.05	0.492	
Students initially identified as struggling with behavior	-0.27	-0.28	0.01	0.08	0.01	0.910			
Students not initially identified	0.15	0.20	-0.04	0.05	-0.04	0.365			
Students initially identified as struggling with behavior (n)									728
Students not initially identified (n)									8,759

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from school years 2014-2015, 2015-2016, and 2016-2017.

NOTES: The analysis includes students from 58 participating schools and 31 non-participating schools. The student sample used in this table is defined as students with all five behavior measures for that year. A student is initially identified as struggling with behavior if his/her teacher reported any office disciplinary referral incidence in the Teacher Survey of Student Behavior (TSSB) in fall 2015.

The impacts on student behavior are estimated using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The impacts on student achievement are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating group members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Second, to examine whether the subgroup findings were sensitive to the selected cut-off values used to define the group of students initially struggling with behavior, the team re-estimated the program's effects on these students using a series of different cut-off points, ranging from 5 percent to 30 percent. Exhibit D.6a summarizes the impact findings on the student disruptive behavior rating for this group of students (as defined by teachers' initial rating on their disruptive behavior) from such exploration for Program Years 1 and 2, respectively. It shows that, in both program years, the estimated program effect on disruptive behavior largely holds when the cut-off value was set at 20 percent or below. As the cut-off value increased, in other words, as this group included more students with relatively less struggles with behavior, the magnitude of the estimated program effect started to dissipate. This pattern was especially pronounced for Program Year 2. Exhibit D.6b shows the impact findings on student reading achievement for students identified by these varying cut-off values for Program Years 1 and 2, as well as the Follow-Up Year. The figure reveals a linear relationship between the magnitude of the impact estimates and the cut-off point choices for Program Year 2 and a non-linear relationship for the other two years.

Exhibit D.6a. Estimated Impacts on Student Disruptive Behavior for Students Initially Identified as Struggling with Behavior with Different Cut Point Choice for Subgroup Definition, by Program Year
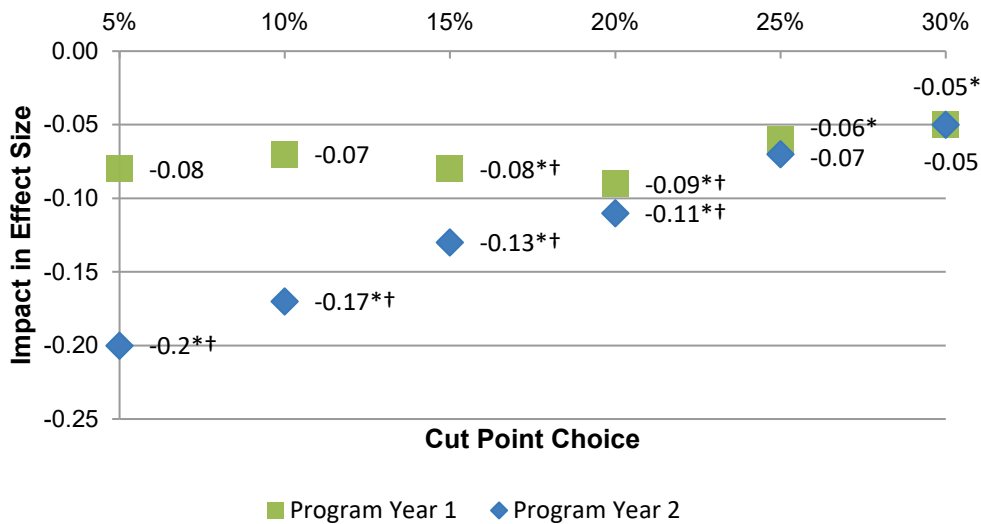
SOURCE: Teacher Survey of Student Behavior data, collected in fall of 2015, spring of 2016 and spring of 2017. District records data for 2015-2016 school year.

NOTES: The analysis sample used in this figure is defined as students with all five behavior measures for that year. Each point in the figure represents the estimated impact on student disruptive behavior rating for students initially identified as struggling with behavior. These students were defined based on their disruptive behavior ratings from the fall of 2015. In the report, this group of students includes students with the highest 15 percent of baseline disruptive behavior ratings. This exhibit presents estimated impacts for such students when they are identified by different cut point values varying from 5 percent to 30 percent.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between students initially identified as strugling with behavior and those who were not identified. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit D.6b. Estimated Impacts on Student Reading Achievement for Students Initially Identified as Struggling with Behavior with Different Cut Point Choice for Subgroup Definition, by Program Year
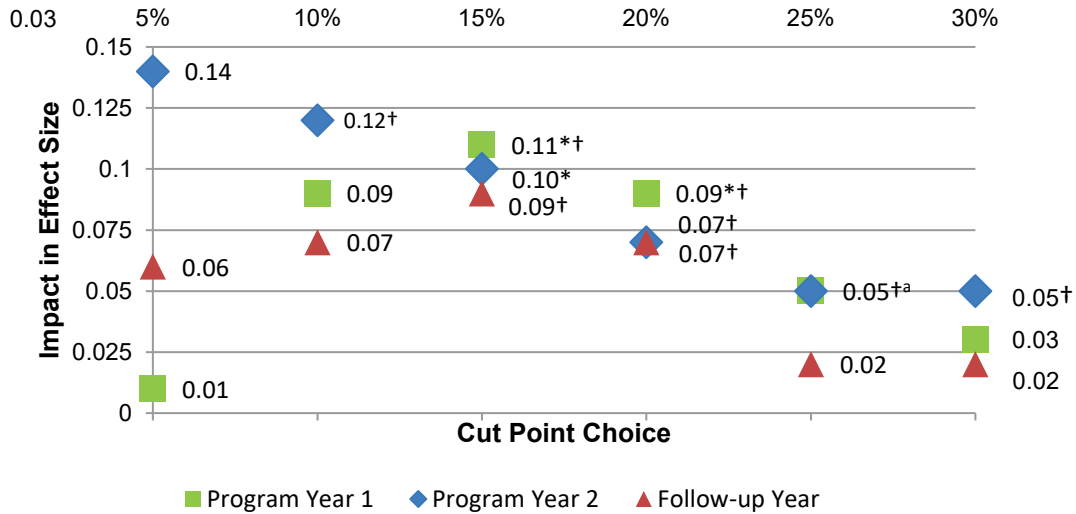

SOURCES: District records data for 2015-2016, 2016-2017, and 2017-2018 school years. Teacher Survey of Student Behavior data from fall 2015.

NOTES: The analysis sample used in this figure is defined as students with state English Language Arts (ELA) test scores for that year. Each point in the figure represents the estimated impact on student ELA scores for students initially identified as struggling with behavior. These students were defined based on their disruptive behavior ratings from the fall of 2015. In the report, this group of students includes students with the highest 15 percent of baseline disruptive behavior ratings. This exhibit presents estimated impacts for such students when they are identified by different cut point values, varying from 5 percent to 30 percent.

All test scores are standardized within school district and grade level using the means and standard deviations of the full non-participating schools' students in each school-by-grade cell. The estimated impacts are therefore in effect size unit.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

[a] For the 25 percent cut point, estimated impacts for both program years have the same magnitude, but the F-test is only statistically significant for Program Year 2.

Overall, these sensitivity checks confirmed that the program effect findings for students initially identified as struggling with behavior were robust to different group definitions.

III. Sensitivity Checks on Potential Attrition Bias

As shown in Exhibit B.5a (in Appendix B), there were differential attritions between students in the participating and non-participating schools across all three rounds of collections for the Teacher Survey of Student Behavior (TSSB). The differential attritions were mostly driven by the differential parental consent rates between the participating and non-participating groups for this data collection rather than differential response rates among students with parental

consent. The team conducted multiple analyses to assess whether or how these differential response rates might affect the interpretation of the program effect findings based on students with teacher ratings on their behaviors.

