Journal of Educational Psychology

Can Effective Classroom Behavior Management Increase Student Achievement in Middle School? Findings From a Group Randomized Trial

Keith C. Herman, Wendy M. Reinke, Nianbo Dong, and Catherine P. Bradshaw

Online First Publication, November 5, 2020. http://dx.doi.org/10.1037/edu0000641

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

http://dx.doi.org/10.1037/edu0000641

Can Effective Classroom Behavior Management Increase Student Achievement in Middle School? Findings From a Group Randomized Trial

Keith C. Herman and Wendy M. Reinke
University of Missouri

Nianbo Dong
University of North Carolina at Chapel Hill

Catherine P. Bradshaw
University of Virginia

This cluster randomized controlled trial evaluated the efficacy of the CHAMPS classroom management program on the social behavioral and academic outcomes of a large diverse sample of middle school students within an urban context. Participants included 102 teachers and 1,450 students in sixth to eighth grade. Two-level hierarchical linear models (HLM) were conducted to examine the overall treatment effects on student behavior and academic outcomes. In addition, mediation analyses examined a hypothesized putative mechanism for observed academic outcomes. Findings indicated that CHAMPS improved teacher ratings of student concentration problems ($d = -0.18$) and classwork completion ($d = 0.18$), observed student time-on-task ($d = 0.16$), and student scores on broad English ($d = 0.14$), and math problem solving ($d = 0.17$) academic achievement tests. Null effects were observed for student prosocial and disruptive behaviors and self-regulation skills as well as reading comprehension and broad math achievement performance. Main effects on the English achievement test scores were partially mediated by student improvements in observed time-on-task. Practical significance of the findings and implications for schools and policymakers are discussed.

Educational Impact and Implications Statement
A teacher classroom behavior management training program, CHAMPS, caused improvements in teacher classroom management practices and student social and academic outcomes. The improvements in academic achievement were, in part, explained by increases in student time-on-task. Improving teacher classroom management training holds promise for increasing student achievement on a large scale.

Keywords: classroom management, teacher training, prevention, middle school, academic achievement

Effective teaching requires a combination of good instruction and classroom behavior management (Lekwa, Reddy, & Shernoff, 2019). Even in the presence of high quality instructional practices, students will not learn if they are not paying attention or are distracted by disruptive behaviors of other students (Gage, MacSuga-Gage, & Crews, 2017).

Keith C. Herman and Wendy M. Reinke, Missouri Prevention Science Institute, University of Missouri; Nianbo Dong, School of Education, University of North Carolina at Chapel Hill; Catherine P. Bradshaw, College of Education, University of Virginia.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130143 to the University of Missouri (PI: Keith C. Herman). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Keith C. Herman, Missouri Prevention Science Institute, University of Missouri, 16 Hill Hall, Columbia, MO 65201. E-mail: hermanke@missouri.edu

Classroom management is defined as "a collection of noninstructional classroom procedures implemented by teachers in classroom settings with all students for the purposes of teaching pro-social behavior and preventing and reducing inappropriate behavior" (Oliver, Wehby, & Reschly, 2011, pp. 7–8). Considerable research has demonstrated that effective classroom management can reduce disruptive and aggressive behavior (Oliver et al., 2011; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). In classic research on effective classroom management, Brophy and Evertson (1976) and Anderson, Evertson, and Brophy (1979) found that effective management techniques were associated with increased student learning. More recently, using a cross-sectional and longitudinal design, Kunter, Baumert, and Köller (2007) found that effective behavior management strategies also promoted student interest in learning.

A large body of literature has also documented the critical features of effective classroom management that can benefit all students. A seminal review identified 20 distinct classroom behavior management practices that met What Works Clearinghouse criteria for evidence-based: (a) maximize structure and predictabil-

ity; (b) post, teach, review, and provide feedback on expectations; (c) actively engage students in observable ways; (d) use a continuum of strategies to acknowledge appropriate behavior; and (e) use a continuum of strategies to respond to inappropriate behavior (Simonsen et al., 2008). In a separate Campbell Review, Oliver, Wehby, and Reschly (2011) expanded on these findings by conducting a systematic meta-analyses of class-wide classroom management packages, those delivered to all students, rather than focusing on discrete strategies delivered to improve individual student behaviors. Class-wide packages such as the Good Behavior Game (GBG; Barrish, Saunders, & Wolf, 1969) and Classroom Organization and Management Program (Evertson, 1989) combine many of the evidence-based strategies into teacher training programs. Oliver et al. (2011) concluded that there is compelling evidence that class-wide classroom management programs significantly reduced student disruptive and aggressive behaviors over time.

Despite the large body of evidence regarding effective classroom management practices, teachers commonly report challenges managing student behaviors (Buell, Hallam, Gamel-Mccormick, & Scheer, 1999; Pavri, 2004). In fact, teachers identify classroom management to be one of the most difficult parts their job, in part because they a receive limited amount of training in this area (Barrett & Davis, 1993; Ingersoll, 2002; Reinke, Stormont, Herman, Puri, & Goel, 2011).

Gaps in the Literature

Classroom Management Effects on Student Achievement

Although it is clear that effective classroom management improves student behavior outcomes, it is less clear whether it also improves student learning. Typically, evaluations of classroom management programs and practices focus on proximal effects such as student behaviors (Oliver et al., 2011). Although many studies have shown links between classroom management practices and student achievement, much of this literature is based on correlational findings. For instance, the classic Anderson and colleagues (1979) study, that in many ways guided the development of effective classroom management programs, found longitudinal associations between specific classroom management practices and reading achievement, but found few significant differences between teachers trained versus untrained in these practices.

Similarly, although the developers of classroom management programs hypothesize academic effects, the literature on the topic is relatively sparse and inconsistent. For instance, the GBG, one of the most widely disseminated classroom management programs, has had two randomized controlled trials (RCTs) that evaluated academic outcomes with one reporting no academic benefit (Dolan et al., 1993) and the other suggesting impact on academic performance (Ialongo et al., 1999). However, the study demonstrating academic benefits was paired with training in effective instructional practices, so the unique effect of GBG on academic performance could not be isolated. A recent RCT of the Incredible Years Teacher Classroom Management (IY TCM) found that teachers in the training condition improved their classroom management practices and increased student prosocial behavior and self-regulation skills (Reinke, Herman, & Dong, 2018). Although there was no main effect on student academic achievement, baseline academic competence moderated IY TCM effects on academic outcomes; specifically, students in IY TCM classrooms with low initial academic competence had significant improvements in their academic skills relative to like peers in wait-list control classrooms.

Effective Classroom Management Practices and Programs in Middle School

An additional limitation of existing literature is that most studies to date have focused on effective classroom management strategies and programs in elementary school settings. Few training programs have been developed explicitly for middle school teachers. In their search of reviews by What Works Clearinghouse (2014) and the Johns Hopkins Best Evidence Encyclopedia (2019), Allen, Pianta, Gregory, Mikami, and Lun (2011) found only two professional development programs for middle school teachers that had any evidence of impacting achievement, and both of these were specific to math education.

Middle school represents an important developmental period for students and presents unique challenges and opportunities for classroom teachers. For instance, deviant and disruptive behaviors peak during this period as students attempt to navigate this new social field where peer influences become particularly salient (Merikangas et al., 2010). Yet, middle school students are only beginning to have fully crystallized cognitive coping skills to independently solve social challenges (Cicchetti & Toth, 1995). Evidence also suggests that middle school represents a time of dramatic decline in academic performance and engagement (Marks, 2000; Pianta & Allen, 2008).

Filling the void of training programs to support teacher skills during this critical developmental stage, Allen et al. (2011) described the evaluation of My Teacher Partner-Secondary (MTP-S), a teacher consultation model, in middle school classrooms. MTP-S is based on the teaching through interactions theoretical framework, which infuses attachment, self-determination, and cognitive theories to explain how teachers influence student development. In particular, MTP-S focuses on helping teachers support student autonomy and relatedness while providing engaging and stimulating academic instruction. As a professional development model, MTP-S is unique in that teachers submit videos of their instruction to coaches who provide ongoing feedback consistent with the program's guiding theory.

In two randomized trials, Allen et al. (2015, 2011) found MTP yielded significant improvements in student achievement. However, these effects did not emerge until one year after the teachers received training in the model. The authors attributed the delay to the time it took for teachers to fully acquire skills that they learned the prior year of consultation in MTP and the full year of exposure their students had to these skills in the subsequent school year. It is important to note that MTP targets instructional skills in addition to relational qualities and that these instructional improvements were not considered in these analyses. Thus, the study did not rule out the possibility that improvements in youth outcomes were principally driven by improvements in teacher instructional skills.

More recently, class-wide function-related intervention teams (CW-FIT), a universal approach to classroom management origi-

nally developed for elementary schools, was modified and evaluated in middle school settings (Monson, Caldarella, Anderson, & Wills, 2020; Orr et al., 2020; Speight, Whitby, & Kucharczyk, 2020; Wills, Caldarella, Mason, Lappin, & Anderson, 2019). Several studies in various middle school settings suggest that CW-FIT improved student behaviors and increased their time-on-task. However, all of these studies used single case designs (ABAB or multiple baseline) not RCTs, and none included student achievement as a primary outcome.

CHAMPS

Another promising teacher training program, CHAMPS, was developed for implementation in middle schools over a decade ago. CHAMPS is an acronym for the six dimensions that the program focuses on for assisting teachers to define expectations in every setting: Conversation, Help, Activity, Movement, Participation, and Signal. CHAMPS is a principle-driven and modular series of training and coaching materials designed to help classroom teachers develop an effective classroom management plan that is proactive, positive, and behaviorally instructional (Sprick, Garrison, & Howard, 2009). CHAMPS is grounded in social learning and behavioral principles, including the research-based principles of effective classroom management. In particular, CHAMPS assumes all behavior is learned and can be taught and altered by the social environment. The acronym STOIC defines the five key principles that guide effective classroom management: Structure classroom, Teach expectations, Observe and supervise, Interact positively, and Correct fluently. Note that these five principles overlap almost completely with the five evidence-based categories of effective classroom management identified by Simonsen and colleagues (2008). Emphasis is placed on helping teachers structure their classrooms in ways that prompt responsible and engaged student behaviors and prevent off-task disruptive behaviors. The program prepares teachers to explicitly teach students how to behave responsibly in every classroom situation consistent with the robust literature on defining and teaching expectations and rules (Alter & Haydon, 2017). Additionally, CHAMPS helps teachers focus more time, attention, and energy on acknowledging responsible behavior than on correcting misbehavior and preplan responses to misbehavior so they will be brief, calm, and consistent. One key indicator of proactive classroom management skills in the model is the ratio of positive to negative interactions between the teacher and each student. Specific program modules and materials have been developed for working with the unique challenges of middle school settings.

