Journal of School Psychology 92 (2022) 227–245

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

A randomized trial examining the effects of paraprofessional behavior support coaching for elementary students with disruptive behavior disorders: Paraprofessional and student outcomes☆

Linda A. Reddy *, Todd A. Glover, Christopher M. Dudek, Alexander Alperin, Nicole B. Wiggs, Briana Bronstein

Rutgers University, The State University of New Jersey, USA

ARTICLE INFO

Editor: Craig A. Albers

Keywords:
Paraprofessionals
Coaching
Professional development
Behavior interventions
Externalizing behaviors

ABSTRACT

There is limited research on job-embedded professional development designed to promote paraprofessionals' use of research-based strategies to support students with disruptive behaviors. This study serves as the first clustered randomized controlled trial to investigate the efficacy of the Behavior Support Coaching for Paraprofessional Model (BSC-P), which is a job-embedded training intervention for elementary schools. BSC-P is a data-driven coaching model designed to enhance paraprofessionals' implementation of behavior interventions for elementary school students with or at risk for disruptive behavior disorders. Within the present study, primary dependent measures included paraprofessionals' behavioral strategy implementation and perceived supports, the paraprofessional-teacher relationship, and work-related stress, as well as observed and rated student behaviors, academic engagement, academic achievement, and social skills. The sample included 259 students, 101 paraprofessionals, and 36 elementary schools randomly assigned to the BSC-P coaching condition or the waitlist control condition. Multilevel models revealed that, relative to those in the waitlist control condition, BSC-P paraprofessionals demonstrated improvements in behavior management practices (antecedent strategies $d = 0.91$, reward appropriate behavior $d = 1.51$) and emotional and instrumental support ($ds = 0.60$ and 0.63, respectively). No between-condition effects were found for perceived teacher-paraprofessional relationships or work-related stress. Relative to students in the waitlist control condition, students supported by BSC-P paraprofessionals exhibited improvements in observed verbal and physical aggression ($ds = -0.68$) and academic engagement ($d = 0.87$), as well as teacher-rated school problems ($d = -0.43$), adaptive skills ($d = 0.44$), and social skills ($d = 0.42$). Paraprofessionals reported that BSC-P was an acceptable and useful professional development model.

Professional development is key for adequately preparing school personnel to support students with significant behavioral

☆ The current study was implemented as part of a US Department of Education – Institute of Education Sciences NCSER efficacy project (awarded to Rutgers, The State University of New Jersey, #R324A170069). The positions and opinions expressed in this article are solely those of the authors.

* Corresponding author at: Graduate School of Applied and Professional Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA.

E-mail address: LReddy@rutgers.edu (L.A. Reddy).

https://doi.org/10.1016/j.jsp.2022.04.002
Received 27 February 2021; Received in revised form 11 December 2021; Accepted 5 April 2022
Available online 25 April 2022
0022-4405/© 2022 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

concerns, including students with or at risk for disruptive behavior disorders (DBDs). Although paraprofessionals are among those most likely to be assigned to students with challenging and complex behavioral needs, they receive very limited training, if any, to prepare them for this important responsibility (e.g., Ashbaker & Morgan, 2006; Giangreco et al., 2013). Furthermore, very few studies have investigated the efficacy of the limited professional development approaches available to these individuals (e.g., Reddy, Alperin, & Glover, 2021; Walker & Smith, 2015). Additional research is needed to identify methods with demonstrated efficacy for preparing paraprofessionals to adequately address students' significant behavioral needs.

1. Need to support students with DBDs

Students with or at risk for DBDs are vulnerable to long-term social, behavioral, and academic difficulties. DBDs are the most common reason for referrals to school teams and mental health clinics, utilizing approximately 30% of special education services in schools nationwide (Allen, 2016; Pikard et al., 2018). Students with DBDs who receive special education services often are classified under Other Health Impairment or Emotional Disturbance (Reddy et al., 2013). These students tend to have complex behavioral, emotional, and academic needs and are often at risk for short- and long-term negative outcomes. A significant body of research has underscored that without adequate behavior interventions and supports, they are at high risk for academic failure, school dropout, family and peer difficulties, driving accidents, teenage pregnancy, drug use, and suicide (e.g., Barker et al., 2010; Delligatti et al., 2003; Reddy et al., 2009). Thus, early behavior interventions and supports offered in school are critical to curtailing negative outcomes in this population.

In addition, students with DBDs often adversely impact classroom teachers' instructional delivery and their classmates' ability to profit from the learning environment. In a recent survey, teachers reported, on average, losing 2.5 h of instruction each week due to disruptive classroom behaviors (Education Advisory Board, 2019), which also is consistent with a national poll conducted by the American Federation of Teachers (Walker et al., 2004). Lost instructional time often has significant effects on learning and social behavior for students. Similarly, disruptive behaviors have adverse consequences on teachers' sense of safety, job retention, and mental and physical wellbeing (e.g., Bottiani et al., 2019; Evers etal., 2004; Reddy et al., 2018). Taken together, students with DBDs represent critical educational and mental health concerns for schools (Alperin et al., 2020). It is essential that research-based professional development is available for teachers and paraprofessionals to assist in the implementation of school-based interventions for meeting the needs of this vulnerable population (e.g., Eyberg et al., 2008; Reddy, Alperin, & Glover, 2021).

2. Paraprofessionals as important classroom support agents

Paraprofessionals are one of the most prevalent providers of classroom-based behavior interventions and strategies and offer assistance for reinforcing learning and social activities led by classroom teachers (e.g., Chopra et al., 2011; Fisher & Pleasants, 2012; Giangreco et al., 2013; Giangreco & Broer, 2005). Although paraprofessional roles vary (e.g., assemble learning materials, promote peer interaction, manage behavior, supplemental instruction; Sobek, 2016) in school contexts, they are nonetheless vital to supporting students and classroom teachers. For example, paraprofessionals provide instructional and behavioral supports, either individually, in small groups, or periodically class-wide, to maximize engagement and learning for students with greater behavioral and/or academic learning difficulties. In some instances, under the direction of teachers, paraprofessionals provide accommodations for academics (e.g., Riggs & Mueller, 2001) or social behavior (e.g., Fisher & Pleasants, 2012) to promote students' on-task behavior and access to learning activities. When provided with adequate training and supervision, paraprofessionals can implement school-based behavior interventions and accommodations that can lead to reductions in classroom disruptive behavior, increases in on-task behavior, and improvements in learning and social development (e.g., Alperin et al., 2020; Giangreco et al., 2013; Penno et al., 2000; Wiggs et al., 2020).

Recognition of paraprofessionals' contributions to the classroom environment is evidenced by increased hiring rates and legislation. Approximately one million paraprofessionals are employed by schools in the United States and hiring rates are currently higher than special education teachers in over 40 states in the U.S (e.g., Mrachko and Kaczmarek, 2017; U.S. Department of Education [USDOE], National Center for Education Statistics, 2017). Likewise, the 2004 Individuals with Disabilities Education Improvement Act (IDEIA, section 1412) requires districts to train and supervise paraprofessionals and the Every Student Succeeds Act (ESSA, 2015) encourages districts to designate funds for ongoing training and professional development. Specifically, ESSA (SEC 1112.20 U.SC. 6312) highlights the use of funds for timely, meaningful, and sustained school consultation/professional development for paraprofessionals and other school personnel (e.g., teachers, school leaders).

Despite these noteworthy developments, paraprofessionals often receive limited training opportunities to improve their knowledge and skills, particularly in regard to managing student behavior (e.g., Ashbaker & Morgan, 2006; Giangreco et al., 2013; Reddy, Alperin, & Glover, 2021). Furthermore, given an absence in training and ongoing support, paraprofessionals are vulnerable to strained relationships with classrooms teachers and are at high risk for stress and burnout. This gap in the provision of paraprofessional training and ongoing support is concerning given their role in addressing the substantial academic and social behavior risks associated with DBDs. Job-embedded professional development, such as school-based instructional coaching, is vital for paraprofessionals to meet the needs of this vulnerable student population.