To start, the team checked the *internal validity* of the TSSB analysis samples for both program years. Recall that the TSSB analysis sample for a given program year included all Grade 1-5 students with teacher ratings in the spring of that program year. Exhibits B.12 and B.13 show that, despite the differential response rates between the participating and non-participating groups, students in the TSSB analysis sample from these two groups of schools were similar to each other across a wide range of baseline characteristics, including the pre-program measures of their behaviors and academic performances. These findings provide supportive evidence for the internal validity of the TSSB analysis samples that the study used to assess program effects on student behavior.

However, there was evidence that the students whose teacher ratings were available for Program Year 1 or Program Year 2 (the "respondents") and those who did not have teacher ratings for the program years (the "non-respondents") were different in terms of their background characteristics (see Exhibits D.7 and D.8 for Program Years 1 and 2, respectively). For baseline achievement measures and most of the demographic characteristics available for both respondents and non-respondents, the differences between the two groups are statistically significant. These comparisons were based on eight of the nine study districts. One study district did not provide the study with student records data for the non-consented students and was excluded from this analysis.

The team conducted two sensitivity checks to assess whether the impact findings based on the respondent sample could be generalized to the broader sample of all enrolled students. These checks were only possible for eight of the nine study districts because one study district did not provide records data for non-consented students.

The team first used student characteristics information collected through district records to create weights for all students with teacher ratings of behavior and checked the program impacts on behavioral outcomes with and without this weight. The team used the propensity score approach to generate the response weights. The team identified a set of predictors for whether a student received teacher ratings in a given program year, including all demographic characteristics, random assignment block indicators, and grade levels. The team then determined the probability that someone would have the teacher ratings in a program year using a logistic regression with the identified predictors. Lastly, the team calculated an individual's weight as the inverse of their predicted probability. The results show that the findings were similar with or without the use of these weights (see Exhibits D.9 and D.10), suggesting that even though the response rates differ significantly by program status, the differential response rate did not appear to have affected the impact findings in substantive ways.

The team then estimated the program effects on academic achievement for students with and without teacher ratings of behavior separately. The results show that except for Program Year 1 math achievement, the estimated impacts did not differ between these two groups of students (Exhibit D.11).

The findings presented in this section suggest that, even though the respondents and non-respondents differed in their background characteristics, the program did not appear to have affected them differently.

IV. Program Effects on an Alternative Consistent Student Sample

This section explores the program's effects on student behavior and achievement outcomes using a consistent student sample across years to assess whether the program effect findings presented in the report were sensitive to changing sample definitions and composition. The results presented below show that the impact findings in the report were robust to these changes.

The program effect findings presented in the report were based on the largest available student sample for a given outcome each year. In other words, the analysis samples for estimating program effects on student behavior included all Grade 1-5 students with valid behavior outcomes in a given program year. The analysis samples for estimating program effects on student academic achievement included all Grade 3-5 students with reading and math state test scores in a given year. Defining the analysis samples this way maximized the sample size and the statistical power of the estimation.

Exhibit D.7. Background Characteristics Comparison of Students With and Without Behavior Ratings, Program Year 1 Sample

Characteristic (%)	Students With Teacher Ratings	Students Without Teacher Ratings	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics						
Age (year)	6.99	7.07	-0.07 *	1.52	-0.05	0.000
Male	51.19	53.10	-1.91 *	0.53	-0.04	0.000
Race/ethnicity						
Black, non-Hispanic	8.83	11.68	-2.85 *	0.26	-0.10	0.000
White, non-Hispanic	15.21	13.86	1.35 *	0.30	0.04	0.000
Hispanic	67.38	67.11	0.26	0.36	0.01	0.462
Asian	6.42	5.14	1.28 *	0.23	0.06	0.000
Other	2.16	2.20	-0.04	0.15	0.00	0.784
Free/reduced lunch status	69.83	71.12	-1.29 *	0.43	-0.03	0.003
English language learner status	40.79	39.96	0.83	0.49	0.02	0.093
Special education status	9.66	10.91	-1.25 *	0.34	-0.04	0.000
Baseline grade levels						
Grade K (Grade 1 in Year 1)	19.51	17.61	1.90 *	42.12	0.05	0.000
Grade 1 (Grade 2 in Year 1)	20.24	20.05	0.19	42.38	0.00	0.652
Grade 2 (Grade 3 in Year 1)	20.19	20.80	-0.61	42.47	-0.02	0.152
Grade 3 (Grade 4 in Year 1)	20.25	20.42	-0.17	42.69	0.00	0.686
Grade 4 (Grade 5 in Year 1)	19.81	21.13	-1.31 *	42.31	-0.03	0.002
Baseline achievement measures[a]						
Reading (in effect size unit)	0.03	-0.13	0.16 *	1.84	0.15	0.000
Math (in effect size unit)	0.03	-0.13	0.16 *	1.82	0.16	0.000

(continued)

Exhibit D.7 (continued)

SOURCES: Teacher Survey of Student Behavior from the fall of 2015 and spring of 2016, and district records data collected for the 2014-2015 and 2015-2016 school years.

NOTES: The sample includes 39,956 eligible Grades 1-5 students in 80 of the 89 study schools in the spring of 2016 who had district records data for spring 2016. The numbers of observations vary by baseline characteristics due to missing values.

    The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within schools. The models control for indicators of random assignment blocks. Rounding may cause slight discrepancies in calculating sums and differences.

   *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

   [a]The state standardized test scores were only available for students who were in Grades 3 and 4 in the baseline year (2014-2015 school year).

Exhibit D.8. Background Characteristics Comparison of Students With and Without Behavior Ratings, Program Year 2 Sample

Characteristic (%)	Students With Teacher Ratings	Students Without Teacher Ratings	Estimated Difference	Standard Error of Estimated Difference	Estimated Difference in Effect Size	p-value of Estimated Difference
Student demographic characteristics						
Age (year)	6.02	6.00	0.02	1.55	0.01	0.247
Male	51.39	52.94	-1.55 *	0.54	-0.03	0.004
Race/ethnicity						
Black, non-Hispanic	8.42	12.09	-3.66 *	0.26	-0.12	0.000
White, non-Hispanic	15.14	13.89	1.25 *	0.31	0.04	0.000
Hispanic	67.58	66.19	1.39 *	0.36	0.03	0.000
Asian	6.52	5.69	0.83 *	0.24	0.04	0.000
Other	2.34	2.13	0.21	0.16	0.01	0.192
Has free/reduced-price lunch status	70.89	72.24	-1.35 *	0.49	-0.03	0.006
Has English language learner status	42.79	40.86	1.93 *	0.58	0.04	0.001
Has special education status	9.68	11.66	-1.97 *	0.40	-0.06	0.000
Baseline grade levels						
Grade Pre-K (Grade 1 in Year 2)	19.61	20.27	-0.65	42.90	-0.02	0.127
Grade K (Grade 2 in Year 2)	19.56	18.18	1.38 *	42.80	0.04	0.001
Grade 1 (Grade 3 in Year 2)	20.39	21.66	-1.27 *	43.49	-0.03	0.004
Grade 2 (Grade 4 in Year 2)	20.31	20.93	-0.62	43.52	-0.02	0.154
Grade 3 (Grade 5 in Year 2)	20.12	19.04	1.08 *	43.50	0.03	0.013
Baseline achievement measures[a]						
Reading (in effect size unit)	0.05	-0.10	0.15 *	2.80	0.15	0.000
Math (in effect size unit)	0.06	-0.10	0.17 *	2.74	0.17	0.000

(continued)

Exhibit D.8 (continued)

NOTES: The sample includes 38,601 eligible Grades 1-5 students in 80 of the 89 study schools in spring 2017 who had district records data for spring 2017. The number of observations varies by baseline characteristics due to missing values.