CHAMPS is a fully developed and widely disseminated prevention program. It is delivered as large group training workshops followed by in-person coaching and feedback. Books, planning materials, and DVD's to support implementation of this program in precise and repeatable ways are provided (Sprick, 2010; Sprick & Howard, 1996; Sprick et al., 2009). Over 100,000 classroom teachers have been trained in the CHAMPS model (R. Sprick, personal communication, June 1, 2017).

Time-on-Task as a Mediator of CHAMPS Impact on Student Achievement

CHAMPS theory of change is rooted in behavioral and social learning theories. The STOIC principles that guide teacher classroom management behaviors encompass the aspects of effective environments for supporting student behavior change (Simonsen et al., 2008). In particular, teachers provide structured environments that minimize distractions and opportunities for misbehavior. By providing students with clear expectations in all settings and increasing rates of positive attention (social reinforcement) for meeting these expectations relative to negative attention for not meeting them, teachers reduce off-task disruptive behaviors and increase on-task, engagement in learning (Gettinger & Seibert, 2002; Partin et al., 2009; Stronge, Ward, & Grant, 2011).

In other words, effective classroom structure and management promote student engagement and increased time-on-task (see Gettinger & Walter, 2012; Stronge et al., 2011). Time has been conceptualized as a critical educational resource for increasing student academic attainment (Carroll, 1989). Academic learning is comprised of several elements, including total amount of time spent in instructional activities and the amount of time students spend attending to instruction (Reeve, Jang, Carrell, Jeon, & Barch, 2004; Spanjers, Burns, & Wagner, 2008). Although overall instructional time is important, engaged time, or instructional time during which students are paying attention, has a stronger relationship with student achievement (Gromada & Shewbridge, 2016). Thus, one key mechanism by which CHAMPS may improve student academic outcomes is by reducing off-task disruptive behaviors so that students spend more time in engaged, on-task learning.

Accordingly, CHAMPS is intended to promote student achievement in the following manner: CHAMPS implementation improves the teacher's ability to (a) structure classrooms and prevent problem behaviors; (b) reduce disruptive, off-task behaviors to allow more time for instruction; and (c) deliver higher rates of attention for socially appropriate and on-task behavior which increases individual student attention/concentration and access to instruction and learning time. Increases in student on-task behavior and learning time leads to improvements in their academic achievement.

Rationale for the Current Study

The U.S. Department of Education has prioritized evaluating educational programs and practices that are widely disseminated yet without rigorous empirical evaluation (Institute of Education Sciences, 2020, p. 48). Despite its popularity and wide-scale adoption, CHAMPS has never been the subject of a rigorous RCT. None of the CHAMPS studies completed to date meet the highest standards of evidence established by review groups, including What Works Clearinghouse (Institute of Education Sciences, 2017; e.g., no randomized trials or rigorous single case studies have evaluated it).

Moreover, although classroom management programs are ubiquitous in elementary school settings, very few programs exist to support middle school teachers. If found to be effective CHAMPS could help fill that void. Finally, although abundant research supports the impact of classroom behavior management programs in improving student behaviors much less is known about whether these programs and practices, in themselves, improve student learning.

The current study moves beyond prior studies by evaluating the specific impact of the CHAMPS program on the development of

effective classroom behavior management practices in middle school teachers from a large urban district. In particular, the trial was the first to determine the unique effects of the program on student behaviors and academic outcomes during the middle school years, sixth through eighth grade. A teacher-focused classwide intervention such as CHAMPS would be cost-effective and easily disseminated if it is shown to have significant effects on academic outcomes.

Research Questions and Hypotheses

Research Question 1: Does CHAMPS improve teacher implementation of effective classroom management skills? Based on previous correlational research suggesting the promise of the program (e.g., Madigan & Cross, 2011), we hypothesized that teachers in classrooms receiving CHAMPS (Madigan, Cross, Smolkowski, & Strycker, 2016) training would demonstrate a significantly larger increase in their implementation of proactive classroom management skills compared with teachers in classrooms that did not receive CHAMPS training.

Research Question 2: Does CHAMPS implementation improve student behavior and academic outcomes? Prior research suggests that effective classroom management practices and programs based on similar strategies employed by CHAMPS significantly reduce student inattention and problem behaviors and improves student self-regulation and prosocial skills (Oliver et al., 2011; Simonsen et al., 2008; Reinke et al., 2018). Thus, we hypothesized that students in the classrooms of teachers who received the CHAMPS intervention would demonstrate reductions in concentration problems, disruptive behaviors, and emotional dysregulation in comparison with students in classrooms of the control group teachers. We also expected students in classrooms of teachers who received the CHAMPS intervention to demonstrate increases in prosocial behavior, time-on-task, and academic achievement in comparison to students in the classroom of the control group teachers. We included student age, grade, sex, race/ethnicity, free and reduced-price lunch, and special education status as covariates in these analyses given the documented association between many sociodemographic characteristics and academic and behavior outcomes.

Research Question 3: Does time-on-task mediate any observed main effects on student academic achievement? Consistent with the theory of change guiding CHAMPS, we expected improvements in time-on-task to mediate intervention effects on academic outcomes.

Method

Participants

Middle school teacher and student participants were recruited from two urban school districts in the Midwest United States. Participants were recruited as part of a group RCT of the CHAMPS behavior management and coaching program. Eligible teacher participants included sixth- to eighth- grade English or math teachers who consented to participate in the project. Parent consent and student assent were obtained for student participants recruited from classrooms of participating teachers.

A final teacher sample of 102 and student sample of 1,450 agreed to participate in the present study. Student participants were 50.8% female and 78.2% African American, 17.8% White, 2.1%, and 4.0% other. The percentage of students in sixth, seventh, and eighth grade was equal to 35.4%, 38.7%, and 25.9%, respectively. Overall, 69.66% of students qualified for free/reduced-priced lunch, and 6% of the sample received special education services. Teacher participants were 79.1% female and 70.9% White, 25.6% African American, and 3.5% other. Teachers' ages ranged from 23 to 63 years ($M = 37.8$, $SD = 8.8$), whereas teaching experience ranged from 1.0 to 23.0 ($M = 10.4$, $SD = 6.3$).

As indicated in the participant flowchart (see Figure 1), the study had high rates of enrollment for eligible teachers (91%) and students (75%). A blocked cluster random assignment design was utilized. Teachers were randomly assigned to receive CHAMPS or to a wait-list, business as usual control group within school, with the constraint that the number of intervention teachers to be no more than one more or less than the number of control teachers. Teacher participants were recruited and randomized across four cohorts—Year 1: 26 teachers (13 interventions), 437 students; Year 2: 36 teachers (18 interventions), 453 students; Year 3: 24 teachers (12 interventions), 337 students; Year 4: 16 teachers (eight interventions), 223 students.

Procedure

The University Institutional Review Board and the participating school district approved the study protocol. Teachers and students were recruited at the beginning of the school year. Data were collected at the beginning of the school year, prior to the intervention, and at the end of the school year, postintervention. All preintervention assessments occurred in mid-September to mid-October. Postintervention assessments were collected in late April and May of the same academic year. Observations were also collected at baseline (Time 1) and three times following intervention: November (Time 2), February (Time 3), and April/May (Time 4).

Intervention condition. In four sequential, annual cohorts of between eight to 18 teachers in the CHAMPS condition attended two full-day group trainings, back-to-back sessions in late October and an additional session in late November/early December. All trainings were facilitated by a certified CHAMPS trainer supervised by the program developer. Additionally, an on-site doctoral-level coach who was trained and supervised by the program developer supported teacher implementation following sessions.

CHAMPS is a comprehensive curriculum for improving teacher classroom management and relationship skills. The CHAMPS model targets teachers' use of effective classroom management strategies by promoting positive relationships with all students and by strengthening the relevance and engagement of instruction. The key principles for an organized and effective classroom are summarized by the acronym STOIC, mentioned previously: Structure classroom, Teach expectations, Observe and supervise, Interact positively, and Correct fluently. The training and subsequent coaching support focuses on building teacher competence in each of these five domains. Training occurs in seven modules: developing a vision, organization, developing and teaching expectations, proactive teaching, student motivation, data-based decisions, and calm and consistent corrections. CHAMPS includes a host of

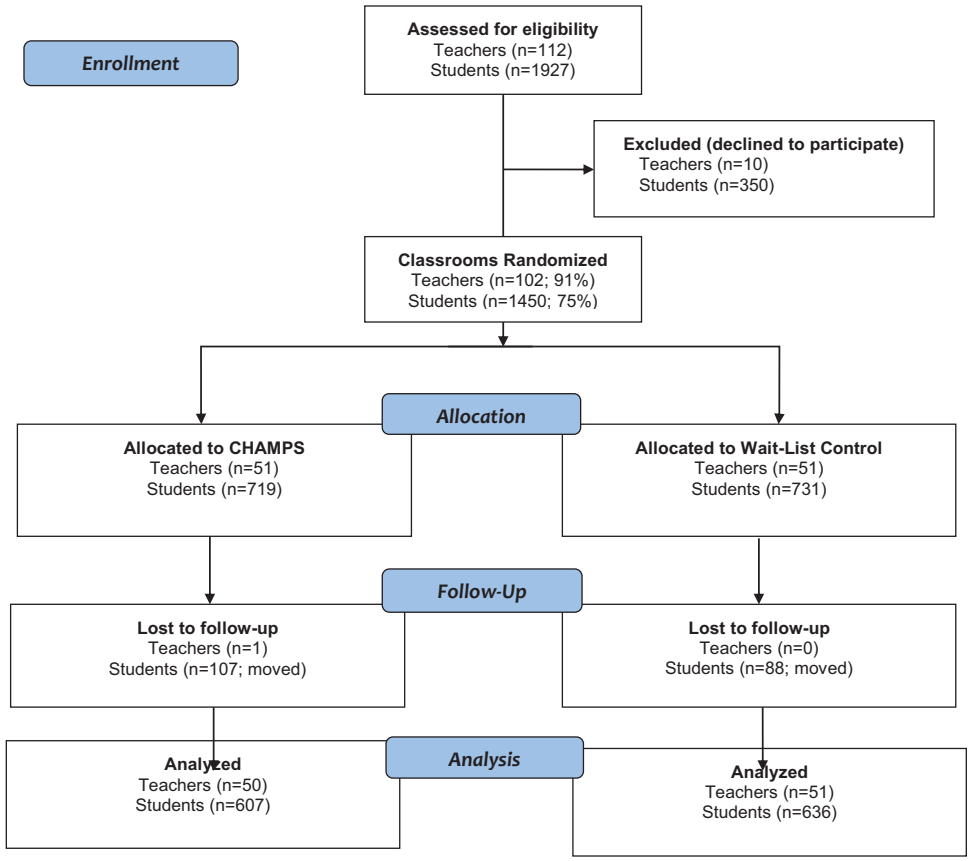

Figure 1. CHAMPS randomization participant flowchart. See the online article for the color version of this figure.

well-developed and user-friendly materials to support teacher implementation of the practices. These include the companion books, *CHAMPS: A Proactive and Positive Approach to Classroom Management* and the *Teacher's Encyclopedia of Behavior Management: 100 Problems/500 Plans*; *CHAMPS Teacher Planners* for keeping on track with the approach; and the *Making Every Second Count* DVD series.