Most scholars and practitioners view coaching as a job-embedded, individualized data-driven and sustained practice that targets instructional needs, formulates goal-based plans, modeling, opportunities for practice, and ongoing feedback to enhance the fidelity of practices in school settings (e.g., Denton & Hasbrouck, 2009; Glover, Reddy, & Crouse, n.d; Kurz et al., 2017; Reddy, Shernoff, & Lekwa, 2021). Erchul (2015) eloquently differentiated coaching from consultation by highlighting that coaching is often designed to

support educators' continuous professional development. Although coaching and consultation models vary in their implementation, these approaches have typically differed with respect to who holds power/authority. Traditionally, consultants have maintained an expert orientation providing advice and directives to consultees, whereas coaches have focused on supporting coachees' growth and professional development. Consultation is typically more time limited. Consistent with this distinction, many coaches are employed in schools and were former teachers, whereas consultants have specialized training in psychology and/or other fields (e.g., occupational therapy, behavior interventions) and are often employed outside of the school.

3. Need for school-based coaching for paraprofessionals

The effectiveness of paraprofessional professional development has been significantly understudied. Existing research has almost exclusively focused on classroom settings for students with severe developmental disabilities, with many studies demonstrating poor methodological quality (Walker & Smith, 2015). In a recent review of the efficacy of paraprofessional training, Reddy, Alperin, and Glover (2021) found only 16 investigations that assessed the efficacy of professional development approaches for paraprofessionals serving students with or at risk for DBDs. Most studies used single-case design (with the majority not utilizing What Works Clearinghouse standards; Kratochwill et al., 2010) or pre/post design, with no control designs to examine professional development methods (e.g., one-on-one teaching sessions, workshop-based group lectures). No study in this review utilized an experimental design or assessed the implementation of job-embedded professional development models such as instructional coaching. Although a recent study by Sobeck et al. (2020), utilizing an alternating treatment design to investigate the impact of didactic instruction (e.g., behavior specific praise, opportunities to respond, effective instruction and commands) and performance feedback found effects on paraprofessional's behavioral strategy use, no data were collected on outcomes for students. Given an absence of research, there is an urgent need for investigations evaluating evidence-based job-embedded professional development models such as coaching models designed for paraprofessionals working with students with or at risk of DBDs.

Instructional coaching has emerged as a promising approach for supporting school personnel to meet their students' needs. Prior research has demonstrated the effectiveness of job-embedded coaching in enhancing teachers' implementation of research-based interventions, as well as improving student learning and behavior difficulties (e.g., Kraft et al., 2018; Kretlow & Bartholomew, 2010; Pianta et al., 2022). In a meta-analysis of 60 school-based coaching randomized controlled trials (RCTs; Kraft et al., 2018), coaching yielded positive effects on teacher practices and student outcomes, with pooled effect sizes of 0.49 *SD* and 0.18 *SD*, respectively). However, the majority of the investigations in this meta-analysis focused on literacy coaching for teachers in elementary school and no study tested the effects of coaching on paraprofessional practices and student outcomes. Likewise, consultation models have been found to be efficacious in large-scale RCTs at improving student outcomes (e.g., Sheridan et al., 2012, 2017). For example, Sheridan et al. (2017) found that, relative to control students, students of teachers assigned to a consultation condition had greater improvements in observed behaviors (e.g., Cohen's *d* ranging from 0.28 to 0.46) and greater reductions in teacher-rated negative behaviors (e.g., Cohen's *d* ranging from −0.38 to −0.45). Despite evidence supporting the efficacy of instructional coaching and

Table 1
BSC-P model phases, sessions, and objectives.

Phases	Sessions	Objectives
Identify Behavior Needs	1	The coach reviews the goals of the coaching model with the paraprofessional and classroom teacher. Coaches were trained to be paraprofessional-focused, using relationship alliance skills and motivational interview strategies to recognition and motivate paraprofessional coaching engagement throughout coaching. Teachers are invited to attend all sessions but encouraged to at minimum be present for sessions 1–3 and 8. Coach discusses the importance of using data to identify student behavior needs and match interventions to student needs. Participants discuss individual student behavior needs and strengths based on data sources (observations, anecdotal reports, work products). The session concludes with identification of one to two high frequency behaviors. Possible functions of behavior are initially discussed.
Confirm Behavior Needs and Set Goals	2	Student behavior needs are reconfirmed. SMART goals are formulated using the coach's observational data and anecdotal teacher and paraprofessional information. Behavior functions are reviewed and identified for each student. Two to three evidence-based interventions were selected and discussed from the behavior intervention toolkit.
Select and Prepare for Intervention Implementation	3	The coach, paraprofessional and teacher meet to reconfirm student goals, set criterion, select intervention(s) from toolkit and prepare for intervention implementation for each target student. Interventions are chosen to fit the function of student behaviors. Intervention fidelity checklists are reviewed, coach model's intervention steps, creates opportunities for practice, and offers performance feedback on implementation. Possible barriers and opportunities for implementation are discussed.
Monitor and Support Implementation	4–7	Sessions focused on supporting intervention implementation and monitoring of implementation fidelity and goal progress through (1) the coach's observation data, (2) intervention fidelity checklists (including graphed percent of fidelity), (3) paraprofessional strategy logs and self-reflection, and (4) the coach's encouragement and feedback. Modeling, practice (e.g., role playing) and performance feedback are provided to enhance paraprofessional implementation.
Evaluate Implementation & Goal Progress	8	The coach, teacher, and paraprofessional meet to discuss and evaluate the coaching process. Data on intervention implementation and goal progress are graphed and discussed. Examples of success are highlighted via self-reflection; discuss plans for generalization of strategies to other contexts are reviewed. Effectiveness of interventions are determined and plans for the future are discussed such as discontinue, continue, or adjust interventions and supports.

consultation approaches, the utility of coaching has not been evaluated with paraprofessionals who support individual student's behavior needs in classroom settings. The present investigation aimed to address this gap in the literature by rigorously examining the efficacy of a unique instructional coaching model (i.e., Behavior Support Coaching for Paraprofessionals [BSC-P]) on paraprofessionals' behavior practices and student outcomes for elementary school students with or at risk for DBDs.

4. Behavior support coaching for paraprofessionals

The Behavior Support Coaching for Paraprofessionals (BSC-P) coaching model is provided to paraprofessionals, with participation from classroom teachers, to enhance knowledge and implementation of research-based behavior management practices to improve elementary school (i.e., Grades K–5) students' behavior and learning in the classroom (see Table 1). BSC-P is data-driven and grounded in research on teacher coaching and behavioral consultation (e.g., Bergan & Kratochwill, 1990; Kraft et al., 2018; Reddy, Shernoff, & Lekwa, 2021; Sheridan et al., 2012). The BSC-P components and processes are supported by other large-scale RCTs focused on coaching teachers to improve their use of classroom practices to address student reading difficulties and disruptive behaviors (i.e., Reddy, Lekwa, & Shernoff, 2021; Reddy, Shernoff, & Lekwa, 2021; Sheridan et al., 2012). BSC-P embraces a child-centered and solution-oriented framework and utilizes a five-phase problem solving process, with coaching provided during eight, 45 min sessions (see Table 1).

BSC-P includes five phases and involves supporting paraprofessionals in using observational assessment data to (a) identify student needs and the conditions responsible for those needs [1 - *needs identification* and *needs analysis*]; (b) set specific, observable, and measurable goals [2 - *goal setting*]; and (c) select research-based interventions matched to individual students' needs from a toolkit [3 - *plan development*]. Coaches then support paraprofessionals during (4) *intervention implementation* via modeling, practice, encouragement, observation, and performance feedback. Through the continued use of observational data, the paraprofessional and coach monitor and evaluate intervention implementation and students' progress in meeting goals [5 - *evaluation of progress*]. Throughout the coaching process, paraprofessionals are afforded ongoing encouragement and support for identifying and addressing a wide range of student behavior strengths and needs in classrooms.

The BSC-P model has several unique features that build upon existing coaching/consultation models. First, the model is designed specifically for empowering paraprofessionals to use data to support students with or at risk for DBDs. Second, BSC-P is conducted in authentic classroom settings where coaches and paraprofessionals, along with the periodic involvement of classroom teachers, engage in data-based reflection and decision-making for selecting and implementing interventions matched to student needs. Third, coaching for paraprofessionals targets the identification of common functions of behavior (e.g., skill deficit, attention seeking, escape), the use of a toolkit of research-based behavior interventions, and the use of observational data to drive decisions about adaptations to behavioral strategy implementation. Unlike traditional behavioral consultation approaches, the session protocols for the BSC-P define very specific steps for guiding paraprofessionals in observing students and using cycles of observational data to analyze the functions of their behaviors, evaluate student progress, and make adjustments to behavioral intervention planning (see Table 1 for details).