The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. Rounding may cause slight discrepancies in calculating sums and differences.

*Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed test.

[a] The state standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015 school year).

Exhibit D.9. Estimated Impacts on Student Behavior Outcomes With or Without Response Weights, Overall Sample, by Program Year

Measures (0- to 5-point scale)	Without Response Weight				With Response Weight			
	Participating Schools	Non-Participating Schools	Estimated Impact	p-value of Estimated Impact	Participating Schools	Non-Participating Schools	Estimated Impact	p-value of Estimated Impact
Program Year 1								
Disruptive behavior	0.58	0.61	-0.03	0.104	0.58	0.60	-0.02	0.160
Emotional dysregulation	1.13	1.16	-0.03	0.217	1.13	1.16	-0.03	0.225
Internalizing behavior	0.67	0.67	0.01	0.726	0.67	0.67	0.00	0.828
Attention to schoolwork	3.45	3.47	-0.02	0.477	3.45	3.47	-0.02	0.337
Pro-social behavior	3.92	3.91	0.01	0.789	3.92	3.91	0.01	0.754
Number of schools	52	28			52	28		
Program Year 2								
Disruptive behavior	0.55	0.57	-0.02	0.396	0.55	0.57	-0.02	0.435
Emotional dysregulation	1.12	1.13	-0.02	0.586	1.12	1.13	-0.01	0.629
Internalizing behavior	0.67	0.69	-0.02	0.431	0.67	0.70	-0.02	0.362
Attention to schoolwork	3.45	3.44	0.01	0.635	3.45	3.44	0.01	0.740
Pro-social behavior	3.94	3.90	0.04	0.227	3.94	3.90	0.04	0.239
Number of schools	52	28			52	28		

(continued)

Exhibit D.9 (continued)

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from the 2014-2015 school year.

NOTES: The respondent group includes all students with valid teacher ratings for all five teacher rating measures in Program Year 1 and Program Year 2, respectively. There are 80 schools in this analysis because one district did not provide records information for students who did not consent to the data collection.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. All models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, and Individualized Education Plan status; baseline standardized math and reading test scores; and baseline teacher rating of student behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values for the participating schools are the weighted average of the observed district means for schools or students randomly assigned to the participating schools (using number of participating group schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The response weights are the inverse of the probability of having behavior ratings in a given program year as predicted by student background demographic characteristics.

A two-tailed t-test was applied to each estimated difference. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit D.10. Estimated Impacts on Student Behavior Outcomes With or Without Response Weights by Whether Students Were Initially Identified as Struggling with Behavior, by Program Year

Measures (0- to 5-point scale)	Without Response Weight						With Response Weight						Number of Observations
	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	
Program Year 1													
Disruptive behavior					0.06	0.004 †					0.07	0.019 †	
Students initially identified as struggling with behavior	1.64	1.71	-0.07 *	0.006			1.64	1.71	-0.07 *	0.023			
Students not initially identified	0.38	0.39	-0.01	0.689			0.38	0.38	0.00	0.897			
Emotional dysregulation					0.02	0.416					0.03	0.407	
Students initially identified as struggling with behavior	2.17	2.21	-0.04	0.228			2.17	2.21	-0.04	0.272			
Students not initially identified	0.94	0.95	-0.02	0.471			0.94	0.95	-0.01	0.609			
Internalizing behavior					0.01	0.713					0.02	0.543	
Students initially identified as struggling with behavior	1.10	1.10	-0.01	0.779			1.10	1.11	-0.01	0.626			
Students not initially identified	0.58	0.58	0.00	0.962			0.58	0.58	0.00	0.913			
Attention to schoolwork					-0.02	0.476					-0.02	0.665	
Students initially identified as struggling with behavior	2.43	2.42	0.01	0.842			2.43	2.42	0.00	0.941			
Students not initially identified	3.67	3.69	-0.02	0.463			3.67	3.69	-0.01	0.545			

(continued)

Exhibit D.10 (continued)

Measures (0- to 5-point scale)	Without Response Weight						With Response Weight						Number of Observations
	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	
Pro-social behavior					-0.03	0.352					-0.04	0.303	
Students initially identified as struggling with behavior	2.93	2.88	0.05	0.182			2.93	2.87	0.06	0.149			
Students not initially identified	4.12	4.10	0.02	0.435			4.12	4.10	0.02	0.442			
Students initially identified as struggling with behavior (n)													2,696
Students not initially identified (n)													15,717
Program Year 2													
Disruptive behavior					0.08	0.010 †					0.08	0.074	
Students initially identified as struggling with behavior	1.18	1.26	-0.08 *	0.018			1.18	1.26	-0.08	0.056			
Students not initially identified	0.41	0.41	0.00	0.946			0.41	0.41	-0.01	0.773			
Emotional dysregulation					0.08	0.051					0.07	0.157	
Students initially identified as struggling with behavior	1.74	1.79	-0.05	0.262			1.74	1.80	-0.05	0.329			
Students not initially identified	0.97	0.95	0.02	0.488			0.97	0.96	0.02	0.617			
Internalizing behavior					0.03	0.334					0.06	0.199	
Students initially identified as struggling with behavior	0.86	0.89	-0.03	0.414			0.86	0.92	-0.06	0.221			
Students not initially identified	0.63	0.63	0.00	0.969			0.63	0.63	0.00	0.916			

(continued)

Exhibit D.10 (continued)

Measures (0- to 5-point scale)	Without Response Weight						With Response Weight						Number of Observations
	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Participating Schools	Non-Participating Schools	Estimated Impacts	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	
Attention to schoolwork					-0.04	0.476					-0.05	0.488	
Students initially identified as struggling with behavior	2.68	2.66	0.02	0.790			2.68	2.65	0.03	0.711			
Students not initially identified	3.65	3.67	-0.02	0.556			3.65	3.67	-0.02	0.598			
Pro-social behavior					-0.07	0.094					-0.10	0.073	
Students initially identified as struggling with behavior	3.30	3.23	0.07	0.191			3.30	3.20	0.10	0.112			
Students not initially identified	4.09	4.09	0.00	0.937			4.09	4.09	0.00	0.971			
Students initially identified as struggling with behavior (n)													1,686
Students not initially identified (n)													10,176

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from the 2014-2015 school year.

NOTES: The respondent group includes all students with valid teacher ratings for all five teacher rating measures in Program Year 1 and Program Year 2, respectively. There are 80 schools in this analysis because one district did not provide records information for students who did not consent to the data collection. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. All models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, and Individualized Education Plan status; baseline standardized math and reading test scores; and baseline teacher rating of student behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values for the participating schools are the weighted average of the observed district means for schools or students randomly assigned to the participating schools (using number of the participating group schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The response weights are the inverse of the probability of having behavior ratings in a given program year as predicted by student background demographic characteristics.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit D.11. Estimated Impacts on Student Academic Achievement by Whether Students Had Behavior Outcomes, by Program Year

Measure	MTSS-B Program Schools	Non-MTSS-B Program Schools	Estimated Impact	Standard Error of Estimated Impact	p-value of Estimated Impact	Estimated Subgroup Difference	p-value of Estimated Subgroup Difference	Number of Observations
Program Year 1								
Standardized reading scores						0.02	0.451	
Students with behavior outcomes	0.07	0.05	0.02	0.03	0.532			
Students with no behavior outcomes	-0.07	-0.07	0.00	0.03	0.938			
Standardized math scores						0.05	0.041 †	
Students with behavior outcomes	0.07	0.05	0.02	0.03	0.589			
Students with no behavior outcomes	-0.10	-0.07	-0.04	0.03	0.294			
Students with behavior outcomes (n)								14,432
Students with no behavior outcomes (n)								9,894
Program Year 2								
Standardized reading scores						0.00	0.985	
Students with behavior outcomes	0.07	0.08	-0.01	0.03	0.818			
Students with no behavior outcomes	-0.03	-0.02	-0.01	0.03	0.831			
Standardized math scores						-0.02	0.426	
Students with behavior outcomes	0.08	0.09	-0.01	0.04	0.753			
Students with no behavior outcomes	-0.03	-0.04	0.01	0.04	0.803			
Students with behavior outcomes (n)								14,154
Students with no behavior outcomes (n)								9,415

(continued)

Exhibit D.11 (continued)

SOURCE: Teacher Surveys of Student Behavior (TSSB) conducted in fall 2015, spring 2016 and spring 2017 and district records collected in 2015, 2016, and 2017.