To minimize risk of contamination across conditions, we asked all intervention teachers to sign a confidentiality agreement acknowledging that they would not share intervention materials with other teachers. In all recruitment visits with principals and teachers and in many postassignment interactions with participants, we emphasized the importance of adherence to condition. We also asked teachers if they had shared or acquired information about the intervention over the course of the study, and all indicated that they had not. We based these strategies off of the guidance of others who have conducted classroom management trials (see Kellam et al., 2008) and our own prior studies in which we found no evidence of contamination across control classrooms. It is also important to note that abundant literature suggests that sustained improvements in classroom management skills typically requires ongoing modeling, feedback, and support (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Such skills are not likely to change through minimal contact, such as social conversations.

CHAMPS coaching. In this study, the CHAMPS coach was a doctoral-level special educator. The coaching model is manualized, partnership-oriented, and involves giving teachers ongoing explicit feedback about their implementation (see Sprick, 2010). In between each workshop session, the CHAMPS coach observed the teachers in the classroom and met with them individually for up to 1 hr on a weekly basis. We defined a minimal dose that each teacher needed to receive as a total of four visits with the coach. The first visit focused on establishing rapport and setting goals. The second visit focused on providing the teacher with explicit feedback based on the coach's classroom observations and developing a plan based on the teacher's own goals. Subsequent visits were tailored to each teacher based on this goal setting and planning. The coach recorded any contact with teachers, including brief check-ins, to reviewing strategies and schedule the next meeting. During the individual coaching sessions, the coach reviewed workshop content and supported goal setting for use of strategies, provided feedback on teacher skills and interpersonal teaching processes, modeled effective practice, role-played potential barriers and challenges, and supported action planning. CHAMPS is a universal intervention for teachers, meaning that the intervention is intended for all teachers regardless of skill level. However, the CHAMPS coach differentiated the amount of coaching provided to teachers based on their need for supports. The mean time spent with a teacher by the coach, outside of classroom observations was 147 min (range = 48 to 358 min).

Control condition. Teachers assigned to the wait-list control condition continued their business as usual teaching and profes-

sional development opportunities during the study period. Due to the wait-list design, control condition teachers were offered the CHAMPS intervention immediately after their period of participation in the evaluation component of the project ended. Teachers in both conditions were compensated for their time and effort in completing surveys as part of the project.

Measures of Implementation Fidelity and Teacher Practices

Fidelity to CHAMPS workshop trainings. Adherence and fidelity of implementation to the CHAMPS workshops were monitored over the course of the year, including dose/exposure to training and coaching, teacher ratings of workshop quality, and observation of classroom implementation. Regarding dose/exposure to the training, teachers in the intervention were all exposed to the training workshops; nearly all teachers attended all three workshops (attendance rate was 92–100% for each workshop) and the few teachers who missed a workshop due to illness or other reason met with the CHAMPS coach to review missed material. Teachers rated the workshops with high satisfaction and likelihood of recommending the training to others (mean ratings of 4.80 and 4.87, respectively, on a scale from 1–5 with high scores indicating greater satisfaction). In addition, teachers reported that they expected good results from receiving the training (4.60), that they agreed with the approach to behavior change (4.69), and that they were confident it would be helpful with current (4.33) and future (4.38) behavior problems in their classrooms.

Direct observations. Classroom observations were conducted by independent observers blind to intervention condition. Classroom-level observations including measures of teacher implementation fidelity and adherence were collected at four time points. The first observation occurred in October prior to receiving CHAMPS training or coaching. The second observation in November after teachers received workshop Sessions 1 and 2 and at least one coaching visit. The third observation occurred in February after all three workshops were completed and the minimal dose of coaching delivered. The final observation occurred at the end of the school year (April/May). All observations occurred in classrooms during instructional times. The pre- and postobservations were an aggregation of a series of four 5-min observations made by the same observer on a single classroom visit, whereas the second and third observations were both 20 min in length. Student-level observations were collected on two occasions, at baseline and at the end-of-the-school-year.

Teacher implementation fidelity to CHAMPS. Independent observers conducted direct observations of teacher implementation fidelity using the *STOIC Rating Form* across the four time points described previously (Sprick, 2010). STOIC provides global ratings of each of the five key domains of CHAMPS practices: Structure classroom, Teach expectations, Observe and supervise, Interact positively, and Correct fluently. Independent observers rated each of these five domains on a 0 (*no evidence*) to 4 (*full evidence*) rating scale, and we computed a summary score of these ratings as a measure of adherence. The STOIC was not gathered at baseline for Cohort 1 of the study because the measure was not available at the start of the project, but all other time points were gathered. Analyses examining changes on the STOIC used other similar measures described below to adjust for baseline differences. Prior to data collection, observers attended a two hour training focused on using the STOIC and practiced coding videos of actual classrooms. They were allowed to collect data only after reaching agreement with a master coder. The ICC (one-way random effects absolute agreement [OWREAA]) for STOIC summary scores ranged from .92 to .97 at each measurement time point. We used the OWREAA because it is consistent with our data collection method where all raters did not score all participants or classrooms; ICCs at or above .75 are considered excellent (Ladd, Tomlinson, Myers, & Anderson, 2016).

In addition, we conducted 20-min classroom observations using the *Classroom Assessment Scoring System-Secondary* (CLASS-S; Pianta, Hamre, Hayes, Mintz, & LaParo, 2008) at baseline and across the same direct observation time points as the STOIC. The CLASS-S asks observers to provide global ratings of specific aspects of a classroom's emotional support, organization, and instructional support on a 7-point scale with higher scores indicating more adaptive environments. All observers attended two full day trainings led by a CLASS-S master trainer. They then completed an online coding test of actual classroom interactions and needed to reach a high level of agreement with the CLASS-S master coder before being certified to collect data. Additionally, observers needed to repeat the certification each year of the project. Because we only collected postintervention STOIC ratings for the first cohort, we used baseline climate subscale as a covariate to equate classrooms on baseline climate. The CLASS-S scales have been shown to be highly reliable and to predict student achievement and social outcomes in a number of studies of large numbers of fifth graders (NICHD ECCRN, 2005) and work with teachers in secondary settings (Allen et al., 2013). The ICC for the climate subscale across all time periods was .75.

Teacher use of proactive strategies. Independent observers also conducted direct observations of teacher use of proactive strategies using the *Multi-Option Observation System for Experimental Studies* (MOOSES; Tapp, 2004) interface for hand held computers to gather real time data using the *Brief Classroom Interaction Observation Revised* observation code (BCIO-R; Reinke, Stormont, Herman, Wachsmuth, & Newcomer, 2015). These observations occurred at the same time points as the STOIC and CLASS-S, but not by the same observer who collected those observations.

The BCIO-R is a 20-min class-wide observation of the frequency of teacher use of proactive classroom management strategies, including praise statements and precorrections, and reactive strategies (i.e., use of reprimands), gathered simultaneously during each observation. Prior studies have shown that these single 20-min observations are significantly correlated with teacher self-reported classroom management self-efficacy and emotional exhaustion and are sensitive to change over time (Reinke et al., 2015). That is, teachers who received training to increase their use or proactive strategies had significantly better BCIO-R scores compared to those who did not, controlling for baseline observations (Reinke et al., 2018; Reinke, Stormont, Herman, & Newcomer, 2014; Reinke et al., 2015).

Observers attended a 2-hr training where they reviewed the operational definitions of the targeted behaviors and then practiced coding videos of actual classrooms. Observers continued to practice after the training until they reached a minimum threshold of agreement with master coders (85%). Observers than practiced in

the field with a master coder and only began collecting data after reaching the minimum threshold of agreement. The observation team met regularly during study collection periods to minimize drift.

The MOOSES program utilizes second-by-second comparison of raters to determine reliability for each variable by determining a match between observers within a 5-s window. If a match was found, then an agreement for that variable was tallied. Variables that were not matched were tallied as disagreements. An agreement ratio was then reported for each variable (agreements divided by the sum of agreements plus disagreements). Ongoing reliability checks were conducted randomly for between 32% to 42% of the observations across time points. The mean percentage agreement across time points on the BCIO-R was 92.3%, ranging from 90% to 95% for the four time points. An overall reliability of 80% is considered acceptable (Tapp, 2004).

Student Outcome Measures

Teacher report of child social behavior and academics. The *Teacher Observation of Classroom Adaptation-Checklist* (TOCA-C; Koth, Bradshaw, & Leaf, 2009) is a 54-item measure of child behavior. It was completed by the classroom teachers for each child. Teachers rated each student at the beginning (September) and end (April/May) of the school year. They rate each child on the items referencing the past 3 weeks. The four subscales of the TOCA-C included in the present study were disruptive behaviors ("breaks rules," "harms others"), concentration problems ("pays attention," "works hard"), emotional dysregulation ("stops and calms down when angry or upset"), and prosocial behavior ("is friendly," "shows empathy"). The item responses ranged from 1 (*never*) to 6 (*almost always*). Prior studies support the TOCA's internal consistency, consistent factor structure over time, predictive and current validity, and sensitivity to change across elementary and secondary school samples (Bradshaw, Waasdorp, & Leaf, 2012; Koth et al., 2009; Petras, Masyn, & Ialongo, 2011; Stormshak, Bierman, Bruschi, Dodge, & Coie, 1999). For instance, longitudinal data from the Prevention Intervention Research Center at Johns Hopkins University indicated that concentration problems scores in first grade predict likelihood of high school dropout and ratings on the disruptive behavior subscale in elementary school are a strong predictor of violence in adolescence and adulthood (e.g., Petras, Chilcoat, Leaf, Ialongo, & Kellam, 2004). Prosocial behaviors, concentration problems, and disruptive behaviors also all significantly predict office discipline referrals (Pas, Bradshaw, & Mitchell, 2011). Previous research of the TOCA-C has found internal consistency estimates ranging from .86 to .96. For the current study, the internal consistency (computed using Cronbach's alpha) for each subscale ranged from .77 to .96.