5. Purpose of study

The primary purpose of this investigation was to examine the efficacy of the BSC-P for enhancing (a) paraprofessional's implementation of behavioral strategies; (b) perceptions about supports, the paraprofessional-teacher relationship, and work-related stress; and (c) student behavioral outcomes. Specific research questions were:

1. What is the effect of BSC-P on the implementation of paraprofessionals' behavior strategy implementation? Based on improvements in coached educators' behavioral management practices from previous research on instructional coaching and behavior consultation (Fabiano et al., 2018; Reddy, Shernoff, & Lekwa, 2021; Sheridan et al., 2012), it was hypothesized that coached paraprofessionals would have greater use of antecedent and consequent strategies than waitlist (WL) control participants.
2. What are the effects of BSC-P on paraprofessionals' perceived supports, paraprofessional-teacher relationship, and work-related stress? Based on findings from previous RCT investigations of behavioral support coaching/consultation with teachers and parents, it was hypothesized that BSC-P paraprofessionals would report greater perceived instrumental and emotional supports and paraprofessional-teacher relationships than WL participants (Reddy, Shernoff, & Lekwa, 2021; Sheridan et al., 2017).
3. What are the effects of BSC-P on the behavior, social, and academic outcomes for students with or at risk for DBDs? It was hypothesized that students assigned to BSC-P paraprofessionals would have greater improvements in their behavior, academic engagement, and social skills than those assigned to WL paraprofessionals. Because BSC-P focused on students' challenging behaviors, the impact of more distal academic performance was unknown.

Finally, we examined paraprofessional and teacher perceived acceptability, usability, and feasibility of BSC-P as a professional development intervention in elementary schools.

6. Method

6.1. Participants

Data for the present study were obtained from 101 paraprofessionals and 98 classroom teacher partners that taught 259 unique

students with or at risk of DBDs. A small number of classrooms included more than one paraprofessional. All paraprofessional, teacher, and student pairings were in elementary school (Grades K–5) and came from 36 schools in a northeast state in the US.

All schools were in urban or suburban settings with 60.49% being the average percent (*SD* = 26.00%, range = 5.60%–95.50%) of students receiving free and reduced lunch. Accordingly, the 36 schools participating in this study served a diverse student body composed of approximately 22.14% White, 39.63% Hispanic/Latinx, 30.00% Black, 6.08% Asian, and 1.70% multi-racial students. All participants in this RCT study were volunteers consented to participate according to the research institution's approved Institutional Review Board IRB procedures.

6.1.1. Paraprofessionals

Paraprofessionals (n = 101) participating in the study were on average 45.40 years old (*SD* = 13.37), with a range of 23–69 years. Most paraprofessionals self-identified as female (94.73%) and reported their race/ethnicity as Black (21.53%), White (23.41%), or Hispanic/Latinx (19.95%). Approximately 3% self-identified as Asian and 4% as multiple races/ethnicities. Paraprofessionals' years of experience varied between as little as 1 year of experience to more than 10 years of experience, with the majority of paraprofessionals being more experienced in their role; 10.76% possessed fewer than 2 years of experience as a paraprofessional, whereas 20.58% possessed between 2 and 5 years, 8.79% possessed between 5 and 10 years, and 34.41% possessed over 10 years of experience. Similarly, education level was varied, with approximately 29% of participants indicating they possessed some college education but no degree conferral. Approximately 18% possessed a high school diploma, whereas 8% possessed an Associate's degree, and 16% possessed a 4-year college degree or greater. Table 2 presents demographics by participant and condition including data not reported.

6.1.2. Teachers

Teachers' (n = 98) ages ranged from 24 to 69 years old, with an average age of 39.92 years (*SD* = 11.75). Teachers predominantly self-identified as female (93.33%) and either White (50%), Black (11.38%), Hispanic/Latinx (7.44%), or as having multiple race/ethnicities (4.71%). Teachers reported varying levels of experience with most teachers (36.98%) reporting greater than 10 years of experience. Approximately 12% of teachers had less than 2 years of experience, 8% possessed between 2 and 5 years of experience, and 14% possessed between 5 and 10 years of experience. Teachers' education level was less varied than that of paraprofessionals, with most teachers (73%) indicating they possessed a 4-year college degree or greater. Table 2 presents demographics by participant and condition, including data not reported here.

Table 2
Demographics of paraprofessional and teacher participants by condition.

	Paraprofessionals		Teachers	
	Coaching (n = 53)	Waitlist (n = 48)	Coaching (n = 53)	Waitlist (n = 45)
Age	M = 45.81	M = 44.97	M = 40.11	M = 39.71
	SD = 12.62	SD = 14.28	SD = 12.55	SD = 10.99
Not Reported	30.19%	29.17%	30.19%	24.44%
Gender				
Female	94.33%	97.91%	100%	86.66%
Male	5.77%	2.09%	0.00%	13.44%
Race/Ethnicity				
Hispanic/Latino	11.32%	28.57%	3.77%	11.11%
Black	22.64%	20.41%	9.43%	13.33%
White	26.42%	20.41%	49.06%	51.11%
American Indian/Alaskan Native	0.00%	0.00%	0.00%	0.00%
Asian	5.66%	0.00%	1.89%	0.00%
Native Hawaiian / Pacific Islander	0.00%	0.00%	0.00%	0.00%
Middle Eastern	0.00%	0.00%	1.89%	2.22%
Other	1.89%	0.00%	0.00%	0.00%
Multiple	3.77%	4.08%	9.43%	0.00%
Not Reported	28.30%	26.53%	24.53%	22.22%
Education				
High School Diploma	9.43%	26.53%	0.00%	4.44%
Some college: non-degree	26.42%	30.61%	0.00%	0.00%
Associate's degree	11.32%	4.08%	0.00%	2.22%
Bachelor's degree	22.64%	8.16%	37.74%	33.33%
Graduate degree	1.89%	0.00%	37.74%	37.78%
Not Reported	28.30%	30.61%	24.53%	22.22%
Years of Experience				
0–2	11.32%	10.20%	11.32%	13.33%
2–5	20.75%	20.41%	9.43%	6.67%
5–10	9.43%	8.16%	15.09%	13.33%
More than 10	32.08%	36.73%	33.96%	40.00%
Not Reported	26.42%	24.49%	30.19%	26.67%

6.1.3. Students

Students (n = 259) were on average 7.63 years old (*SD* = 1.89) and most were identified by their classroom teacher as male (77.07%). Grade level distribution of the students was 58 students in kindergarten, 41 students in first grade, 27 students in second grade, 47 students in third grade, 43 students in fourth grade, 33 students in fifth grade, and 10 students in multi-grade classrooms. Other individual student demographics were not obtained due to participating school district procedures requiring signed consent documents from parent/guardians for the release of this sensitive student data. The individual student demographics that were collected were data points observable and reportable by the classroom teacher or study staff (i.e., sex assigned at birth, grade level, and age).

6.2. Research design

A two-cohort, cluster randomized, experimental design was used to evaluate the impact of BSC-P on (a) paraprofessionals' implementation of behavioral strategies; (b) paraprofessionals' perceptions about supports, the paraprofessional-teacher relationship, and work-related stress; and (c) student behavioral outcomes. Given the importance of within-school consistency in paraprofessional support practices, a total of 36 schools were randomly assigned to one of two conditions: a business-as-usual waitlist (WL) control condition or BSC-P intervention (experimental condition).

6.3. Measures

6.3.1. Behavioral and emotional screening system, third edition

The Behavioral and Emotional Screening System (BESS-3; Kamphaus & Reynolds, 2015) Teacher Form (TF) was used as a screening measure to assess students for project eligibility and was administered as part of the screening phase of the study. The BESS-3 TF is a brief emotional and behavioral screener for children in Grades K–12. The BESS-3 is a 20-item instrument that is scored using a 4-point Likert scale (0 = *Never*, 1 = *Sometimes*, 2 = *Often*, and 3 = *Almost Always*), in which higher scores indicate a higher risk for behavior and emotional difficulties. The BESS-3 TF produces a summed score of item ratings, which is then transformed into a t-score ($M = 50$, $SD = 10$). T-scores are used to classify children's risk status for behavioral and emotional problems into three categories: normal ($t < 59$), elevated risk ($t = 60$–69) or extreme risk ($t \geq 70$). In the present study, participating students scored in the elevated to extreme risk ranges. High levels of split-half reliability (0.96), test-retest reliability (0.91), and acceptable internal consistency (0.83) have been

Table 3
Modified BOSS behavior code definitions.