NOTES: The analysis sample used in this exhibit is defined as students with both reading and math state test scores for that year. The subgroup is defined by whether a student has teacher ratings of behavior measures for a given year.

The estimated impacts are regression-adjusted using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating schools (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impact effect sizes are calculated as a proportion of the standard deviation of the full non-MTSS-B program school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Alternatively, the study could use a narrower sample definition to estimate the program's effects on student outcomes. The study team defined a consistent student sample that included all students with behavioral outcomes and reading and math achievement outcomes for *both* Program Year 1 and Program Year 2, and with disruptive behavior ratings from the TSSB collection in the fall of 2015. This consistent sample was much smaller than the sample used for the analyses in the report, with 5,870 students for the overall student sample and 785 students initially identified as struggling with behavior. It included only students with complete data for both types of outcomes across both program years and baseline behavior ratings needed to form the subgroup. It also only included students who were in Grades 3 or 4 in Program Year 1 and remained in study schools through Program Year 2. Students below Grade 3 in Program Year 1 would not have achievement data in that year, and students in Grade 5 in Program Year 1 exited the study sample in Program Year 2. On the other hand, students in this consistent sample stayed with the program from the start of the study to the end of Program Year 2 and thus had the maximum amount of exposure to the program.

Findings for this consistent sample largely corroborate the findings presented in the report. For the consistent sample of overall students, by and large, the program did not affect any of the student outcomes with one exception in Program Year 1 (see Exhibit D.12). For students initially identified as struggling with behavior in the consistent sample, the program reduced students' disruptive behavior and increased their reading achievement in both program years (see Exhibit D.13). The magnitudes of these changes appeared to be larger than those based on the broader student sample. Students in this consistent sample were exposed to the program the most.

V. Program Effects on Student Behavior for Additional Student Subgroups

To explore possible heterogeneity in program effects across different student populations, the team examined additional student-level subgroups based on students' gender, age, special education status, and English learner status. It estimated the program effects on the behaviors of these groups of students separately for both program years.

Gender subgroup: It is well documented that students' developmental trajectories differ by gender, and boys tend to have more behavioral issues than girls in general.[41] For this reason, one might expect the program to affect boys and girls differently. However, this study showed that, while the boys in the study on average exhibited more behavioral issues than the girls in both years, by and large the program did not change the behaviors for either the boys or the girls overall. The only exception was that the program appeared to have reduced the disruptive behavior of boys by a small amount (effect size = 0.05), but the differences in impacts between these two subgroups were not statistically significant (Exhibit D.14).

Age subgroup: Age plays an important role in students' behavioral development: students in Grades 1 and 2 are new to the school environment, and their focus is on the basic tools of learning in an academic setting; students in Grades 3-5 are transitioning from learning to read to reading to learn and therefore are exposed to instructional practices that differ from lower grades.[42] Consequently, one would expect that students in different grade levels interact with the program differently, potentially generating different finding patterns. The team estimated the program effects on student behavior outcomes separately for students in lower grades (Grades 1 and 2) and upper grades (Grades 3-5) and found that the program significantly reduced the disruptive behavior ratings for program students in Grades 3-5 in Program Year 1 (estimated impact = -0.06 SD, p-value = 0.037). However, the difference in impacts between lower- and upper-grade students on this measure was not significant. Such a pattern was not observed for Program Year 2 (Exhibit D.15).

Special education status subgroup: Policymakers are interested in whether the MTSS-B program could reduce behavioral problems among students receiving special education services. Ideally the study would like to define the subgroups by students' behavior-related special education categories. However, such information was not available to the study, so the authors used the broad categorization of whether a student receives an Individualized Education Program (IEP) prior to the program to define the subgroups. In general, the program did not affect students receiving special education differently from those who did not receive special education (Exhibit D.16). The only exceptions are that the program appeared to have reduced students' emotional dysregulation behavior for students receiving special education in Program Year 1 and have increased the pro-social behaviors for these students in Program Year 2. However, these findings were not statistically different from those for students not receiving special education, and these effects were not observed consistently across the two program years.

Exhibit D.12. Estimated Impacts on Student Outcomes for a Consistent Sample of Students, Overall Sample, by Program Year

Outcome	Program Year 1			Program Year 2		
	Estimated Impacts	Effect Size of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Effect Size of Estimated Impacts	p-value of Estimated Impacts
Teacher rating of student behavior (0- to 5-point scale)						
Disruptive behavior	-0.06 *	-0.08	0.020	-0.05	-0.08	0.114
Emotional dysregulation	-0.03	-0.04	0.351	-0.03	-0.04	0.467
Internalizing behavior	-0.04	-0.05	0.200	-0.03	-0.04	0.494
Attention to schoolwork	0.01	0.01	0.697	0.00	0.00	0.929
Pro-social behavior	0.02	0.02	0.535	-0.01	-0.01	0.905
Academic achievement (standardized score)						
Reading	0.01	0.01	0.714	0.02	0.01	0.685
Math	0.01	0.01	0.789	0.08	0.08	0.074

SOURCES: Teacher Survey of Student Behavior data, collected in fall 2015, spring 2016, and spring 2017. Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample consists of 58 participating schools and 31 non-participating schools. It includes all students with behavioral outcomes and reading and math achievement outcomes for both Program Year 1 and Program Year 2, and with disruptive behavior ratings from the fall of 2015 TSSB collection (n=5,870).

The impacts are estimated using multi-level hierarchical linear models to account for the nested structure of the data. The models control for the blocking of random assignment and for baseline differences between students in the MTSS-B participating and non-participating schools with respect to the following variables: grade, age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The estimated impact effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit D.13. Estimated Impacts on Student Outcomes for a Consistent Sample of Students Initially Identified as Struggling with Behavior, by Program Year

Outcome	Program Year 1			Program Year 2		
	Estimated Impacts	Effect Size of Estimated Impacts	p-value of Estimated Impacts	Estimated Impacts	Effect Size of Estimated Impacts	p-value of Estimated Impacts
Teacher rating of student behavior (0- to 5-point scale)						
Disruptive behavior	-0.13 *	-0.19	0.001	-0.18 *	-0.27	0.000
Emotional dysregulation	-0.05	-0.06	0.318	-0.06	-0.08	0.357
Internalizing behavior	-0.01	-0.01	0.864	-0.09	-0.13	0.117
Attention to schoolwork	0.00	0.00	0.986	0.15	0.13	0.066
Pro-social behavior	0.04	0.04	0.462	0.10	0.11	0.175
Academic achievement (standardized score)						
Reading	0.20 *	0.20	0.009	0.15 *	0.14	0.037
Math	0.04	0.04	0.582	0.08	0.08	0.290

SOURCES: Teacher Surveys of Student Behavior data, collected in fall 2015, spring 2016, and spring 2017. Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample consists of 58 participating schools and 31 non-participating schools. It includes students who are initially identified as struggling with behavior and who have behavioral outcomes and reading and math achievement outcomes for both Program Year 1 and Program Year 2, and have disruptive behavior ratings from the fall of 2015 TSSB collection (n=785). A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools.