Standardized academic achievement. Grade-Level Assessments (GLAs) are assessed using the Missouri Assessment Program (MAP) which is a standardized, state-wide assessment administered to students in Grades 3 through 8 in the spring of every school year. This criterion-referenced test was designed to measure student achievement toward state-level standards. Data included in the current study are from the end-of-year mathematics and English arts subtests of the MAP. Since 2014 the GLA assessments are online assessments administered by the district's testing vendor. Scale scores produced for each student describes achievement on a continuum that spans third to eighth grades. MAP scaled scores had acceptable Cronbach's alpha coefficients. Specifically, reliability of the communication arts test was 0.87 for sixth grade, 0.90 for seventh grade, and .91 for eighth grade, and the mathematics test produced reliability coefficients of 0.88 for sixth grade, 0.90 for seventh grade, and 0.87 for the eighth grade versions of the test (Missouri Department of Elementary & Secondary Education, 2015). Within a content area MAP scores of adjacent grades can be compared.

Additionally, we administered subtests of the *Stanford Achievement Test Tenth Edition* (SAT-10; Harcourt Assessment, Inc., 2004) pre, post, and in the spring of the following year. The SAT-10 is a widely used group-administered standardized measure of academic achievement developed around national and state curriculum standards as well as those trends promoted by national professional educational groups (Harcourt Assessment, Inc., 2004). It is designed to estimate academic achievement in reading, math, language arts, and science. Extensive research documents the reliability and construct validity of the SAT-10 (Harcourt Assessment, Inc., 2004). Subtest coefficient alphas all exceed .80. We used two subtests, the reading comprehension subtests for students in reading/English classes and the problem solving subtest for students in math classes. Assessment occurred post intervention in April and May of the same school year.

Student classroom and homework completion. Teachers rated students' work completion at the beginning and end of the year using two single-item scales: "In general, what percentage of assigned homework does this student complete fully?" and "In general, what percentage of classwork does this student complete fully?" Items were rated from 0% to 100%.

Student time-on-task. Student time-on-task was measured at baseline and end-of-year using the direct observation tool, *Student Teacher-Classroom Interaction Observation* (ST-CIO; Reinke, Herman, & Newcomer, 2016). The ST-CIO is a 5 min observation of each student during classroom instruction. During the observations, observers use a duration code to indicate when the student is on-task on handheld device running the MOOSES system. Time-on-task was operationalized as, "Student is engaging with instructional content or activity via choral response, raising hand, responding to teacher instruction, listening, writing, reading, or otherwise completing assigned task," and was distinguished from time-off-task (i.e., "Student is obviously not working on assigned task or attending to task or lesson"). Research supports the concurrent and predictive validity of these 5-min observations. Recent studies using the ST-CIO found significant relations with teacher ratings of disruptive behaviors concurrently and over time (Lewis et al., 2017; Reinke et al., 2016).

Observers were trained for 2 weeks using videos and practice sessions in live classrooms to a criterion of 85% reliability with a master coder prior to conducting observations in study classrooms. Reliability checks were conducted for between 32% to 42% of the observations across time points. To determine reliability, two observers began the observation of an individual student at precisely the same time. One was considered the primary coder. The primary coder's data were used for the study. The secondary coder's data were used for reliability purposes only. The kappa for time-on-task ranged from .94-.97 across time points. A kappa of .80 is acceptable and considered strong agreement (McHugh, 2012).

Student demographics. Free and reduced lunch status (FRL), race, sex, age, grade, and special education status were obtained from the school district for all participating students. Students were coded as 1 if they received FRL and 0 if not. Student sex was coded as 1 for female and 0 for male. Students receiving special education were coded as 1 and if not 0. For the purposes of this study, student race was coded as two dummy variables indicating African American or other race with White serving as the reference group for both.

Analytic Approach

For the data analysis, we used multiple imputation for handling missing data (Rubin, 1987; Schafer & Olsen, 1998). We checked covariate balance by calculating the effect sizes of the covariates between the treatment and control groups to make sure that randomization worked to produce equivalent groups. We then used hierarchical linear models (HLM) to count for nested data structure (e.g., students nested within teachers) for main effect analysis of the teacher and student outcomes. Finally, we used multilevel mediation analysis to test our theory. The details are below.

Missing data. The original sample included 102 teachers and 1,450 students in nine schools. One teacher moved out of the district during the fall semester of the school year. Missing data occurred primarily on the outcome measures. The missing rates for the pretests of four social and behavioral outcome measures is 0.5% while the missing rates for the posttests of four social and behavioral outcome measures is 14.2% in the overall sample; the vast majority of this missing data was the result of students moving out of the school district during the year. The differential missing rates between the treatment and control groups are −0.7% for the pretest and 2.7% for the posttest. Based on the What Works Clearinghouse attrition standard, the combination of an overall attrition rate of 14.2% and a differential attrition rate of 2.7% would result in low levels of potential bias (i.e., greater than 0.05 of standard deviation) even under the more conservative assumptions (What Works Clearinghouse, 2014). Hence, the results from the analysis of the students who have complete posttests will have good internal validity. The literature also showed that when the outcome is included in the imputation model, there are very small differences between models that impute the outcome compared with those that do not (Kontopantelis, White, Sperrin, & Buchan, 2017). The final analytic samples included nine schools 101, teachers, and 1,244 students for the analyses of social and behavioral outcomes; among 101 teachers, 47 teachers in math class (587 students for problem solving and 594 for MAP math) and 54 teachers in reading class (632 students for reading comprehension and 646 students for MAP English) for the analyses of academic achievement outcomes.

The maximum overall data missing rate and differential missing rate between the treatment and control group in the final analytic samples for the analysis of social and behavioral outcomes were 0.6% and −0.8%, respectively. The maximum overall data missing rate and differential missing rate between the treatment and control group in the final analytic samples for the analysis of academic outcomes were 11.2% and 3.0%, respectively. Multiple imputation using a Markov chain Monte Carlo (MCMC) method in SAS PROC MI was used to impute missing data by including posttest, pretest, and other covariates. We imputed five times for the final analytic samples for the analysis of social and behavioral outcomes and 30 times for academic outcomes based on the missing rates (Rubin, 1987; Schafer & Olsen, 1998).

Analysis of teacher implementation. First, to evaluate whether teacher implementation of proactive classroom management skills increased following receipt of the CHAMPS intervention, we conducted the longitudinal analysis. Because of the small school sample size ($K = 9$) and small nonsignificant variance at the school level, we chose to fit a linear growth curve model using two-level hierarchical linear modeling (HLM) using SAS PROC MIXED. The repeated measures (Level 1) are nested within teachers (Level 2). We controlled for the baseline pretest in evaluating the treatment effects on teacher implementation of proactive classroom management skills. We also calculated the mean rate of praise, precorrections, and reprimands observed at each time point to demonstrate any changes in the base rate of the teacher behaviors.

Analysis of main effects on student outcomes. Because of the small school sample size ($K = 9$) and small nonsignificant variance at school level (e.g., the intraclass correlation coefficients at school level ranged from 0.001 to 0.034), we chose to fit a two-level HLM for each of the five imputed data sets, in which students (Level 1) are nested within teachers (Level 2). We used SAS PROC MIXED to estimate the overall treatment effects student behavior and academic outcomes. Each student's pretest and demographic information were included at Level 1, and the treatment variable was at Level 2. SAS PROC MIANALYZE was used to combine the results from the analyses of five data sets. The statistical model is below:

Level 1 (student):

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qj} + r_{ij},\ r_{ij} \sim N(0, \sigma^2)$$

Level 2 (class):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Intervention)_j + u_{oj},\ u_{oj} \sim N(0, \tau)$$

$$\beta_{qj} = \gamma_{01}, q = 1, \ldots, Q$$

where X_{qj} represents student-level covariates, which include pretest, age at pretest, gender, race, FRL, special education status, grade level, and cohort year in the study. Specifically for race, we included two dummy variables representing African American and other race in the model with White serving as the reference race group for each. $(Intervention)_j$ is a binary variable indicating treatment condition ($Intervention = 0$ for control group and $Intervention = 1$ for treatment group). The parameter, γ_{01}, estimates the overall treatment effect. σ^2 and τ are variance components for Level 1 and Level 2 residuals conditional on these covariates. Similar analyses were conducted for the analysis of the academic achievement outcomes, except that we analyzed 30 imputed data sets.

Mediation analyses. We used Baron and Kenny's (1986) framework to conduct multilevel mediation analysis (Zhang, Zyphur, & Preacher, 2009). This multilevel mediation analytic approach was recently used by Curenton, Dong, and Shen (2015) to test the hypothesis the former early childhood education attendees' fifth grade achievement was mediated by the aggregate schoolwide achievement of their elementary school, and we applied a similar procedure for our 2–1–1 mediation analyses (see Figure 2): (a) Path *a*: examined if the CHAMP intervention (Level 2 variable

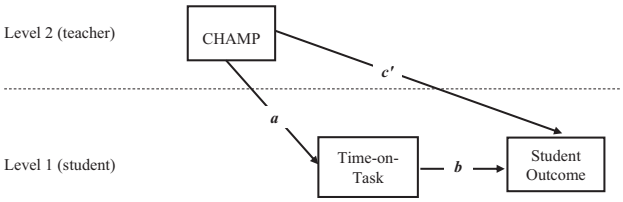

Figure 2. 2–1–1 multilevel mediation model.

at Time 1) changed the percentage student time-on-task (Level 1 mediator at Time 4); (b) Path *b*: examined if the percentage of observation student on task predicted the individual student outcomes (i.e., Level 1 outcomes of achievement at Time 4) while controlling for the CHAMP intervention status; (c) Path *c*: examined if the CHAMP intervention changed student outcomes at Time 4; and (d) Path *c′*: examined if the CHAMP intervention at Time 1 and the mediators at Time 4 predicted the student outcomes at Time 4 (i.e., the statistical model is same as in Path *b*). Two-level HLMs were used in all path analyses after the outcome and mediator variables were standardized with means of 0 and standard deviations of 1. The coefficient of CHAMP intervention in Path *c′* indicates the direct effect of CHAMP intervention on student academic achievement. The product of the coefficient of CHAMP intervention in Path *a* and the coefficient of the percentage of observation student on task in Path *b* indicates the indirect effect of CHAMP intervention on student academic achievement, that is, the mediated effect of CHAMP intervention on student outcomes through changing the percentage of observation student was on task. The coefficient of Path *c* indicated the total effect of CHAMP intervention. In all the path analyses, we controlled for the covariates. We calculated the 95% confidence interval through Monte Carlo simulation (Zhang et al., 2009).