Behavior	Operational Definitions	Examples
Active Engagement	Times when the student is actively attending to the assigned work; on-task and actively participating	Writing; reading aloud; raising a hand to answer a teacher's question; talking to the teacher about the assigned material; talk to a peer about the assigned material; looking up a word in a dictionary; typing essay on computer
Passive Engagement	Times when the student is passively attending to the assigned work; on task and passively participating	Listening to a lecture; looking at an academic worksheet; reading assigned material silently; looking at the blackboard during teacher instruction listening to a peer respond to a question, watching a video, viewing lesson on the smartboard
Inappropriate Physical Behavior	A forceful movement directed at another person, either directly or by utilizing a material object as an extension of the hand	Hitting, biting, kicking, pinching, raising a clenched fist or open hand and swinging toward another, throwing objects toward another person
	A forceful movement directed at an inanimate object or inflicts physical damage on an object	Damaging property (tearing papers) or objects (ripping book bags); kicking a bookcase; throwing an object forcefully onto the floor
	Physical behavior that interferes with or disrupts classroom functioning and/or makes it difficult for others to perform their work	Taking objects from others without asking permission; excessive or inappropriate motor activity either in or out of his/her seat; out of seat; lack of body control; moving around the room; leaving an assigned area and/or leaving the group, tapping peers, making inappropriate gestures, drawing on assignment, playing with objects at desk
Inappropriate Verbal Behavior	Verbalizations that are abusive or threatening and directed toward other people; all negative, noncontact communication	Verbal bullying, verbal threats, tattling, teasing, name-calling; cursing
	Verbal behavior that interferes with or disrupts classroom functioning and/or makes it difficult for others to perform their work	Initiating conversations during quiet work periods; calling out in class; making noises that distract others (e.g., animal noises, grunts); verbal interruption; talk outs; inappropriate elevated voice level; lack of emotional control (e.g., crying)
Non-Compliance	If (a) the child is *given an instruction/direction* by an aide or teacher and does not comply with the request by the end of the next observation interval (i.e., within 15 s), or (b) the *teacher reprimands the child or reminds* them to follow earlier instructions	Not following directions; not completing or starting assigned work
Disruptive Academic Behavior	For at least 3 consecutive seconds: Attending to any stimulus or activity other than the one assigned; not directing focus toward the teacher during presentation of a lesson; "does nothing"	Off-task; looking out of the window when directed to complete work, looking at teacher while she is talking to another student

Note. BOSS behavior codes and definitions were adapted and modified from Shapiro (2011) and Sheridan et al. (2012).

demonstrated for the BESS-3 TF and convergent validity has been demonstrated with other teacher behavior rating scales and report-card outcomes (Kamphaus & Reynolds, 2007; Renshaw et al., 2009).

6.3.2. Behavioral observation of students in schools

The Behavioral Observation of Students in Schools (BOSS; Shapiro, 1996, 2011; Shapiro & Heick, 2004) was used as an outcome measure for student behavior and was administered during the baseline and post-assessment phases of the study. The BOSS captured behavioral outcome data on students' academic engagement and disruptive classroom behaviors as measured by both independent observers and coaches. The BOSS utilizes an interval time-sampling procedure to measure counts of individual students' levels of on-task (i.e., engagement) and off-task behaviors (i.e., disruptive behaviors). Behavior codes were modified for purposes of this RCT to meet the project objectives and the range of student behavior needs (see Table 3) and included (a) active engagement (AE; e.g., answering a teacher's question), (b) passive engagement (PE; e.g., listening to the teacher), (c) inappropriate physical (IP; e.g., throwing object, hitting paraprofessional), (d) inappropriate verbal (IV; e.g., using profanity, call out inappropriately), (e) non-compliance (NC; e.g., refusing to follow a directive), and (f) disruptive academic behaviors (DA; e.g., off-task).

The interval time-sampling procedure was not modified in the current study. As per the original procedure, 15-s interval and partial interval momentary time sampling were used to capture the presence of student on-task and off-task behaviors. BOSS observations were each 15 min in duration, equating to a total of 60 intervals measured. Using an in-the-ear audio sound device, observers received an audio prompt at 15-s intervals (i.e., 60 prompts), and immediately following the prompts, observers coded the identified student's behavior as either AE or PE. Partial interval recording was conducted for the remainder of each interval to code for observed disruptive behaviors (i.e., IP, IV, NC, and/or DA). Following individual observations, scores for each behavior category were calculated. For each observed behavior, the total number of intervals observed was summed together and then divided by the total number of possible intervals (i.e., 60) to represent the percentage of time students engaged in each behavior. For the purposes of the present study, total engagement (TE) scores were created by combining AE and PE into one code. Furthermore, observations of student behavior problems and academic engagement conducted by independent observers were averaged together to compute baseline and post-assessment BOSS index scores.

A small number of studies report psychometric information for the original BOSS (Shapiro, 1996, 2011; Shapiro & Heick, 2004). Strong inter-observer agreement (Sheridan et al., 2012; Volpe et al., 2005; i.e., 90% to 100%) and kappa coefficients (DuPaul et al., 2004; Kappas = 0.93–0.98) have been recorded by investigators. Additionally, discriminant validity (DuPaul et al., 2004), as well as treatment sensitivity (Ota & DuPaul, 2002) for students with ADHD have been demonstrated for the BOSS. In the present study, the modified BOSS inter-observer reliability was strong with correlations ranging from .92 to 0.80 for the six modified behavior code categories (see Procedures section). Specifically, inter-observer reliability for each coded behavior were acceptable (i.e., AE = 0.88; PE = 0.88, IP = 0.92, IV = 0.86, NC = 0.80, and DA = 0.86). Relatedly, the current RCT evidenced convergent validity between the modified BOSS and the Behavioral Assessment System for Children-3 (BASC-3) screener assessment and full assessment (Alperin et al., under review). Similarly, convergent validity between the modified BOSS and Social Skills Improvement System (SSIS) was found in the current RCT (Dudek et al., under review).

6.3.3. Woodcock-Johnson IV test of achievement

The Woodcock-Johnson Tests of Achievement, Fourth Edition (WJ-IV ACH; Schrank et al., 2014, Woodcock-Johnson, 2022) was used as an outcome measure to assess changes in students' academic skills at baseline and post-assessment. The WJ-IV ACH is a norm-referenced, standardized test of academic achievement. For this study, the WJ-IV Brief Achievement cluster (WJ-IV BA), an academic proficiency cluster that includes three subtests (i.e., Letter-Word Identification, Applied Problems, and Spelling), was used and administered individually in pencil-and-paper format. Internal consistency for cluster scores has been reported as high (0.92–0.97), with test-retest correlations (0.83–0.95) as being in the acceptable to excellent range (Villarreal, 2015).

6.3.4. Behavior assessment system for children-third edition teacher rating scale

The Behavior Assessment System for Children (3rd ed.) Teacher Rating Scale-Child Form (BASC-3 TRS-C; Kamphaus & Reynolds, 2015), which was designed for students 6–11 years of age, was used as an outcome measure to obtain teachers' perceptions of changes in students' behavioral and emotional problems in the classroom at baseline and post-assessment. The BASC-3 TRS-C includes 156 items that contain descriptions of observable positive and negative behaviors that a student may display in the classroom. The 156 items are organized into five composite scores: (1) Adaptive Skills measures how students develop positive social behaviors and interact with peers; (2) Behavioral Symptoms Index measures hyperactivity, aggression, depression, attention problems, atypicality, and withdrawal; (3) Externalizing Problems measures the presence of aggressive, hyperactive, and inattentive behaviors in the classroom; (4) Internalizing Problems measures observed students' feelings of anxiety, worry, and stress; and (5) School Problems measures students' ability to attend and interact in the school environment. Teacher raters respond to the frequency of observed behavior items through a 4-point Likert scale (N = *Never*, S = *Sometimes,* O = *Often*, A = *Almost Always*). Each scale point is then converted to a numeric equivalent of 0, 1, 2, or 3 and items are summed according to the representative scale. Raw scores for each scale are then converted to a t-score based on the norming samples governing the BASC-3 development. The BASC-3 TRS composite scales have strong internal consistency estimates (0.92–0.97), test-retest reliabilities (0.77–0.91, *Mdn* = 0.87), and inter-rater reliabilities (0.37–0.73, *Mdn* = 0.68). The BASC-3 has demonstrated convergent correlations with related student behavior rating scales (e.g., Achenbach System of Empirically Based Assessment, Achenbach et al., 2001; Conners 3rd Edition, Conners, 2008).