The impacts are estimated using multi-level hierarchical linear models to account for the nested structure of the data. The models control for the blocking of random assignment and for baseline differences between students in the MTSS-B participating and non-participating schools with respect to the following variables: grade, age, gender, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

Exhibit D.14. Estimated Impacts on Student Behavior, by Student Gender and Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							0.02	0.173	
Male	0.70	0.74	-0.04 *	0.02	-0.05	0.041			
Female	0.46	0.48	-0.02	0.02	-0.02	0.322			
Emotional dysregulation							0.03	0.089	
Male	1.30	1.33	-0.03	0.02	-0.04	0.154			
Female	1.00	1.01	0.00	0.02	-0.01	0.843			
Internalizing behavior							0.01	0.502	
Male	0.70	0.71	-0.01	0.02	-0.01	0.709			
Female	0.65	0.65	0.00	0.02	0.00	0.925			
Attention to schoolwork							-0.03	0.229	
Male	3.18	3.18	0.00	0.02	0.00	0.969			
Female	3.72	3.75	-0.03	0.02	-0.02	0.283			
Pro-social behavior							-0.03	0.141	
Male	3.76	3.74	0.02	0.03	0.02	0.511			
Female	4.05	4.06	-0.01	0.03	-0.01	0.755			

(continued)

Exhibit D.14 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Male (n)									12,852
Female (n)									12,456
Program Year 2									
Disruptive behavior							0.01	0.428	
Male	0.68	0.71	-0.02	0.02	-0.03	0.228			
Female	0.43	0.44	-0.01	0.02	-0.02	0.557			
Emotional dysregulation							-0.03	0.074	
Male	1.30	1.30	0.00	0.03	-0.01	0.875			
Female	0.97	1.01	-0.04	0.03	-0.04	0.184			
Internalizing behavior							-0.03	0.046 †	
Male	0.71	0.73	-0.01	0.03	-0.02	0.607			
Female	0.64	0.68	-0.04	0.03	-0.06	0.083			
Attention to schoolwork							0.03	0.276	
Male	3.18	3.18	0.00	0.03	0.00	0.968			
Female	3.73	3.70	0.03	0.03	0.02	0.393			
Pro-social behavior							0.01	0.570	
Male	3.77	3.73	0.03	0.03	0.04	0.287			
Female	4.09	4.04	0.05	0.03	0.05	0.159			

(continued)

Exhibit D.14 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Male (n)									12,629
Female (n)									12,158

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The analysis sample used in this exhibit is defined as students with all five measures for that year and gender information from the 2014-2015 school year.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating group (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit D.15. Estimated Impacts on Student Behavior, by Student Grade Group and Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							-0.04	0.174	
Grades 1-2	0.61	0.61	0.00	0.02	-0.01	0.864			
Grades 3-5	0.57	0.61	-0.04 *	0.02	-0.06	0.037			
Emotional dysregulation							-0.05	0.244	
Grades 1-2	1.15	1.15	0.01	0.03	0.01	0.824			
Grades 3-5	1.15	1.19	-0.04	0.03	-0.05	0.144			
Internalizing behavior							-0.07	0.053	
Grades 1-2	0.67	0.63	0.04	0.03	0.06	0.150			
Grades 3-5	0.69	0.72	-0.03	0.02	-0.04	0.219			
Attention to schoolwork							0.04	0.356	
Grades 1-2	3.46	3.50	-0.03	0.03	-0.03	0.331			
Grades 3-5	3.44	3.43	0.01	0.03	0.01	0.797			
Pro-social behavior							0.01	0.821	
Grades 1-2	3.95	3.95	0.00	0.04	0.00	0.914			
Grades 3-5	3.86	3.85	0.01	0.03	0.01	0.661			
Grades 1-2 (n)									10,198
Grades 3-5 (n)									15,146

(continued)

Exhibit D.15 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 2									
Disruptive behavior							0.00	0.971	
Grades 1-2	0.58	0.60	-0.02	0.03	-0.02	0.543			
Grades 3-5	0.55	0.57	-0.02	0.02	-0.02	0.513			
Emotional dysregulation							0.04	0.374	
Grades 1-2	1.14	1.19	-0.04	0.04	-0.05	0.273			
Grades 3-5	1.14	1.14	0.00	0.03	0.00	0.972			
Internalizing behavior							0.01	0.762	
Grades 1-2	0.66	0.70	-0.04	0.04	-0.05	0.305			
Grades 3-5	0.69	0.71	-0.02	0.03	-0.03	0.459			
Attention to schoolwork							-0.01	0.899	
Grades 1-2	3.47	3.45	0.01	0.04	0.01	0.745			
Grades 3-5	3.44	3.43	0.01	0.04	0.01	0.849			
Pro-social behavior							-0.05	0.401	
Grades 1-2	3.99	3.92	0.07	0.04	0.07	0.142			
Grades 3-5	3.89	3.87	0.02	0.04	0.02	0.624			
Grades 1-2 (n)									9,701
Grades 3-5 (n)									15,126

(continued)

Exhibit D.15 (continued)

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The analysis sample used in this exhibit is defined as students with all five measures for that year and grade information.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, race, free-/reduced-price lunch status, English language learner status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating group (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit D.16. Estimated Impacts on Student Behavior, by Individualized Education Plan (IEP) Status and Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							0.03	0.229	
Students with IEP status	0.83	0.89	-0.05	0.03	-0.07	0.078			
Students without IEP status	0.55	0.57	-0.02	0.02	-0.03	0.248			
Emotional dysregulation							0.12	0.000 †	
Students with IEP status	1.61	1.73	-0.13 *	0.04	-0.14	0.001			
Students without IEP status	1.10	1.10	-0.01	0.02	-0.01	0.757			
Internalizing behavior							0.00	0.941	
Students with IEP status	1.05	1.06	-0.01	0.03	-0.01	0.815			
Students without IEP status	0.63	0.64	-0.01	0.02	-0.01	0.780			
Attention to schoolwork							0.02	0.623	
Students with IEP status	2.58	2.62	-0.03	0.04	-0.03	0.464			
Students without IEP status	3.56	3.57	-0.01	0.02	-0.01	0.609			
Pro-social behavior							0.04	0.295	
Students with IEP status	3.38	3.41	-0.03	0.04	-0.03	0.500			
Students without IEP status	3.97	3.96	0.01	0.03	0.01	0.685			
Students with IEP status (n)									2,133
Students without IEP status (n)									21,060

(continued)

Exhibit D.16 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 2									
Disruptive behavior							0.02	0.691	
Students with IEP status	0.76	0.81	-0.05	0.04	-0.07	0.200			
Students without IEP status	0.52	0.55	-0.04	0.02	-0.05	0.080			
Emotional dysregulation							0.00	0.978	
Students with IEP status	1.58	1.61	-0.03	0.05	-0.03	0.560			
Students without IEP status	1.07	1.10	-0.03	0.03	-0.04	0.296			
Internalizing behavior							-0.02	0.600	
Students with IEP status	1.00	1.02	-0.02	0.04	-0.03	0.613			
Students without IEP status	0.63	0.67	-0.04	0.03	-0.06	0.112			
Attention to schoolwork							-0.06	0.339	
Students with IEP status	2.68	2.61	0.07	0.07	0.06	0.302			
Students without IEP status	3.54	3.54	0.01	0.03	0.01	0.815			
Pro-social behavior							-0.07	0.168	
Students with IEP status	3.44	3.33	0.11 *	0.06	0.12	0.046			
Students without IEP status	3.99	3.95	0.04	0.03	0.05	0.196			
Students with IEP status (n)									1,623
Students without IEP status (n)									16,244

(continued)

Exhibit D.16 (continued)

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The analysis sample used in this exhibit is defined as students with all five measures for that year and IEP status information from the 2014-2015 school year.