Results

Descriptive Statistics and Covariate Balance Checking

Table 1 provides descriptive statistics and covariate balance checking for the analytic sample of social behavioral outcomes at baseline. Effect sizes are also provided in Table 1, indicating that the baseline measures were equivalent between the two conditions.

Teacher Adherence to CHAMPS

A two-level HLM of STOIC ratings from ratings from postintervention in October and December, and at the end of the school year, adjusting for baseline CLASS-S climate scores, revealed a significant intervention effect on STOIC summary scores across three postintervention observations adjusting for pretest scores ($b = 0.22$, $p = .007$). CHAMPS teachers had consistently higher ratings on STOIC and the average differences across three posttest observations represented a medium effect ($d = 0.62$), indicating teachers in the CHAMPS condition demonstrated adherence to the intervention.

Teacher Implementation of Proactive Classroom Management

To evaluate whether teachers receiving CHAMPS demonstrated an increase in their implementation of proactive strategies in comparison to control teachers, a two-level HLM was conducted on BCIO-R positive (praise and precorrections) to negative (reprimands) ratios across three time points (see Table 2) controlling for the baseline pretest positive-negative implementation ratio. Analyses revealed a significant intervention effect on positive-negative ratios across three postintervention observations (October, December, and end of the school year), adjusting for pretest

Table 1
Covariate Balance Checking for the Analytic Sample of Social Behavioral Outcomes at Baseline

	Control		Treatment		
Baseline variable	*M*	*SD*	*M*	*SD*	Effect size
Female	0.49	0.50	0.53	0.50	0.08
Lunch	0.69	0.46	0.66	0.47	−0.06
Special education	0.10	0.30	0.08	0.27	−0.07
White	0.20	0.40	0.18	0.39	−0.04
Black	0.76	0.43	0.77	0.42	0.02
Other race	0.04	0.19	0.05	0.21	0.04
Year 1	0.29	0.46	0.34	0.47	0.10
Year 2	0.30	0.46	0.31	0.46	0.01
Year 3	0.27	0.44	0.20	0.40	−0.15
Year 4	0.14	0.34	0.15	0.35	0.03
Grade 6	0.32	0.47	0.41	0.49	0.19
Grade 7	0.37	0.48	0.41	0.49	0.08
Grade 8	0.31	0.46	0.18	0.39	−0.31
Math class	0.42	0.49	0.54	0.50	0.22
Age (month)	12.72	0.89	12.50	0.82	−0.25
TOCA–concentration problems	3.02	1.28	2.93	1.24	−0.07
TOCA–disruptive behavior	1.89	0.76	1.78	0.71	−0.15
TOCA–prosocial behavior	4.48	0.94	4.55	0.93	0.07
TOCA–emotion regulation	2.38	1.08	2.32	0.99	−0.07
N	637		607		

Table 2
Mean Rate and Standard Deviation of Teacher Use of Praise, Precorrective Statements, and Reprimand Across Time Points

	Time 1		Time 2		Time 3		Time 4	
	M (*SD*)		*M* (*SD*)		*M* (*SD*)		*M* (*SD*)	
Teacher behavior	CHAMPS	Control	CHAMPS	Control	CHAMPS	Control	CHAMPS	Control
Praise	0.68 (0.40)	0.64 (0.35)	1.23 (0.64)	0.82 (0.68)	1.20 (0.63)	0.90 (0.55)	1.03 (0.71)	0.63 (0.38)
Precorrection	0.02 (0.03)	0.02 (0.03)	0.03 (0.04)	0.03 (0.04)	0.02 (0.05)	0.02 (0.04)	0.02 (0.03)	0.02 (0.03)
Reprimand	0.84 (0.53)	0.70 (0.39)	0.65 (0.45)	0.59 (0.43)	0.61 (0.44)	0.57 (0.39)	0.51 (0.34)	0.58 (0.34)

scores ($b = 12.28$, $p = .006$). Adjusted means scores across three postintervention time points indicated CHAMPS teachers had positive-negative ratio of 50.1% versus 37.6% for wait-list teachers, and this represented a small-to-moderate effect ($d = 0.39$), indicating improvements in use of proactive classroom management strategies for teachers receiving the CHAMPS intervention.

Main Effects on Student Social Behavior

Tables 3–5 provides the fixed and random effects of two-level HLM analysis of the main effects of the intervention on social behavioral outcomes and teacher reported outcomes. In addition to the conditional intraclass correlation coefficients (ICCs), for reference we reported the unconditional ICCs that were calculated using the two-level HLM without including any predictors. The ICCs indicate the proportion of the variance at teacher level in the total variance. Most of the unconditional ICCs of social behavioral outcomes and teacher reported outcomes varied from 0.185 to 0.297 except for the student time-on-task measure (0.052). Main effect analyses demonstrated that students in CHAMPS classrooms showed a significant reduction on teacher reported concentration problems ($b = -0.18$, $p = .017$, $d = -.14$; Table 3) compared with students in the control condition. Although all social behavior effects were in the expected direction, there were no significant effects of CHAMPS on teacher reported emotional dysregulation, disruptive behaviors, or prosocial behaviors. There was a significant main effect on student classwork completed ($b = 4.32$, $p = .008$, $d = 0.18$) but not for homework completed in comparison with control classrooms (see Table 4). Finally, the intervention had a significant main effect on student time-on-task ($b = 2.84$; $p = .014$, $d = 0.16$; see Table 5).

Main Effects on Student Academic Outcomes

Tables 6 and 7 show the fixed and random effects of two-level HLM analysis of the main effects of intervention on student academic outcomes. The intervention had a significant main effect on MAP English ($b = 0.12$, $p = .044$, $d = 0.14$) and SAT-10 problem solving ($b = 5.66$, $p = .049$, $d = 0.17$) scales controlling for prior year scores. However, there were no significant effects on the MAP math ($b = 0.15$, $p = .11$, $d = 0.16$) or SAT-10 reading comprehension ($b = 3.04$, $p = .36$, $d = 0.08$) scores in comparison with control classrooms.

Table 3
Two-Level HLM Results for the Effects of CHAMPS on Social Emotional and Behavior Outcomes

	Concentration problems			Disruptive behavior			Emotional dysregulation			Prosocial behavior		
Variable	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Fixed effect												
Intercept	−0.58	0.69	.400	0.22	0.50	.659	0.70	0.65	.280	1.35	0.63	.033
Age	0.11	0.06	.053	0.02	0.04	.694	0.00	0.06	.967	−0.02	0.06	.755
Female	−0.16	0.06	.004	−0.05	0.03	.142	−0.01	0.05	.822	0.09	0.04	.024
Lunch status	0.04	0.06	.524	0.05	0.03	.129	0.00	0.04	.935	−0.05	0.05	.349
Special education	−0.05	0.07	.493	−0.05	0.07	.459	−0.01	0.09	.911	−0.06	0.09	.468
African American	0.24	0.06	<.001	0.11	0.04	.005	0.09	0.06	.091	−0.11	0.06	.062
Other race	−0.04	0.12	.731	−0.02	0.07	.794	−0.12	0.10	.233	0.18	0.09	.040
Year 2	−0.04	0.09	.685	−0.02	0.06	.784	0.00	0.07	.989	0.09	0.08	.227
Year 3	0.15	0.10	.128	0.07	0.06	.265	0.06	0.08	.494	−0.10	0.08	.209
Year 4	0.27	0.14	.051	0.10	0.10	.284	0.08	0.12	.512	0.02	0.11	.879
Grade 7	−0.24	0.11	.026	−0.05	0.08	.520	−0.10	0.10	.323	0.11	0.09	.236
Grade 8	−0.15	0.15	.301	0.07	0.10	.506	0.15	0.15	.328	−0.03	0.14	.862
Pretest	0.72	0.02	<.001	0.80	0.03	<.001	0.75	0.03	<.001	0.74	0.03	<.001
Intervention	**−0.18**	**0.07**	**.017**	**−0.07**	**0.05**	**.169**	**−0.09**	**0.07**	**.192**	**0.08**	**0.06**	**.195**
Random effect												
Teacher	0.075	0.019	<.001	0.039	0.009	<.001	0.058	0.015	<.001	0.051	0.013	<.001
Student	0.642	0.027	<.001	0.239	0.010	<.001	0.499	0.021	<.001	0.448	0.019	<.001
Conditional ICC	0.105			0.139			0.104			0.101		
Unconditional ICC	0.185			0.245			0.216			0.242		

Note. Sample size: 101 teachers and 1,244 students in nine schools. HLM = hierarchical linear model. Bold indicates treatment effects.

Table 4
Two-Level HLM Results for the Effects of CHAMPS on Teacher Reported Outcomes

	Classwork completion			Homework completion		
Variable	*b*	*SE*	*p*	*b*	*SE*	*p*
Fixed effect						
Intercept	36.64	17.15	.035	55.64	19.51	.005
Age	−1.22	1.40	.384	−2.48	1.63	.128
Female	2.93	0.97	.003	4.36	1.17	<.001
Lunch status	−0.44	1.19	.711	−1.06	1.51	.482
Special education	−3.49	2.05	.089	−0.34	2.12	.874
African American	−3.08	1.47	.036	−4.33	1.65	.009
Other race	1.15	2.36	.627	−1.02	3.05	.739
Year 2	−0.77	2.06	.708	−2.81	3.34	.399
Year 3	−3.64	2.11	.084	−1.57	3.54	.658
Year 4	−4.69	2.45	.056	−4.05	3.64	.266
Grade 7	3.09	2.41	.200	5.26	4.15	.205
Grade 8	1.03	3.53	.770	4.17	4.43	.347
Pretest	0.70	0.04	<.001	0.62	0.04	<.001
Intervention	**4.32**	**1.63**	**.008**	**2.45**	**2.75**	**.373**
Random effect						
Teacher	42.81	10.14	<.001	159.02	28.56	<.001
Student	261.73	11.05	<.001	359.81	15.16	<.001
Conditional ICC	0.141			0.307		
Unconditional ICC	0.231			0.297		

Note. Sample size: 101 teachers and 1,237 students in nine schools. HLM = hierarchical linear model. Bold indicates treatment effects.