6.3.5. Social skills improvement system–rating scale teacher form

The Social Skills Improvement System–Rating Scale (SSIS-RS; Gresham & Elliott, 2008; Gresham et al., 2011) Teacher Form was used as an outcome measure to assess teachers' perceptions of changes in students' social skills at baseline and post-assessment. The SSIS-RS is a behavior rating scale measuring teachers' perceptions of student prosocial behavior and skills, problematic behaviors, and academic competence in the classroom. The present study used the SSIS-RS Teacher Form designed for Grades K–5. The SSIS-RS Teacher Form includes 83 rating scale items organized into the three larger domain areas of Social Skills (46 questions), Problem Behaviors (30 questions), and Academic Competence (7 questions). Items are rated by the teacher on a 4-point Likert scale (0 = *Never*, 1 = *Seldom*, 2 = *Often*, 3 = *Almost Always*) that measures the frequency of the observed behavior or skill. In addition, a second frequency scale uses a 3-point Importance rating (0 = *Not Important*, 1 = *Important*, 2 = *Critical*) to identify behaviors requiring immediate intervention. In the current study, the Social Skills composite score was used. Items are summed respectively to create raw scale scores which are then corrected for missing items and adjusted into standard scores ($M = 100$, $SD = 15$). The SSIS-RS has adequate internal consistency estimates (above 0.90) and test-retest reliabilities (0.82–0.92).

6.3.6. Paraprofessional demographics questionnaire

A brief demographics questionnaire was provided to participating paraprofessionals and their partner classroom teacher to obtain relevant participant demographic information. The demographics questionnaire contained questions about their age, gender, race/ethnicity, years of experience, and classroom setting. Additionally, the questionnaire gathered information on paraprofessionals' and classroom teachers' quantity of PD and training hours, along with their level of knowledge and need for additional training in behavior management principles. Paraprofessionals in the WL control and BSC-P conditions reported receiving mostly between 2 and 4 h of professional development. No differences between ratings of baseline knowledge and need for training in behavior management strategies were found between groups. Both groups reported moderate levels of knowledge and need for training in behavior management (Wiggs et al., 2021).

6.3.7. Paraprofessional and teacher relationship scale

The Paraprofessional and Teacher Relationship Scale (PTRS) was used as an outcome measure to assess changes in the paraprofessional-teacher relationship at baseline and post-assessment. The PTRS is a modified version of the Parent-Teacher Relationship Scale (Vickers & Minke, 1995), which included 24-items and results in a total score and two subscales (Joining – 15 items; Communication – 9 items). The PTRS is a 24-item Likert-type scale that assesses the quality of paraprofessional and teacher relationship. Paraprofessionals rate their level of agreement on a 5-point scale (1 = *Almost Never, 2 = Once in a While, 3 = Sometimes, 4 = Frequently, and 5 = Almost Always*). In the present study, the PTRS Total, Joining subscale, and Communication subscale had strong internal consistency (Cronbach alphas of 0.90, 0.88, and 0.87, respectively). The PTRS scales are theoretically and factor analytically derived (confirmatory factor analyses using diagonally weighted least squares) yielding fair to acceptable fit to the data (e.g., X^2/df, goodness-of-fit index, root mean square error of approximation; Reddy, Shernoff, & Lekwa, 2021). The PTRS Total scores are summed across the 24 item ratings, and the Joining and Communication subscale scores are summed across the nested item ratings (i.e., 15 and 9 items, respectively).

6.3.8. Behavior strategy logs

Behavior Strategy Logs (BXL) were used as an outcome measure to monitor paraprofessionals behavioral strategy use in both BSC-P and WL conditions throughout the duration of the study. The BXL is a self-report retrospective log that was developed as part of the present study to track students' behavioral incidents and paraprofessional strategy use. When a disruptive behavior incident occurred in the classroom for any of the target students, paraprofessionals were instructed to complete one BXL form corresponding to the specific incident. The BXL form contains questions about the date, time, classroom, and target student followed by prompts to provide a description of the behavioral incident that occurred and identify the type of behavioral strategy used.

Paraprofessionals were asked to select from a list of Antecedent Strategies (e.g., moving students' seat ahead of time, student self-monitoring; $\alpha = 0.79$) and Consequence Strategies (e.g., ignoring target behavior, providing a time out, enforcing classroom rules, removing token/privilege; $\alpha = 0.68$). BXL scores represent a sum score of the frequency of practices reported by paraprofessionals. All logs were completed retrospectively by assessing strategies paraprofessionals used before the behavioral incidence occurred (i.e., antecedent strategy) or after the behavior incidence (i.e., consequence strategy).

6.3.9. Paraprofessional stress and support assessment

The Paraprofessional Stress and Support Assessment (PSSA; Reddy, Lekwa, & Shernoff, 2021) was used as an outcome measure to assess paraprofessional perceptions of instructional support, emotional support, and stress at baseline and post-assessment. The Instrumental Support Scale (8 items; $\alpha = 0.96$) assesses the amount and type of guidance and concrete support that teachers received from peers and instructional leaders. The Emotional Support Scale (6 items; $\alpha = 0.94$) assesses paraprofessional beliefs regarding the amount of acknowledgement, respect, and encouragement they receive at their respective schools. Paraprofessionals rated their level of agreement (1 = *Strongly Disagree* to 5 = *Strongly Agree*) on both of these scales. The Stress Scale (14 items, $\alpha = 0.93$) assesses paraprofessionals' feelings of strain and pressure stemming from the classroom and school context. Each item is rated on a 5-point scale (1 = *No Stress* to 5 = *Extremely High Stress*). Items on each scale are summed and divided by the number of items completed to compute a total scale score. The PSSA has good reliability, content, and construct validity. An exploratory factor analysis conducted as part of the present study suggested that the PSSA is comprised of three factors (i.e., scales; factor loadings ranging from 0.80 to 0.92 for ISS, 0.71 to 0.93 for ESS, and 0.36 to 0.86 for Stress). Similarly, internal consistency estimates (Cronbach alphas) in the current study for

the Instrument Support Scale, Emotional Support Scale, and Stress Scale were all above 0.90.

6.3.10. Usage rating profile–intervention revised

The Usage Rating Profile–Intervention Revised (URPI-IR; Briesch et al., 2013; Chafouleas et al., 2009) was used as an outcome measure at post-assessment for paraprofessionals and teachers in the BSC-P condition only and was used to assess BSC-P usability and feasibility. The URP-IR is a 29-item rating scale measuring the usability and feasibility of an intervention. In this study, the URP-IR was used to assess paraprofessional and teacher feedback on the acceptability of the BSC-P intervention. The URP-IR yields a 4-factor model (i.e., Acceptability, Understanding, Feasibility, and Systems Support) with items being rated on a 6-point Likert scale (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Slightly Disagree*, 4 = *Slightly Agree*, 5 = *Agree*, 6 = *Strongly Agree*). Previous coaching RCT investigations have used the URP-IR and reported acceptable levels of reliability (alphas above 0.85; Fabiano et al., 2018) as well as acceptable omega values for the Acceptability, Understanding, and Feasibility factors (0.94, 0.79, and 0.96, respectively). Concurrent validity with other acceptance-based measures has also been reported (Miller et al., 2014). In the present study, internal consistency estimates (Cronbach alphas) for paraprofessionals ranged from acceptable to fair for the Acceptability ($\alpha = 0.86$), Understanding ($\alpha = 0.78$), and Feasibility ($\alpha = 0.72$) scales. For teachers, internal consistency estimates for the same scales evidenced strong results with all values at or above 0.90.