    The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, race, free-/reduced-price lunch status, English language learner status, IEP status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

    The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating group (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

    The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

    A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

    An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

English language learner subgroup: Even though the program was a behavioral intervention, its implementation was carried out through communications of behavioral expectations. Therefore, it was of interest to see if a student's language background affected the program's potential effect. The study checked the program's effect on English language learners (ELLs) and non-ELLs and found that the program effects did not differ by these two groups of students and there was no statistically significant impact on either of these two subgroups (Exhibit D.17).

VI. Program Effects on Student Behavior for Random Assignment Block Level Subgroups

As shown in Appendix B (for example, Exhibits B.2), one key feature of the current study sample is that it has a higher proportion of Hispanic students than the national average or other similar studies of MTSS-B.[43] To explore whether this difference contributed to the impact findings, the study team intended to explore whether the program impacts vary by students' ethnicity identification. However, it was not feasible to carry out such an analysis at the student level. This is because the proportions of Hispanic students were close to 100 percent in some of the study schools. If subgroups were to be defined at the student level, certain schools would drop out of the non-Hispanic subgroup.

To preserve all schools in the sample for the subgroup analysis while still exploring the impact variation by student ethnicity, the team used aggregated-level student ethnicity information to create relevant subgroups. Specifically, the team calculated the average proportion of Hispanic students by random assignment blocks,[44] ranked the 15 blocks by this average proportion, and then defined the seven blocks with the highest proportion as the subgroup with a high concentration of Hispanic students, and the remaining eight blocks as the subgroup with a low proportion of Hispanic students. Exhibit D.18 shows that the program effects on student behavior did not differ between places with a high or low proportion of Hispanic students. This finding suggests that the program tested in this study could be applicable across a broad range of settings, including schools with varying proportions of Hispanic students.

Exhibit D.17. Estimated Impacts on Student Behavior, by English Language Learner (ELL) Status and Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							-0.02	0.495	
English language learners	0.51	0.51	0.00	0.03	-0.01	0.890			
Non-English language learners	0.60	0.62	-0.02	0.02	-0.03	0.250			
Emotional dysregulation							0.03	0.324	
English language learners	1.07	1.11	-0.04	0.03	-0.04	0.253			
Non-English language learners	1.18	1.18	-0.01	0.02	-0.01	0.740			
Internalizing behavior							0.01	0.808	
English language learners	0.67	0.68	-0.01	0.03	-0.02	0.666			
Non-English language learners	0.68	0.68	-0.01	0.02	-0.01	0.765			
Attention to schoolwork							0.04	0.352	
English language learners	3.42	3.47	-0.04	0.04	-0.04	0.265			
Non-English language learners	3.51	3.52	-0.01	0.02	-0.01	0.761			

(continued)

Exhibit D.17 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Pro-social behavior							-0.01	0.879	
English language learners	3.93	3.91	0.01	0.04	0.01	0.719			
Non-English language learners	3.93	3.92	0.01	0.03	0.01	0.756			
English language learners (n)									9,292
Non-English language learners (n)									13,900
Program Year 2									
Disruptive behavior							-0.03	0.296	
English language learners	0.50	0.50	0.00	0.03	0.00	0.962			
Non-English language learners	0.55	0.58	-0.03	0.02	-0.05	0.128			
Emotional dysregulation							-0.05	0.238	
English language learners	1.07	1.06	0.01	0.04	0.01	0.813			
Non-English language learners	1.15	1.19	-0.03	0.03	-0.04	0.270			
Internalizing behavior							0.00	0.994	
English language learners	0.67	0.71	-0.04	0.04	-0.06	0.268			
Non-English language learners	0.67	0.72	-0.04	0.03	-0.06	0.147			

(continued)

Exhibit D.17 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Attention to schoolwork							0.06	0.268	
English language learners	3.40	3.45	-0.04	0.05	-0.04	0.421			
Non-English language learners	3.48	3.47	0.02	0.04	0.01	0.671			
Pro-social behavior							0.02	0.701	
English language learners	3.92	3.89	0.02	0.05	0.02	0.622			
Non-English language learners	3.94	3.90	0.04	0.04	0.04	0.267			
English language learners (n)									7,405
Non-English language learners (n)									10,460

(continued)

Exhibit D.17 (continued)

SOURCES: Teacher Survey of Student Behavior data, collected in the fall of 2015, spring of 2016, and spring of 2017. Student records data from the 2014-2015, 2015-2016, and 2016-2017 school years.

NOTES: The sample includes 58 participating schools and 31 non-participating schools. The analysis sample used in this exhibit is defined as students with all five measures for that year and English language learner status from the 2014-2015 school year.

The estimated impacts are regression-adjusted using three-level hierarchical linear models to account for the nested structure of the data, with students nested within teachers, and teachers nested within schools. The models control for the blocking of random assignment and for baseline differences between students in the participating and non-participating schools with respect to the following variables: grade, age, race, free-/reduced-price lunch status, ELL status, Individualized Education Plan status, baseline standardized math and reading test z-scores, and student baseline behavior measures. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

The values in the column labeled "Participating Schools" are the weighted average of the observed district means for students from the participating group (using number of participating schools in each district as weight). The non-participating schools' values are calculated by subtracting the estimated impacts from the participating school averages. Rounding may cause slight discrepancies in calculating sums and differences.

The estimated impacts' effect sizes are calculated as a proportion of the standard deviation of the full non-participating school members in the analysis sample.

A two-tailed t-test was applied to each estimated impact. Statistical significance is indicated by (*) when the p-value is less than 5 percent.

An F-test was applied to the difference in the estimated impacts between the two subgroups. Statistical significance is indicated by (†) when the p-value is less than 5 percent.

Exhibit D.18. Estimated Impacts on Student Behavior, by Subgroups Defined by Block-Level Hispanic Concentration and Program Year

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Program Year 1									
Disruptive behavior							0.02	0.572	
Blocks with a high concentration of Hispanic students	0.58	0.62	-0.04	0.03	-0.05	0.119			
Blocks with a low concentration of Hispanic students	0.59	0.61	-0.02	0.02	-0.03	0.363			
Emotional dysregulation							-0.03	0.456	
Blocks with a high concentration of Hispanic students	1.09	1.10	0.00	0.03	-0.01	0.890			
Blocks with a low concentration of Hispanic students	1.21	1.25	-0.04	0.03	-0.04	0.196			
Internalizing behavior							0.03	0.482	
Blocks with a high concentration of Hispanic students	0.61	0.63	-0.02	0.03	-0.02	0.576			
Blocks with a low concentration of Hispanic students	0.74	0.73	0.01	0.03	0.01	0.667			
Attention to schoolwork							-0.03	0.548	
Blocks with a high concentration of Hispanic students	3.42	3.41	0.01	0.03	0.00	0.864			
Blocks with a low concentration of Hispanic students	3.48	3.50	-0.02	0.03	-0.02	0.471			

(continued)