Mediation Effects

We examined the students' observed time-on-task as a potential mediator of the CHAMPS intervention on student achievement (see Table 8). We found that the percentage of time during direct observation that a student was on-task mediated the effect of the CHAMPS intervention on the MAP English scores. Path *a* analysis indicated a significant effect of the CHAMPS intervention on the observed student time-on-task posttest (mediator; $b = 0.24$, $p =$

Table 5
Two-Level HLM Results for the Effects of CHAMPS on Student Time-on-Task

	Percentage of observation student on task		
Variable	*b*	*SE*	*p*
Fixed effect			
Intercept	101.07	19.22	<.001
Age	−0.80	1.64	.625
Female	0.43	1.14	.705
Lunch status	−0.46	1.27	.716
Special education	−0.53	1.97	.788
African American	−2.05	1.25	.101
Other race	−1.98	2.84	.486
Year 2	−0.56	1.27	.662
Year 3	−6.65	1.80	<.001
Year 4	1.54	1.80	.392
Grade 7	0.73	2.25	.747
Grade 8	0.75	3.82	.844
Pretest	0.04	0.03	.159
Intervention	**2.84**	**1.15**	**.014**
Random effect			
Teacher	6.28	4.52	.083
Student	308.58	12.93	<.001
Conditional ICC	0.020		
Unconditional ICC	0.052		

Note. Sample size: 101 teachers and 1,235 students in nine schools. HLM = hierarchical linear model. Bold indicates treatment effects.

Table 6
Two-Level HLM Results for the Effects of CHAMPS on MAP Academic Outcomes

	MAP math			MAP english		
Variable	*b*	*SE*	*p*	*b*	*SE*	*p*
Fixed effect						
Intercept	0.83	0.95	.379	1.51	0.77	.051
Age	−0.07	0.08	.353	−0.14	0.07	.027
Female	0.01	0.04	.850	0.10	0.05	.050
Lunch status	−0.08	0.06	.131	−0.10	0.05	.074
Special education	−0.24	0.09	.009	−0.23	0.11	.028
African American	−0.22	0.07	.003	0.00	0.08	.977
Other race	0.04	0.11	.737	0.25	0.11	.031
Year 2	0.01	0.13	.925	0.08	0.06	.195
Year 3	0.01	0.10	.898	−0.16	0.07	.024
Year 4	−0.31	0.15	.033	−0.27	0.10	.005
Grade 7	0.18	0.12	.146	0.26	0.10	.012
Grade 8	0.50	0.19	.007	0.47	0.16	.004
Pretest	0.61	0.06	<.001	0.71	0.05	<.001
Intervention	**0.15**	**0.10**	**.105**	**0.12**	**0.06**	**.044**
Random effect						
Teacher	0.078	0.023	<.001	0.010	0.007	.078
Student	0.321	0.019	<.001	0.304	0.018	<.001
Conditional ICC	0.196			0.031		
Unconditional ICC	0.333			0.288		

Note. Sample size: 47 teachers and 594 students for MAP math, and 54 teachers and 646 student for MAP English in nine schools. HLM = hierarchical linear model. Bold indicates treatment effects.

Table 7
Two-Level HLM Results for the Effects of CHAMPS on SAT-10 Academic Outcomes

	SAT-10 Problem solving			SAT-10 Comprehension		
Variable	*b*	*SE*	*p*	*b*	*SE*	*p*
Fixed effect						
Intercept	641.32	36.97	<.001	694.48	31.04	<.001
Age	−0.03	3.19	.993	−4.16	2.61	.112
Female	2.54	2.09	.224	3.57	2.60	.170
Lunch status	−4.68	2.48	.059	−4.55	2.95	.123
Special education	4.47	3.73	.231	−2.87	3.70	.438
African American	−8.99	2.94	.002	−2.69	3.81	.480
Other race	7.16	5.64	.204	11.48	5.92	.052
Year 2	3.97	4.05	.326	−2.29	3.81	.548
Year 3	2.73	4.60	.553	−8.33	3.66	.023
Year 4	1.40	5.06	.782	−7.61	4.13	.065
Grade 7	13.51	5.42	.013	15.85	4.97	.001
Grade 8	23.16	8.31	.005	43.91	6.87	<.001
Pretest	18.90	1.82	<.001	26.02	1.96	<.001
Intervention	**5.66**	**2.87**	**.049**	**3.04**	**3.31**	**.359**
Random effect						
Teacher	34.09	17.90	.029	39.37	19.64	.024
Student	659.85	39.99	<.001	724.91	42.44	<.001
Conditional ICC	0.049			0.052		
Unconditional ICC	0.155			0.271		

Note. HLM = hierarchical linear model. Bold indicates treatment effects.

Table 8
HLM Results for 2–1–1 Mediation Effects of CHAMPS on MAP English Through Observation Student on Task

Variable	Path *a* (Treat → Observation student on task posttest [mediator])			Path *b* (Path *c'*) (Observation student on task posttest [mediator] → MAP English)			Path *c* (Treat → MAP English)		
	b	*SE*	*p*	*b*	*SE*	*p*	*b*	*SE*	*p*
Intercept	1.85	1.75	.290	1.77	0.88	.045	1.81	0.88	.040
Age	−0.16	0.15	.280	−0.16	0.08	.035	−0.16	0.08	.030
Female	−0.05	0.08	.560	0.11	0.06	.080	0.11	0.06	.075
Special education	−0.01	0.15	.954	−0.25	0.12	.040	−0.25	0.12	.041
Other race	0.12	0.13	.359	0.25	0.12	.030	0.26	0.11	.025
Year 3	−0.34	0.12	.007	−0.23	0.08	.007	−0.24	0.08	.004
Year 4	0.07	0.10	.485	−0.40	0.11	<.001	−0.40	0.11	<.001
Grade 7	0.27	0.20	.171	0.29	0.12	.015	0.30	0.12	.014
Grade 8	0.35	0.35	.313	0.52	0.18	.005	0.53	0.19	.005
MAP English Pretest	NA			0.67	0.05	<.001	0.67	0.05	<.001
Intervention	**0.24**	**0.08**	**.004**	**0.13**	**0.07**	**.060**	**0.14**	**0.07**	**.038**
Observation student on task pretest	0.08	0.04	.085	0.02	0.02	.449	NA		
Observation student on task posttest	NA			**0.04**	**0.02**	**.013**	NA		

Note. Sample size: 54 teachers and 641 students in nine schools. HLM = hierarchical linear model. Bold indicates treatment effects.

.004), and Path *b* analysis indicated a significant effect of the observed student time-on-task posttest (mediator) on MAP English scores ($b = 0.04$, $p = .013$). The indirect (mediation) effect was 0.01 with a 95% confidence interval of (0.001, 0.023) while the direct effect was 0.13 ($p = .06$).

Discussion

This group RCT investigated the efficacy of CHAMPS among teachers in middle school classrooms on student social behavior and academic outcomes. Observational data indicated that CHAMPS teachers demonstrated an increase in proactive classroom management skills targeted by the intervention. Additionally, students in CHAMPS classrooms had significant improvements in teacher-reported concentration problems and classwork completed, observed time-on-task, and academic achievement. Finally, subsequent analyses suggested that improvements in English achievement were partially mediated by the observed increases in student time-on-task.

It was expected that CHAMPS would produce significant impacts on student prosocial behaviors, disruptive behaviors, and emotion regulation problems in addition to the effects on concentration problems such as has been found with other universal behavior interventions (Bradshaw et al., 2012; Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Reinke et al., 2018). Although all effects were in the expected direction, only main effects on teacher-reported concentration problems were statistically significant. In a prior trial in elementary schools with a related but different classroom management training program (IY TCM), we found nearly the opposite findings, namely significant effects on teacher-reported prosocial behaviors and emotion dysregulation but not on concentration problems (Reinke et al., 2018). Some noteworthy differences in the training programs and the studies may have accounted for these differences. First, the prior study was conducted in K–3 classrooms compared with the middle school context of the present study. In the elementary trial, students spent nearly their entire school day with the teacher being trained in IY TCM. In the present study, given the nature of middle schools, students only spent one class period a day with the teacher trained in CHAMPS. Thus, the potential for teacher impact on student social outcomes in middle schools is somewhat reduced by the lower amount of time teachers spend with students. Second, in addition to an emphasis on classroom structure and routines, the IY TCM training program supports teachers in developing social-emotional coaching skills to improve student social and emotion regulation skills. On the other hand, CHAMPS emphasizes the importance of classroom structure and predictability without explicitly teaching social and emotional development skills. Thus, improvements in prosocial and emotion regulation skills may require explicit instruction and/or social coaching.

In addition to the null findings on several social behavioral outcomes, not all academic outcomes were significant. The intervention did not impact student math performance on the state achievement test or a comprehension subtest of the SAT-10. These differential outcomes on distinct academic achievement tests suggest that the program had selective benefit for specific math skill development (problem solving) whereas it had global benefit for comprehensive English skill development. It is worth noting that, although not significant, the effect size of CHAMPS on the broad state math achievement test (.16) was comparable with the effect sizes of both significant achievement findings. Additionally, the program had a significant benefit for teacher ratings of student classwork but not homework completion. This indicates that the effects of effective classroom management are proximal to the classroom environment and may not increase student academic engagement outside of the class. Other strategies may be needed to foster increases in homework completion.

That academic outcomes were achieved in the context of this study within a single academic year is important. In a prior trial evaluating the MTP program in middle schools, effects on academic outcomes were not observed within the first cohort of students; rather, effects emerged during the year after the training program. It is especially striking that CHAMPS created positive student academic outcomes within the same school year given an artifact of the study design. Because baseline data was needed,

CHAMPS teachers were not fully trained in the model until December and may not have achieved full implementation until the Spring semester. This limited window of impacting student learning suggests that CHAMPS may hold potential for even larger effects on student outcomes when implemented over the course of an entire school year and across multiple teachers and classrooms.

Some of the findings were consistent with CHAMPS theory of change. For instance, student engagement as indicated by observed time-on-task, partially mediated CHAMPS effects on English achievement. Although it is well-established that effective classroom management strategies, including those embedded in CHAMPS, significantly reduce student problem behaviors (Oliver et al., 2011), much less literature has examined whether these strategies alone improve achievement. Unlike prior studies, we did not find a reduction in student disruptive behavior, but instead found improvements in student concentration and attention. This finding suggests that effective classroom management can improve student achievement, in part, by increasing the amount of time students are attentive, on-task, and exposed to instruction (Fredricks, Blumenfeld, & Paris, 2004; Gromada & Shewbridge, 2016). This complements prior research on the effects of MTP-S, which found that outcomes were mediated by improvements in student autonomy (Allen et al., 2011). It is possible, though speculative, that student increases in autonomy led to improvements in student attention and time-on-task, which in turn led to increased academic performance. Unlike MTP-S which includes both relational supports and instructional skill training, CHAMPS principally focuses on classroom behavior management. Thus, the current findings suggest that improving behavior management skills of teachers by itself leads to better student academic performance.