6.3.11. Coaching evaluation scale

The Coaching Evaluation Scale (CES; Reddy et al., 2017) was used as an outcome measure at post-assessment for paraprofessionals and teachers in the BSC-P condition only to assess satisfaction with the BSC-P intervention. Paraprofessionals and their partner teachers separately completed the CES, which is a 14-item scale assessing satisfaction with the BSC-P. Teachers and paraprofessionals complete the CES by indicating their ratings using a 7-point Likert agreement scale (1 = *Strongly Disagree*, 4 = *Neutral*, 7 = *Strongly Agree*). Previous coaching RCTs utilizing the CES have demonstrated good internal consistency of 0.87–0.90 (Fabiano et al., 2018; Reddy, Shernoff, & Lekwa, 2021). In the present study, internal consistency estimates for paraprofessionals ratings were strong ($\alpha = 0.94$) and similarly were strong for teachers ($\alpha = 0.95$).

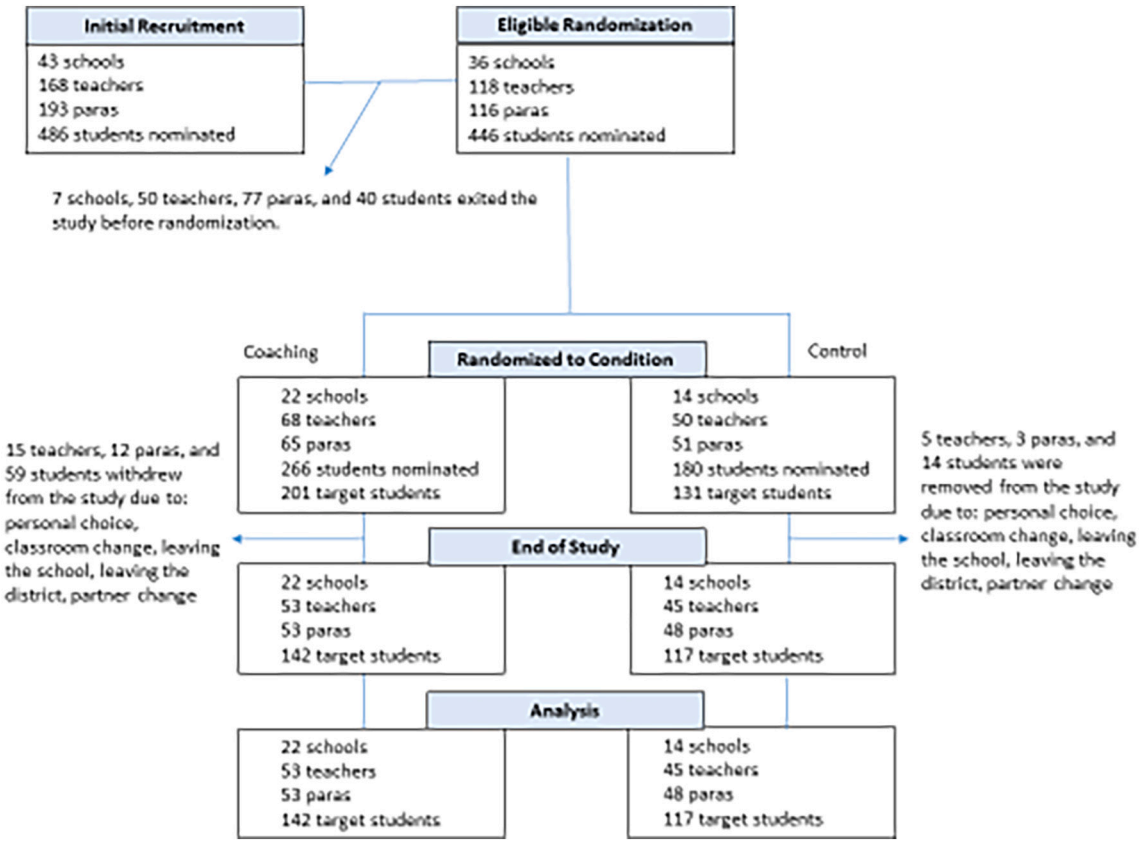

Fig. 1. Participant flow diagram.

6.4. Procedures

All study procedures were approved by the research institution's IRB. Informed consent was secured from all participants. Data were collected as part of two cohorts occurring participating during the 2017–2018 and 2018–2019 school years.

6.4.1. Inclusion/exclusion criteria

Schools were required to have a minimum of two participating classrooms in Grades K–5 that contained an assigned paraprofessional and teacher partner pair. For each participating classroom, teachers and paraprofessionals were required to identify approximately 3–5 students with or at risk for DBDs using study screening procedures (outlined below). If these inclusion criteria were not met, then a participating school building, or individual classroom within a school building, was not eligible to participate. Based on these criteria and the volunteer nature of the study, it was possible that not all school buildings in a district, and/or not all classrooms in a school building were eligible despite expressing interest in participating in the study. Given the focus of the study was on paraprofessionals working with students with or at risk for DBDs, students diagnosed with autism spectrum disorder, other developmental disabilities, or intellectual disabilities, along with paraprofessionals whose sole responsibilities involved support for such students, were not eligible for study participation.

6.4.2. Project timeline

For each cohort, a staggered implementation process was used whereby recruitment and randomization typically occurred during the July through November months. Screening procedures began following school randomization (approximately October through January). Once the screening procedures and school randomization were completed, baseline assessment began, which required approximately 2 weeks per school. Baseline assessment typically began in November of each year. Following baseline procedures, those in the BSC-P condition began receiving the coaching intervention, whereas those in the WL condition engaged in business-as-usual procedures. BSC-P was implemented across approximately 12 weeks, with coaching sessions occurring on average once every 9 days. Similarly, the WL condition non-intervention period mirrored the same duration. Once completed, both the BSC-P and WL conditions began the post-assessment process, which lasted approximately 2 weeks.

6.4.3. Recruitment and consent procedures

Recruitment involved a multi-step procedure to obtain research study permission from local school districts, school buildings, classrooms, and parents (see Fig. 1). First, school district leaders were contacted via phone and email about the benefits of the study and provided with project flyers. Second, a 45-min in-person presentation was then provided to school district administration and school building leaders that outlined study procedures, timelines, and inclusion/exclusion criteria in detail. Third, a copy of the approved IRB protocol, consent forms, and any school district required forms were provided to school district officials and legal teams for review. Fourth, once district-level approval was granted, interested school leaders then contacted study staff to schedule in-person recruitment presentations for qualifying paraprofessionals and teachers. Presentations were approximately 20 min long and included a detailed description of the study. Fifth, following the recruitment presentation, interested paraprofessionals and their partner teachers provided informed consent as per IRB procedures. Sixth, once a partner paraprofessional and classroom teacher pair consented to participate, the parent/guardian informed consent process was then implemented using a classroom-wide approach. Specifically, school leaders and classroom teachers were provided an IRB approved parent consent letter to send to all parents in the participating classroom that described the study, contact information for addressing questions, and instructions on how to opt-out of the study. If parents did not want their child to participate in the study, they were instructed to contact study authors and the classroom teacher via email or telephone. Following distribution of the parent/guardian consent letter, classrooms had to wait 2 weeks (i.e., for possible parent communications to the authors and/or teachers for opting out of the study) before beginning with the study. Lastly, a student-assent process was used for data collection on the WJ-IV BA. For the participating classrooms only, target students were individually read an assent script describing the larger BSC-P study and the WJ-IV BA. Students had to provide explicit verbal permission to be administered the WJ-IV BA at baseline and post-intervention.

6.4.4. Screening procedures

To ensure students met criteria, a 2-step screening procedure was used. First, participating classroom teachers were asked to nominate five students in their classroom who displayed heightened levels of disruptive behaviors as compared to classroom peers using a nomination form that included a description of common behaviors associated with DBDs. Teachers were also provided an open-ended section to list their behavioral concerns and specify the times of day when behavioral incidents were occurring. Second, for each nominated student, teachers completed the BESS-3. Students scoring at or above the "at-risk" category (t-score of 60 or greater) were eligible to participate. Lastly, the three students with the highest BESS-3 t-scores from each classroom were selected to participate in the study.

6.4.5. Randomization

After eligibility and screening procedures were met, randomization of conditions occurred at the school level. Classroom teachers and their paraprofessionals were notified of their condition status for the purposes of coordinating study activities.

6.4.6. Observer training

Independent observers received 4 h of BOSS training. Training encompassed didactic presentation of modified BOSS behavior codes

definitions, video-based examples, and interactive discussions with a research faculty member proficient in BOSS. Following training, observers completed classroom practice videos using the BOSS and received feedback from the research faculty member. Observers were then required to independently code the three classroom videos and pass reliability testing of 80% or greater. Throughout the study, BOSS inter-observer reliability was measured via two observers in the classroom observing the same student for the same 15-min duration. In the study, a total of 1679 student observations with the modified BOSS were conducted and a second observer was present for inter-rater reliability purposes for approximately 10% of the sample. Pearson correlations were computed for each of the six modified BOSS behavior code categories between the two observers to measure inter-observer reliability, which was strong with values ranging from 0.80 to 0.92.