Exhibit D.18 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Pro-social behavior							-0.04	0.497	
Blocks with a high concentration of Hispanic students	3.89	3.87	0.02	0.04	0.02	0.634			
Blocks with a low concentration of Hispanic students	3.91	3.93	-0.02	0.04	-0.02	0.627			
Blocks with a high concentration of Hispanic students (n)									14,530
Blocks with a low concentration of Hispanic students (n)									10,861
Program Year 2									
Disruptive behavior							0.01	0.772	
Blocks with a high concentration of Hispanic students	0.54	0.56	-0.02	0.03	-0.02	0.544			
Blocks with a low concentration of Hispanic students	0.58	0.58	-0.01	0.03	-0.01	0.798			
Emotional dysregulation							-0.03	0.592	
Blocks with a high concentration of Hispanic students	1.06	1.06	0.00	0.04	0.00	0.986			
Blocks with a low concentration of Hispanic students	1.22	1.25	-0.03	0.04	-0.03	0.426			

(continued)

Exhibit D.18 (continued)

Measure (0- to 5-point scale)	Participating Schools	Non-Participating Schools	Estimated Impacts	Standard Error of Estimated Impacts	Estimated Impacts in Effect Size	p-value of Estimated Impacts	Estimated Difference of Subgroup Impacts	p-value of Estimated Difference of Subgroup Impacts	Number of Observations
Internalizing behavior							-0.03	0.595	
Blocks with a high concentration of Hispanic students	0.59	0.60	-0.01	0.04	-0.02	0.746			
Blocks with a low concentration of Hispanic students	0.76	0.80	-0.04	0.03	-0.05	0.238			
Attention to schoolwork							0.01	0.925	
Blocks with a high concentration of Hispanic students	3.44	3.42	0.01	0.05	0.01	0.788			
Blocks with a low concentration of Hispanic students	3.46	3.45	0.02	0.04	0.02	0.651			
Pro-social behavior							0.03	0.670	
Blocks with a high concentration of Hispanic students	3.93	3.91	0.02	0.05	0.02	0.641			
Blocks with a low concentration of Hispanic students	3.93	3.88	0.05	0.04	0.05	0.242			
Blocks with a high concentration of Hispanic students (n)									13,838
Blocks with a low concentration of Hispanic students (n)									11,004

Exhibit D.18 (continued)

SOURCES: Teacher Ratings of Student Behavior Survey from the fall of 2015, spring of 2016, and spring of 2017. District records data collected for the 2014-2015 and 2016-2017 school years.

NOTES: The sample includes all eligible students in Grades 1-5 in the study schools in spring 2017 who have teacher ratings for all five behavior measures. A student is initially identified as struggling with behavior if his/her initial (fall 2015) disruptive behavior rating is in the highest 15 percent among students in the same grade across all study schools. The Program Year 2 behavior analysis sample has 1,847 such students. The number of observations varies by baseline characteristics due to missing values.

    The estimated differences are regression-adjusted using hierarchical linear models to account for the nested structure of the data with students nested within teachers and teachers nested within schools. The models control for indicators of random assignment blocks. The values for the participating schools are the weighted average of the observed district means for the participating schools, using the number of participating schools in each district as weight. The values for the non-participating schools are the differences between the participating school means and the estimated differences. Rounding may cause slight discrepancies in calculating sums and differences.

    *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

    [a]State standardized test scores were only available for students who were in Grade 3 in the baseline year (2014-2015).

ENDNOTES

[1]Sugai and Horner, 2020.

[2]A similar approach was followed by Bradshaw and others in prior studies. See Bradshaw, Waasdorp, and Leaf (2015) for an example.

[3]Crone, Hawken, and Horner (2010); Hawken et al. (2015).

[4]See the Center for Positive Behavioral Interventions and Supports (https://www.pbis.org/resource/check-in-check-out-a-targeted-intervention).

[5]See Center for Social Behavior Support (http://www.csbspartnership.org/mtssb-materials) and Midwest PBIS (http://www.mid-westpbis.org/home).

[6]Lindstrom Johnson et al. (2020).

[7]Rusby, Crowley, Sprague, and Biglan (2011); Rusby, Taylor, and Milchak (2001).

[8]Abry, Cash, and Bradshaw (2014).

[9]Horner et al. (2004).

[10]Anderson et al. (2012).

[11]The lower consent rate in the non-participating schools was driven by a lower proportion of families returning the consent form and a lower consent rate among those that returned the form as compared with the participating schools. However, it is not clear what might have caused the lower return and consent rates for this group. Among other things, it might be the case that teachers in the non-participating schools did not emphasize the consent process as much as teachers in the participating schools.

[12]Koth, Bradshaw, and Leaf (2009); Bradshaw and Kush (2020).

[13]Bradshaw, Waasdorp, and Leaf (2012).

[14]Mason, Gunersel, and Ney (2014).

[15]Rusby, Crowley, Sprague, and Biglan (2011); Rusby, Taylor, and Milchak (2001).

[16]Abry, Cash, and Bradshaw (2014).

[17]Bradshaw et al. (2018); Debnam, Pas, and Bradshaw (2012).

[18]Bradshaw, Koth, Thornton, and Leaf (2009); Hoy and Tarter (1997).

[19]Bear, Gaskins, Blank, and Chen (2011); Bear, Yang, Pell, and Gaskins (2014).

[20]National Center on Safe Supportive Learning Environment (pulled in 2016 from https://safesupportivelearning.ed.gov/edscls).

[21]Horner et al. (2004).

[22]Bradshaw, Waasdorp, and Leaf (2012).

[23]Debnam, Pas, and Bradshaw (2012).

[24]Bradshaw, Waasdorp, and Leaf (2015).

[25]Horner and Sugai (2015).

[26]See What Works Clearinghouse Standards Handbook, Version 4.1, Table II.6.

[27]Puma, Olsen, Bell, and Price (2009).

[28]Schochet (2010).

[29]What Works Clearinghouse Standards Handbook, Version 4.1, page 15.

[30]All ratings received before December 13, 2015, or 28 days since the start of the TSSB fielding, are included in the early subgroup, and all ratings received after that date are in the late subgroup.

[31]Even though the estimated growth rates are statistically significant for the disruptive behavior and internalizing behavior ratings, the magnitude of such growth is very small and not substantively significant.

[32]What Works Clearinghouse Standards Handbook, Version 4.1.

[33]Bloom, Unterman, Zhu, and Reardon (2020).

[34]Program effects on these outcomes were estimated using linear probability regression models.

[35]Pas, Bradshaw, and Mitchell (2011).

[36]Because schools within each district were randomly assigned to participate in the MTSS-B training and support or continue with their existing practices, the study did not emphasize some of the typical elements of MTSS-B such as strong district support and involvement and outreach to stakeholders external to the school. The higher level of district support observed for participating schools was perhaps the result of study-funded support for staff training on MTSS-B and a district level coach for participating schools.

[37]Lee and Gage (2020).

[38]Bradshaw et al. (2012); Waasdorp, Bradshaw, and Leaf (2012).

[39]Bradshaw et al. (2012).

[40]Horner and Sugai (2015).

[41]Grossman and Grossman (1994); Bertrand and Pan (2013).

[42]Chall and Jacobs (2003).

[43]For example, Bradshaw et al. (2012) reported that 3.9 percent of the study sample were Hispanic students.

[44]Specifically, the team calculated the average proportion of Hispanic students by random assignment blocks. School-level data from the baseline school year (2014-2015) from the Common Core Data (CCD) were used for this calculation.

REFERENCES

Abry, Tashia, Anne H. Cash, and Catherine P. Bradshaw. 2014. "Using Generalizability Theory to Examine Sources of Variance in Observed Behaviors Within High School Classrooms." *Society for Research on Educational Effectiveness*.

Algozzine, B., S. Barrett, L. Eber, H. George, R. Horner, T. Lewis, B. Putnam, J. Swain-Bradway, K. McIntosh, and G. Sugai. 2014. "School-Wide PBIS Tiered Fidelity Inventory." OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports.