Intervention Effect Sizes

The size of intervention effects on observed teacher practices were in the moderate range ($d = 0.39–0.62$). Given that CHAMPS is a universal preventive intervention and all teachers were eligible to participate regardless of risk or baseline functioning, these are promising levels of impact on teacher performance.

On the other hand, the effect sizes on student outcomes were relatively modest, and similar to effects reported in response to MTP (Allen et al., 2011). It is not surprising that CHAMPS would have larger proximal effects on the direct recipients of the training and coaching—teachers—compared with effects on students who are more downstream from the intervention. Notably, small effect sizes are common in longitudinal universal prevention studies (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Flay et al., 2005), given that they are delivered to entire populations with varying degrees of risk (i.e., many individuals would not develop behavior or academic problems even without the intervention). However, very small effects on a population level can result in dramatic improvements in public health outcomes (NRC & IOM, 2009). Based in his review of causal studies including 1,942 effect sizes across 747 RCTs evaluating education interventions, Kraft (2020) suggested the following benchmarks be used for large studies that intervene with teachers and use broad achievement scores as outcomes: 0.05 = small; less than 0.20 = medium; greater than 0.20 = large. Using these benchmarks, CHAMPS' effect sizes on student achievement would be interpreted as medium effects.

Scholars have advocated for different methods for expressing practical significance and interpretability of findings in the context of school-based intervention studies including reporting percentile gains and using empirical benchmarks (Bloom, Hill, Black, & Lipsey, 2008; Dong, Reinke, Herman, Bradshaw, & Murray, 2016; Hill, Bloom, Black, & Lipsey, 2008; Lipsey et al., 2012). In line with these recommendations, on average, the academic outcomes in this study, equated to an increase in English performance from the 50th to the 55th percentile and math problem solving to the 56th percentile for students in CHAMPS classrooms versus students in the control classrooms (Bloom et al., 2008, Table 3). Additionally, Hill, Bloom, Black, and Lipsey (2008) argued that effect sizes should be interpreted with respect to empirical benchmarks that are relevant to the intervention, target population, and outcome measure being considered. In particular, these benchmarks can include policy-relevant performance gaps, such as the performance gaps between White and African American students. Prior research has provided these benchmarks regarding academic achievement (Hill et al., 2008). A recent study reports policy relevant demographic performance gaps on social behavior, including the White–African American, girls–boys, and ineligible–eligible free or reduced-price lunch achievement gaps (see Dong et al., 2016). Educational researchers have suggested that to understand the effect sizes of educational interventions, using these empirical benchmarks that are relevant to the intervention, target population, and outcome measure should be considered (Bloom et al., 2008; Hill et al., 2008; Lipsey et al., 2012). For instance, the magnitude of the effect sizes on math and English performance are equivalent to reducing 25.8% and 17.2% of the achievement gap between students eligible and ineligible for free/reduced-price lunch (Bloom et al., 2008, Table 4).

Implications

The findings from this study provide important implications for promoting effective classroom environments in schools. Many teachers report struggling with classroom management and note that they receive little training in this area prior to entering the classroom (Reinke et al., 2010). One study found that only 27% of teacher preparation programs devoted an entire course to classroom management (Oliver & Reschly, 2010). Moreover, very few classroom management programs exist to support middle school teachers. The present findings suggest that CHAMPS could be an effective approach to improve classroom management practices of inservice middle school teachers. This is important given that so many schools have already adopted CHAMPS in the absence of evidence. Middle schools are encouraged to continue using CHAMPS and/or consider adopting it as a professional development approach for their teachers. Educational psychologists can advocate for universal prevention programs, such as CHAMPS, that have the potential to positively impact large numbers of students behaviorally and academically. This is an important role for educational psychologists, as many schools struggle to identify evidence-based programs and practices (Stormont, Reinke, & Herman, 2011). Additionally, teacher preparation programs should consider infusing CHAMPS principles and practices into their curriculum to prepare aspiring middle school teachers to deliver effective management practices upon graduation.

Further, the findings in this sample of predominantly African American youth from families with lower economic means in an urban context suggests CHAMPS principles and practices may be generalizable to schools across the nation with diverse student populations and those living in higher risk settings. This is particularly encouraging given the heightened interest and attention to longstanding achievement gaps between racial and ethnic groups in the United States. Using benchmark strategies as described in this discussion can provide educators and policymakers with critical information for making decisions about new programs and practices to reduce these gaps.

Study Limitations and Future Directions

While the findings are interesting and important, this study is not without some limitations. The findings from the study on student behavior outcomes were predominantly based on teacher report. Teachers are the most common source of students' social behavior and special education referrals (Zima et al., 2005), thus their perspectives are important in the context of school-based interventions. Further, teacher reports of students' social behavior predict behavior problems (Darney, Reinke, Herman, Stormont, & Ialongo, 2013; Koth et al., 2009; Reinke, Herman, Petras, & Ialongo, 2008; Schaeffer, Petras, Ialongo, Poduska, & Kellam, 2003). Teacher perceptions are important in their own right given that these perceptions influence student referral for special education, teacher stress and burnout levels, as well as teacher self-efficacy and competence. Additionally, an objective measure of student engagement given by direct observation, time-on-task, suggested that the teacher ratings of improved student concentration problems were evident to blind observers.

A second study limitation concerns the mediation analyses. The mediator, observed time-on-task, was collected at the same time point as the outcome, academic performance. Thus temporal sequence cannot be established by the study design. Moreover, as is typical of mediation analyses, the mediator was not experimentally manipulated. Given these limitations, additional studies are needed to confirm the hypothesized causal mechanisms of CHAMPS on academic achievement.

Third, our time-on-task measure only captured a limited aspect of academic learning time, procedural engagement (Spanjers et al., 2008). Procedural engagement refers to the observable indicators of learning (e.g., completing tasks, looking at the teacher while talking). Substantive engagement, on the other hand, refers to a student's involvement and investment in learning (Reeve et al., 2004). Researchers have argued that procedural engagement measures are more obtuse and less sensitive to actual learning differences than deeper aspects captured by substantive engagement. That we found evidence of mediation even using this relative blunt assessment of engagement in learning suggests even more robust effects may be found with more sensitive measures of the construct.

Fourth, the present study did not examine heterogeneity of intervention effects for teachers and students. Because CHAMPS was delivered as a universal prevention intervention available to all teachers and students regardless of risk we did not hypothesize specific benefits of the intervention for subgroups of teachers or students. However, further studies will examine exploratory hypotheses to determine if particular subgroups benefitted more than others. For instance, do teachers or students with greater room for improvement at baseline (e.g., lower classroom management skills or higher levels of behavior problems) benefit more from the intervention. Perhaps fewer workshop days are necessary for teachers who have a higher rate of proactive classroom management practices. Furthermore, future research could focus on a cost-benefit analysis of the program to allow for schools to determine if this universal prevention intervention effects outweigh the resources needed to implement the intervention.

Conclusion

Many teachers struggle with classroom management issues (Reinke et al., 2011), resulting in a large number of teachers leaving the profession early in their careers (Smith & Ingersoll, 2004). The cost of teacher turnover in public schools has been conservatively estimated to be over 7 billion dollars a year (National Commission on Teaching and America's Future, 2007). In light of this, finding ways to improve the classroom management skills of teachers should be a priority for policymakers. In this study, a teacher classroom behavior management training program, CHAMPS, caused improvements in teacher classroom management practices and student social and academic outcomes. The improvements in academic achievement were, in part, explained by increases in student time-on-task. Improving teacher classroom management training holds promise for increasing student achievement on a large scale.

References

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—Secondary. *School Psychology Review, 42,* 76–98.

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: Replication and extension of the My Teaching Partner-Secondary intervention. *Journal of Research on Educational Effectiveness, 8,* 475–489. http://dx.doi.org/10.1080/19345747.2015.1017680

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333,* 1034–1037. http://dx.doi.org/10.1126/science.1207998

Alter, P., & Haydon, T. (2017). Characteristics of effective classroom rules: A review of the literature. *Teacher Education and Special Education, 40,* 114–127.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79,* 193–223.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. http://dx.doi.org/10.1037/0022-3514.51.6.1173

Barrett, E. R., & Davis, S. (1993). Perceptions of beginning teachers' needs in classroom management. *Teacher Education and Practice, 11,* 22–27.

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom 1. *Journal of Applied Behavior Analysis, 2,* 119–124.

Bloom, H. S., Hill, C. J., Black, A. B., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size

benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1,* 289–328. http://dx.doi.org/10.1080/19345740802400072

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130,* e1136–e1145. http://dx.doi.org/10.1542/peds.2012-0243

Brophy, J. E., & Evertson, C. M. (1976). *Learning from teaching: A developmental perspective*. Boston, MA: Allyn & Bacon.

Buell, M. J., Hallam, R., Gamel-Mccormick, M., & Scheer, S. (1999). A survey of general and special education teachers' perceptions and inservice needs concerning inclusion. *International Journal of Disability Development and Education, 46,* 143–156. http://dx.doi.org/10.1080/103491299100597

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher, 18,* 26–31. http://dx.doi.org/10.3102/0013189X018001026

Cicchetti, D., & Toth, S. L. (1995). Developmental psychopathology and disorders of affect. In D. Cicchetti & D. J. Cohen (Eds.), *Wiley series on personality processes. Developmental psychopathology, Vol. 2. Risk, disorder, and adaptation* (pp. 369–420). New York, NY: John Wiley & Sons.

Curenton, S. M., Dong, N., & Shen, X. (2015). Does aggregate school-wide achievement mediate fifth grade outcomes for former early childhood education participants? *Developmental Psychology, 51,* 921–934. http://dx.doi.org/10.1037/a0039295

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. (2013). Children with co-occurring academic and behavior problems in 1st grade: Distal outcomes in 12th grade. *Journal of School Psychology, 51,* 117–158. http://dx.doi.org/10.1016/j.jsp.2012.09.005

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., . . . Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14,* 317–345. http://dx.doi.org/10.1016/0193-3973(93)90013-L

Dong, N., Reinke, W. M., Herman, K. C., Bradshaw, C. P., & Murray, D. W. (2016). Meaningful effect sizes, intraclass correlations, and proportions of variance explained by covariates for planning two-and three-level cluster randomized trials of social and behavioral outcomes. *Evaluation Review, 40,* 334–377. http://dx.doi.org/10.1177/0193841X16671283

Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432. http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Evertson, C. M. (1989). Improving elementary classroom management: A school-based training program for beginning the year. *The Journal of Educational Research, 83,* 82–90. http://dx.doi.org/10.1080/00220671.1989.10885935

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute.