6.4.7. Baseline assessment

Once the target students were identified, baseline assessment procedures began for paraprofessionals, classroom teachers, and target students.

Paraprofessionals. At baseline, BSC-P and WL paraprofessionals completed a demographics questionnaire containing questions about their age, gender, race/ethnicity, years of experience, and classroom setting. BSC-P and WL paraprofessionals also completed the PTRS about their relationship with the participating teacher. Additionally, BSC-P and WL condition paraprofessionals were provided with two BXL forms to complete as part of the baseline phase. BSC-P condition paraprofessionals turned in both forms to their assigned coach at the first coaching session (i.e., approximately 1 week after completion of student observations). WL condition paraprofessionals turned in both forms to study staff approximately 1 week after completion of student observations.

Teachers. Similar to paraprofessionals, partner classroom teachers completed a demographic questionnaire and the PTRS about their relationship with their classroom paraprofessional(s) participating in the study. Teachers also completed the BASC-3 and SSIS-RS measures for the identified target students.

Students. Baseline assessment included three direct observations per student with the BOSS, each 15 min in duration (i.e., 45 min total per student and 135 min total per classroom) within a 10-day period. The BOSS observations were conducted by trained and certified independent observers blind to study condition. Target times for observations were specified by the classroom teacher. Teachers were instructed to identify class periods (e.g., Language Arts) during which instructional activities (i.e., whole-group, small-group, centers, individualized instruction) were occurring in their classroom where these students exhibited disruptive behaviors. Additionally, target observation times were also required to include times when the paraprofessional was present in the classroom and were responsible for working with and monitoring the assigned students. Thus, student observations occurred when the participating partner paraprofessional and classroom teacher were present and the paraprofessional was engaging in their assigned responsibilities with the target students. On average, observations were 2 days apart. When students were absent or not able to be observed, observations were scheduled for the next available day/time indicated by the teacher. Observations were paused, and then resumed, when students left the classroom or received instructional breaks (i.e., 2 min or less). Observations were terminated when students were not present for 2 or more minutes.

Approximately 1 week after BOSS observations were conducted, the WJ-IV BA tests were administered to students by trained and supervised school psychology doctoral students. On average, the WJ-IV BA tests were completed in 30 min.

6.4.8. Coach training

The five BSC-P coaches included three doctoral-level school psychologists, a doctoral-level special educator and behavior analyst, and a bachelor's-level behavioral specialist with over 20 years of school-based behavior consultation experience. Coaches participated in a 3-day training on the purpose and structure of the BSC-P. They were then trained on a manualized approach that included session-specific procedural checklists aligned to the coaching session objectives shown in Table 1. Coaches received weekly supervision from the principal investigators to maximize fidelity and problem solve barriers to coaching implementation. Supervision included role playing and modeling coaching sessions, reviewing observational data and progress in meeting goals, and problem-solving barriers to implementation of paraprofessional-student goals. All coaches were trained to reliability on the BOSS and were required to complete a reliability test (i.e., 80% or higher) on three coded classroom videos.

6.4.9. BSC-P implementation

BSC-P was implemented in schools assigned to the BSC-P condition via eight coaching sessions, over approximately 12 weeks, with coach classroom observations conducted between each coaching session. Specifically, coaching sessions were on average 24 min in length ($SD = 12.11$) and occurred 13 days apart ($SD = 12.03$). In general, classroom teachers were present and active contributors in Sessions 1–3, whereas Sessions 4–7 focused on supporting the individual paraprofessional implement the selected interventions; classroom teachers were also present for the eighth session (i.e., final session) in the model. Table 1 includes an outline of the five-phase coaching process and objectives for each session. The coaches supported the classroom teachers and paraprofessionals in identifying and refining student needs based on observational data from the BOSS, formulating student goals, and appropriately selecting and implementing research-based interventions based on the functions of student behaviors. Regularly observing paraprofessionals, coaches provided feedback and encouragement pertaining to observed strategy implementation and graphed student and paraprofessional implementation progress. They also supported paraprofessionals in making data-informed adjustments to intervention practices.

Interventions were selected by the classroom teacher in collaboration with their paraprofessional and the BSC-P coach from a research-based toolkit specifically designed for the BSC-P. Toolkit interventions were supported by a large body of research (e.g., Bruhn et al., 2015; Kern et al., 2001; Shogren et al., 2004; Simonsen et al., 2008) and organized around common functions of behaviors

(e.g., acquisition intervention strategies, attention-seeking strategies, escape intervention strategies) and whether targeted behaviors were for individual or class-wide needs. Examples of the interventions included in the toolkit were active rule teaching, differential reinforcement, behavior contracts, and token economies.

Coaching Adherence. Approximately 30% of coaching sessions were audiotaped (eight sessions) and reviewed by lead authors using a fidelity checklist. Coaching fidelity was computed based on the percent of procedural steps completed by the coach within and across coaching sessions per case. Approximately 95% of BSC-P session specific components were implemented. In this study, the high degree of fidelity is related to the extensive training and supervision aligned to the checklists prior to implementation, and coaches had fidelity checklists in front of them during coaching.

Business as Usual for WL Controls. Following completion of the baseline student observations, WL condition paraprofessionals, teachers, and students had minimal interaction with study staff other than baseline data collection, requests for additional BXL logs, and study timeline updates.

6.4.10. Post assessment

BSC-P and WL conditions were administered post-assessments approximately 12 weeks after the end of coaching.

Paraprofessionals. Both BSC-P and WL paraprofessionals were provided with a post assessment packet that included a PTRS form and additional BXL forms. BSC-P participants were provided with the CES and URP-IR rating scales. Post assessment forms were provided to BSC-P participants following the conclusion of the final coaching session. WL paraprofessionals received their post assessment packet during post-assessment student observations. Both BSC-P and WL packets were collected from participants approximately 1 week after post-assessment student observations were completed.

Teachers. BSC-P and WL teachers were provided with a post-assessment packet at the same time as their partner paraprofessionals. Teachers were asked to complete the BASC-3 and SSIS-RS forms for each of the target students. BSC-P teachers were also provided with the CES and URP-IR rating scales.

Students. BOSS observations and WJ-IV BA were administered to participating students at post-assessment. Each of the target students received three, 15-min BOSS observations within 2 days of each other (45 min per student; 135 min per classroom). BOSS observations were conducted by independent observers who were blind to study condition. For the WJ-IV BA tests, only one target student was randomly selected at post-assessment due to the large data collection requirements for the study.

6.5. Data management and analytic approach

Differences between the BSC-P and WL control condition with respect to changes from baseline to post-assessment were analyzed for (a) self-report of paraprofessional behavior practices (i.e., BXL), (b) direct observation of student behavior and academic engagement (i.e., BOSS index scores), (c) student academic achievement (i.e., WJ-IV BA), (d) teacher ratings of student behavior (i.e., BASC-3) and social skills (i.e., SSIS), and (e) paraprofessional ratings of school-related supports (i.e., PSSA) and the quality of teacher-paraprofessional relationships (i.e., PTRS; Tables 4–5).

We calculated intra-class correlations (ICC) and associated design effects for paraprofessional behavior practices, student behavior

Table 4
Descriptive statistics for paraprofessional outcome variables.