Algozzine, B., C. Wang, R. White, N. Cooke, and M.B. Marr. 2012. "Effects of multi-tiered academic and behavioral instruction on difficult-to-teach students." *Exceptional Children*, 79, 45-64. https://doi.org/10.1177/001440291207900103(TFI).pdf.

Anderson, C.M., T. Lewis-Palmer, A.W. Todd, R.H. Horner, G. Sugai, and N.K. Sampson. 2012. "Individual Student Systems Evaluation Tool (Version 3.0)." Eugene: University of Oregon, Educational and Community Supports.

Bear, George G., Chunyan Yang, Megan Pell, and Clare J. Gaskins. 2014. "Validation of a Brief Measure of Teachers' Perceptions of School Climate: Relations to Student Achievement and Suspensions." *Learning Environments Research* 17, 3: 339-354.

Bear, George G., Clare Gaskins, Jessica Blank, and Fang Fang Chen. 2011. "Delaware School Climate Survey–Student: Its Factor Structure, Concurrent Validity, and Reliability." *Journal of School Psychology* 49, 2: 157-174.

Bertrand, Marianne, and Jessica Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Economic Journal: Applied Economics* 5, 1: 32-64.

Bloom, Howard S., Rebecca Unterman, Pei Zhu, and Sean F. Reardon. 2020. "Lessons from New York City's Small Schools of Choice about High School Features That Promote Graduation for Disadvantaged Students." *Journal of Policy Analysis and Management* 39, 3: 740-771.

Bradshaw, Catherine P., Christine W. Koth, Leslie A. Thornton, and Philip J. Leaf. 2009. "Altering School Climate Through School-Wide Positive Behavioral Interventions and Supports: Findings from a Group-Randomized Effectiveness Trial." *Prevention Science* 10, 2: 100-115.

Bradshaw, Catherine P., Elise T. Pas, Jessika H. Bottiani, Katrina J. Debnam, Wendy M. Reinke, Keith C. Herman, and Michael S.J. Rosenberg. 2018. "Promoting Cultural Responsivity and Student Engagement Through Double Check Coaching of Classroom Teachers: An Efficacy Study." *School Psychology Review* 47, 2: 118-134.

Bradshaw, C.P., E. Pas, J. Bloom, S. Barrett, P. Hershfeldt, A. Alexander, M. McKenna, A.E. Chafin, and P. Leaf. (2012). "A state-wide collaboration 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. doi:10.1007/s10488-011- 0384-6 ECONOMIC BENEFITS OF PBIS 11.

Bradshaw, Catherine P., and Joseph M. Kush. 2020. "Teacher Observation of Classroom Adaptation-Checklist: Measuring Children's Social, Emotional, and Behavioral Functioning." *Children & Schools* 42, 1: 29-40.

Bradshaw, C.P., M.M. Mitchell, and P.J. Leaf. 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(3), 133-148. doi:10.1177.1098300709334798.

Bradshaw, Catherine P., Tracy E. Waasdorp, and Philip J. Leaf. 2012. "Effects of School-Wide Positive Behavioral Interventions and Supports on Child Behavior Problems." *Pediatrics* 130, 5: e1136-e1145.

Bradshaw, Catherine P., Tracy E. Waasdorp, and Philip J. Leaf. 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.

Center for Positive Behavioral Interventions and Supports. Website: www.pbis.org/resource/check-in-check-out-a-targeted-intervention.

Center for Social Behavior Support. 2015. Website: www.csbspartnership.org/home.

Chall, Jeanne S., and Vicki A. Jacobs. 2003. "The Classic Study on Poor Children's Fourth-Grade Slump." *American Educator* 27, 1: 14-15.

Crone, Deanne A., Leanne S. Hawken, and Robert H. Horner. 2010. *Responding to Problem Behavior in Schools: The Behavior Education Program*. Guilford Press.

Debnam, Katrina J., Elise T. Pas, and Catherine P. Bradshaw. 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*: 142-152.

Grossman, Herbert, and Suzanne H. Grossman. 1994. *Gender Issues in Education*: ERIC.

Hawken, Leanne S., Kaitlin Bundock, Courtenay A. Barrett, Lucille Eber, Kimberli Breen, and Danielle Phillips. 2015. "Large-Scale Implementation of Check-In, Check-Out: A Descriptive Study." *Canadian Journal of School Psychology* 30, 4: 304-319.

Horner, Robert H., Anne W. Todd, Teri Lewis-Palmer, Larry K. Irvin, George Sugai, and Joseph B. Boland. 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.

Horner, Robert H., and George Sugai. 2015. "School-Wide PBIS: An Example of Applied Behavior Analysis Implemented at a Scale of Social Importance." *Behavior Analysis in Practice* 8, 1: 80-85.

Horner, R.H., G. Sugai, K. Smolkowski, L. Eber, J. Nakasato, A.W. Todd, and J. Esperanza. 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.

Hoy, Wayne K., and Clemens John Tarter. 1997. *The Road to Open and Healthy Schools: A Handbook for Change*. Corwin.

Koth, Christine W., Catherine P. Bradshaw, and Philip J. Leaf. 2009. "Teacher Observation of Classroom Adaptation–Checklist: Development and Factor Structure." *Measurement and Evaluation in Counseling and Development* 42, 1: 15-30.

Lee, Ahhyun, and Nicholas A. Gage, 2020. "Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports." *Psychol Schs*. 2020; 57: 783–804.

Lindstrom Johnson, Sarah, Y. Natalia Alfonso, Elise T. Pas, Katrina J. Debnam, and Catherine P.J. Bradshaw. 2020. "Scaling Up Positive Behavioral Interventions and Supports: Costs and Their Distribution Across State, Districts, and Schools." *School Psychology Review* 49, 4: 399-414.

Mason, B.A., A.B. Gunersel, and E.A. Ney. 2014. "Cultural and Ethnic Bias in Teacher Ratings of Behavior: A Criterion-Focused Review." *Psychology in the Schools*, 51: 1017-1030. https://doi.org/10.1002/pits.21800.

Midwest PBIS Network. n.d. "Curriculum & Materials." Website: http://www.midwestpbis.org/home.

National Center on Safe Supportive Learning Environments. 2021. Website: https://safesupportivelearning.ed.gov/edscls.

Pas, Elise T., Catherine P. Bradshaw, and Mary M. Mitchell. 2011. "Examining the Validity of Office Discipline Referrals as an Indicator of Student Behavior Problems." *Psychology in the Schools* 48, 6: 541-555.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price 2009. *What to Do When Data Are Missing in Group Randomized Controlled Trials. NCEE 2009-0049*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Rusby, Julie C., Ryann Crowley, Jeffrey Sprague, and Anthony Biglan. 2011. "Observations of the Middle School Environment: The Context for Student Behavior Beyond the Classroom." *Psychology in the Schools* 48, 4: 400-415.

Rusby, J.C., T. Taylor, and C.J. Milchak. 2001. "Assessing school settings: Interactions of Students and Teachers (ASSIST) Observation System." Unpublished manuscript. Eugene, OR: Oregon Research Institute.

Schochet, Peter. 2010. "The Late Pretest Problem in Randomized Control Trials of Education Interventions." *Journal of Educational and Behavioral Statistics*, 35(4), 379-406.

Sugai, George, and Robert H. Horner. 2020. "Sustaining and Scaling Positive Behavioral Interventions and Supports: Implementation Drivers, Outcomes, and Considerations." *Exceptional Children*: 120-136.

Waasdorp, T.E., C.P. Bradshaw and P.J. Leaf. 2012. "The impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial." *Archives of Pediatrics & Adolescent Medicine*, 166(2), 149- 156. doi:10.1001/archpediatrics.2011.755.

What Works Clearinghouse. 2020. *What Works Clearinghouse Standards Handbook, Version 4.1*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.