Flay, B., Biglan, A., Boruch, R., Castro, F. G., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6,* 151–175. http://dx.doi.org/10.1007/s11121-005-5553-y

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74,* 59–109.

Gage, N. A., MacSuga-Gage, A. S., & Crews, E. (2017). Increasing teachers' use of behavior-specific praise using a multitiered system for professional development. *Journal of Positive Behavior Interventions, 19,* 239–251.

Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 943–954). Washington, DC: National Association of School Psychologists.

Gettinger, M., & Walter, M. J. (2012). Classroom strategies to enhance academic engaged time. In S. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 653–673). Boston, MA: Springer.

Gromada, A., & Shewbridge, C. C. (2016), *Student learning time: A literature review* (OECD Education Working Papers, No. 127). Paris, France: OECD Publishing. http://dx.doi.org/10.1787/5jm409kqqkjh-en

Harcourt Assessment, Inc. (2004). *Stanford Achievement Test series—10th ed. technical data report*. San Antonio, TX: Author.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2,* 172–177. http://dx.doi.org/10.1111/j.1750-8606.2008.00061.x

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27,* 599–641. http://dx.doi.org/10.1023/A:1022137920532

Ingersoll, R. M. (2002, August 15). High turnover plagues schools. *USA Today,* p. 13A.

Institute of Education Sciences. (2017). *What works clearinghouse*. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks#procedures

Institute of Education Sciences. (2020). *Call for applications: CFDA# 84.305A*. Retrieved from https://ies.ed.gov/funding/pdf/2021_84305A.pdf

Johns Hopkins University Best Evidence Encyclopedia. (2019). Empowering educators with evidence on power programs. Retrieved from http://www.bestevidence.org/index.cfm

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10,* 165–185. http://dx.doi.org/10.1017/S0954579498001564

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., . . . Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95,* S5–S28.

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher Observation of Classroom Adaptation-Checklist: Development and factor structure. *Measurement & Evaluation in Counseling & Development, 42,* 15–30. http://dx.doi.org/10.1177/0748175609333560

Kontopantelis, E., White, I. R., Sperrin, M., & Buchan, I. (2017). Outcome-sensitive multiple imputation: a simulation study. *BMC Medical Research Methodology, 17,* 2.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49,* 241–253.

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject related skills. *Learning and Instruction, 17,* 494–509. http://dx.doi.org/10.1016/j.learninstruc.2007.09.002

Ladd, B. O., Tomlinson, K., Myers, M. G., & Anderson, K. G. (2016). Feasibility and reliability of a coding system to capture in-session group behavior in adolescents. *Prevention Science, 17,* 93–101. http://dx.doi.org/10.1007/s11121-015-0588-1

Lekwa, A. J., Reddy, L. A., & Shernoff, E. S. (2019). Measuring teacher practices and student academic engagement: A convergent validity study. *School Psychology, 34,* 109–118.

Lewis, C. G., Herman, K. C., Huang, F. L., Stormont, M., Grossman, C., Eddy, C., & Reinke, W. M. (2017). The utility of single-item readiness screeners in middle school. *Journal of School Psychology, 64,* 1–16. http://dx.doi.org/10.1016/j.jsp.2017.04.003

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., . . . Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013–3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Madigan, K., & Cross, R. W. (2011). *Comparison of two positive behavior supports program models on the rate of change in student academic achievement: A ten-year evaluation*. Unpublished manuscript.

Madigan, K., Cross, R. W., Smolkowski, K., & Strycker, L. A. (2016). Association between schoolwide positive behavioural interventions and supports and academic achievement: A 9-year evaluation. *Educational Research and Evaluation, 22*(7–8), 402–421.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37,* 153–184.

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22,* 276–282. http://dx.doi.org/10.11613/BM.2012.031

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49,* 980–989. http://dx.doi.org/10.1016/j.jaac.2010.05.017

Missouri Department of Elementary and Secondary Education. (2015). Missouri Assessment of Progress test. Jefferson City, MO: DESE.

Monson, K. D., Caldarella, P., Anderson, D. H., & Wills, H. P. (2020). Improving Student Behavior in Middle School Art Classrooms: Initial Investigation of CW-FIT Tier 1. *Journal of Positive Behavior Interventions, 22,* 38–50. http://dx.doi.org/10.1177/1098300719864704

National Commission on Teaching and America's Future. (2007). *The high cost of teacher turnover* (Policy Brief). Retrieved September 5, 2008, from http://www.nctaf.org

National Institute of Child Health and Human Development, Early Child Care Research Network (NICHD ECCRN). (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105,* 305–323. http://dx.doi.org/10.1086/428746

National Research Council and Institute of Medicine (NRC & IOM). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: National Academies Press.

Oliver, R. M., & Reschly, D. J. (2010). Teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders, 35,* 188–199.

Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews, 7,* 1–55. http://dx.doi.org/10.4073/csr.2011.4

Orr, R. K., Caldarella, P., Hansen, B. D., & Wills, H. P. (2020). Managing student behavior in a middle school special education classroom using CW-FIT tier 1. *Journal of Behavioral Education, 29,* 168–187. http://dx.doi.org/10.1007/s10864-019-09325-w

Partin, T. C. M., Robertson, R. E., Maggin, D. M., Oliver, R. M., & Wehby, J. H. (2009). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure, 54,* 172–178. http://dx.doi.org/10.1080/10459880903493179

Pas, E. T., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools, 48,* 541–555. http://dx.doi.org/10.1002/pits.20577

Pavri, S. (2004). General and special education teachers' preparation needs in providing social support: A needs assessment. *Teacher Education and Special Education, 27,* 433–443. http://dx.doi.org/10.1177/088840640402700410

Petras, H., Chilcoat, H., Leaf, P., Ialongo, N., & Kellam, S. (2004). The utility of teacher ratings of aggression during the elementary school years in identifying later violence in adolescent males. *Journal of the American Academy of Child & Adolescent Psychiatry, 1,* 88–96. http://dx.doi.org/10.1097/00004583-200401000-00018

Petras, H., Masyn, K., & Ialongo, N. (2011). The developmental impact of two first grade preventive interventions on aggressive/disruptive behavior in childhood and adolescence: An application of latent transition growth mixture modeling. *Prevention Science, 12,* 300–313. http://dx.doi.org/10.1007/s11121-011-0216-7

Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195327892.003.0002

Pianta, R. C., Hamre, B. K., Hayes, N., Mintz, S., & LaParo, K. M. (2008). *Classroom assessment scoring system - Secondary (CLASS-S)*. Charlottesville, VA: University of Virginia.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28,* 147–169. http://dx.doi.org/10.1023/B:MOEM.0000032312.95499.6f

Reinke, W. M., Herman, K. C., & Dong, N. (2018). The incredible years teacher classroom management program: Outcomes from a group randomized trial. *Prevention Science, 19,* 1043–1054. http://dx.doi.org/10.1007/s11121-018-0932-3

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

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. (2008). Empirically-derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology, 36,* 759–777. http://dx.doi.org/10.1007/s10802-007-9208-2

Reinke, W. M., Stormont, M., Herman, K. C., & Newcomer, L. (2014). Using coaching to support teacher implementation of classroom-based interventions. *Journal of Behavioral Education, 23,* 150–167. http://dx.doi.org/10.1007/s10864-013-9186-0

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26,* 1–13. http://dx.doi.org/10.1037/a0022714

Reinke, W. M., Stormont, M., Herman, K. C., Wachsmuth, S., & Newcomer, L. (2015). The Brief Classroom Interaction Observation–Revised: An observation system to inform and increase teacher use of universal classroom management practices. *Journal of Positive Behavior Interventions, 17,* 159–169. http://dx.doi.org/10.1177/1098300715570640

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys,* New York, NY: John Wiley & Sons, Inc.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33,* 545–571.

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and anti-

social personality disorder. *Developmental Psychology, 39,* 1020–1035. http://dx.doi.org/10.1037/0012-1649.39.6.1020

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

Smith, T. M., & Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American educational research journal, 41,* 681–714.

Spanjers, D. M., Burns, M. K., & Wagner, A. R. (2008). Systematic direct observation of time on task as a measure of student engagement. *Assessment for Effective Intervention, 33,* 120–126. http://dx.doi.org/10.1177/1534508407311407

Speight, R., Whitby, P., & Kucharczyk, S. (2020). Impact of CW-FIT on student and teacher behavior in a middle school. *Journal of Positive Behavior Interventions, 22,* 195–206. http://dx.doi.org/10.1177/1098300720910133

Sprick, R. (2010). *CHAMPS: DVD Inservice* (2nd ed.). Eugene, OR: Pacific Northwest Publishing.

Sprick, R., Garrison, M., & Howard, L. (2009). *CHAMPS: A proactive and positive approach to classroom management* (2nd ed.). Eugene, OR: Pacific Northwest Publishing.

Sprick, R., & Howard, L. (1996). *The teacher's encyclopedia of behavior management: 100 problems/500 plans*. Longmont, CO: Sopris West.

Stormont, M., Reinke, W. M., & Herman, K. C. (2011). Teachers' knowledge of evidence-based interventions and available school resources for children with emotional or behavioral problems. *Journal of Behavioral Education, 20,* 138–147. http://dx.doi.org/10.1007/s10864-011-9122-0

Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., & Coie, J. D. (1999). The relation between behavior problems and peer preference in different classroom contexts. *Child Development, 70,* 169–182. http://dx.doi.org/10.1111/1467-8624.00013

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62,* 339–355. http://dx.doi.org/10.1177/0022487111404241

Tapp, J. (2004). *MOOSES* (Multi-Option Observation System for Experimental Studies). Retrieved from https://mooses.vueinnovations.com/

What Works Clearinghouse. (2014). *Procedures and standards handbook* (Version 3.0). Washington DC: Institute of Education Sciences.

Wills, H. P., Caldarella, P., Mason, B. A., Lappin, A., & Anderson, D. H. (2019). Improving student behavior in middle schools: Results of a classroom management intervention. *Journal of Positive Behavior Interventions, 21,* 213–227. http://dx.doi.org/10.1177/1098300719857185

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods, 12,* 695–719. http://dx.doi.org/10.1177/1094428108327450

Zima, B. T., Hurlburt, M. S., Knapp, P., Ladd, H., Tang, L., Duan, N., . . . Wells, K. B. (2005). Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 130–144. http://dx.doi.org/10.1097/00004583-200502000-00005

Received February 5, 2020
Revision received September 1, 2020
Accepted September 7, 2020 ■