	WL		BSC-P		
	Mean (SD)		Mean (SD)		
	Baseline	Post-coaching	Baseline	Post-coaching	Effect size[a]
Paraprofessional Behavior (Bx) Practices					
Antecedent Total	4.13 (3.26)	9.41 (9.72)	5.52 (5.07)	15.00 (11.69)*	0.91
Gave Praise for Appropriate Bx	1.19 (0.98)	3.33 (3.68)	1.66 (1.37)	3.78 (3.05)	
Rewarded Appropriate Bx	0.88 (0.89)	1.85 (2.2)	1.00 (1.13)	3.59 (2.60)*	1.51
Moved Seat	0.44 (0.73)	1.26 (1.61)	0.76 (1.02)	1.89 (1.91)	
Set up Behavior Contract	0.50 (0.73)	0.74 (1.4)	0.86 (1.64)	1.62 (2.16)	
Had Student Self Monitor	0.56 (0.73)	0.78 (1.48)	0.59 (1.12)	1.68 (2.06)	
Praised Peers	0.56 (0.73)	1.44 (2.29)	0.655 (0.86)	2.43 (2.69)	
Consequence Total	3.88 (2.09)	10.33 (7.36)	5.55 (5.33)	9.78 (7.85)	
Ignored Behavior	0.56 (0.81)	1.07 (1.36)	0.55 (0.74)	1.03 (1.30)	
Provided Time Out	0.69 (0.79)	1.00 (1.00)	0.76 (0.87)	1.59 (1.98)	
Reinforced Rules	1.81 (1.11)	5.48 (4.80)	3.03 (3.47)	5.03 (3.41)	
Removed Tokens/Privileges	0.81 (0.83)	2.78 (2.19)	1.21 (1.29)	2.14(2.86)	
Paraprofessional Ratings: Supports & Stress					
Emotional Support Scale	24.76 (3.83)	25.11 (4.48)	23.06 (5.38)	26.36 4.27)*	0.60
Instrumental Support Scale	25.80 (9.24)	27.88 (7.30)	26.21 (8.17)	33.83 4.70)*	0.63
Stress Scale	26.42 (10.48)	26.80 (9.74)	29.31 (9.17)	28.80 (11.12)	
Para-Teacher Relationship Total	112.60 (7.40)	113.67 (6.72)	113.55 (6.26)	111.33 (8.75)	

Note. WL = Waitlist; BSC-P = Behavior Support Coaching for Paraprofessionals.

* Significant at the $p < .05$ level.

[a] Morris effect size method (2008) for pre-test-posttest-control group designs was used.

Table 5
Descriptive statistics for student outcome variables.

	WL		BSC-P		
	Mean (SD)		Mean (SD)		
	Baseline	Post-coaching	Baseline	Post-coaching	Effect size[a]
Observed Student Behavior: BOSS					
Total Engagement (AP + PE)	0.57 (0.16)	0.66 (0.16)	0.54 (0.18)	0.78 (0.18)*	0.87
Active Engagement (AP)	0.25 (0.11)	0.31 (0.14)	0.24 (0.14)	0.37 (0.19)*	0.54
Passive Engagement (PE)	0.31 (0.14)	0.36 (0.15)	0.30 (,15)	0.40 (0.18)*	0.34
Inappropriate Physical	0.25 (0.15)	0.25 (0.17)	0.24 (0.14)	0.14 (0.13)*	−0.68
Inappropriate Verbal	0.15 (0.12)	0.17 (0.15)	0.17 (0.14)	0.10 (0.13)*	−0.68
Noncompliance	0.02 (0.04)	0.01 (0.02)	0.02 (0.04)	0.01 (0.03)	
Disruptive Academic	0.16 (0.11)	0.12 (0.10)	0.13 (0.11)	0.09 (0.13)	
Teacher Behavior Ratings: BASC-3					
Behavior Symptom Composite	71.51 (11.44)	70.21 (12.27)	70.63 (12.13)	68.43 (13.59)	
Externalizing Problems Composite	70.26 (14.17)	71.17 (16.65)	72.06 (15.30)	71.21 (16.23)	
Internalizing Problems Composite	59.72 (14.97)	56.85 (13.13)	57.73 (14.63)	57.76 (14.09)	
School Problems Composite	64.00 (8.87)	64.72 (10.16)	65.42 (9.73)	62.03 (9.37)*	−0.43
Adaptive Skills Composite	34.47 (6.77)	34.68 (7.75)	34.83 (7.04)	38.13 (8.49)*	0.44
Teacher Ratings: SSIS					
Social Skills Total	73.13 (13.12)	72.72 (15.56)	72.46 (13.28)	77.64 (14.44)*	0.42
Student Achievement: WJ Brief Achievement Composite	441.84 (36.98)	445.97 (34.63)	439.84 (38.92)	451.15 (37.31)	

Note. WL = Waitlist; BSC-P = Behavior Support Coaching for Paraprofessionals.
* Significant at the $p < .05$ level.
[a] Morris effect size method (2008) for pre-test-posttest-control group designs was used.

problems, academic engagement, social skills, and achievement to assess the effects of clustering of paraprofessionals and students within schools (Peugh, 2010). The average ICC was 0.43 (range = 0.09–0.89) and the average design effect was 1.68 (range = 1.15–2.45). Models of student outcomes that included school level effects were initially examined but found to be functionally equivalent to models that included only the paraprofessional level effects. As a result, we assessed the effects of BSC-P on paraprofessional behavior practices and student outcomes via a series of two-level multilevel models (Raudenbush & Bryk, 2002; Snijders et al., 1999) implemented as a mixed effects model in R (R Core Team, 2020) with the package lme4 (Bates et al., 2015). This approach took into account the nested nature of the sampling design with students (Level 1) and students nested within paraprofessionals (Level 2). For paraprofessional models, a two-level model was used along with difference scores (baseline – post-assessment; Level 1) of the outcome measures and a measure of paraprofessional behavior practice logs included as a covariate (Level 1) and condition as Level 2. Difference (i.e., change) scores are common in RCTs and yield similar results as models that control for baseline scores (Senn, 2006; Van Breukelen, 2006).

For each of the models, condition was entered as a dummy coded (0 = *Waitlist*, 1 = *BSC-P*) fixed effect. Timepoint (0 = *baseline*, 1 = *post-assessment*) was also included as a dummy coded fixed effect in the student outcome models. The Time X Condition interaction effect was of primary interest as it indicated the extent to which differences in baseline to post-assessment outcome changes were noted between the BSC-P group and WL control conditions.

6.5.1. *Missing data*

A form of full maximum likelihood estimation was used to account for missing data that were considered to be Missing Completely at Random (MCAR) or Missing at Random (MAR). Specifically, restricted maximum likelihood estimation (REML, Little & Rubin, 2002; Raudenbush & Bryk, 2002) was used in the multilevel models given the unequal number of paraprofessionals in each condition along with the differing number of observed data points in the student outcomes. Thus, under the assumption that data were MCAR or MAR, we applied REML estimation to recover statistical power and to diminish bias in parameter estimation for the effect of coaching on constructs measured in paraprofessional behavior practice logs, as well as teacher and paraprofessional rating scales (Enders & Bandalos, 2001).

Little's MCAR test was conducted to ensure MCAR assumptions were met. Approximately 2.5% or fewer BOSS scores were missing from the total sample and the underlying mechanisms of missingness for these variables were considered negligible sources of bias in analysis (e.g., Little et al., 2014). Missing data for the BOSS and teacher rating scales of BASC-3 and SSIS existed. Results of Little's MCAR Test (Little, 1988, Little et al., 2014) suggested that our data were missing completely at random (MCAR; $\chi^2(130) = 145.10$, $p = .173$). Likewise, missing data existed for the paraprofessional behavior practice logs and paraprofessional rating scales of PSSA and PTRS. Little's MCAR Test (Little, 1988; Little et al., 2014) suggested that our data were missing completely at random (MCAR; $\chi^2(4) = 3.862$, $p = .425$).

We calculated standardized mean difference effect sizes (e.g., Cohen's *d*; Cohen, 1988) following the method described by Morris (2008) for pre-test-posttest-control group designs, in which the difference between the mean baseline-post assessment of the control and BSC-P conditions was divided by the pooled baseline standard deviation (Tables 4–5).

7. Results

Primary outcome measures included (a) paraprofessional behavior practices; (b) paraprofessional perceived supports, stress, and relationship quality with classroom teachers (Table 4); and (c) student behavior, academic engagement, social skills, and achievement (Table 5). Tables 4 and 5 present descriptive statistics, parameter estimates for fixed effects of interest (i.e., Condition for models evaluating paraprofessional outcomes; Timepoint X Condition interaction for models evaluating student outcomes), and associated effect sizes.

7.1. Effects of BSC-P on paraprofessional behavior practices

Baseline descriptive statistics indicated some variability in paraprofessional behavior practice logs between conditions (Table 4). Results from the two-level multilevel models indicated BSC-P paraprofessionals reported greater implementation of antecedent practices than WL control paraprofessionals at post-assessment (i.e., Antecedent Total: $b = -8.99$, $p \leq .05$, $d = 0.91$; Rewarded Appropriate Behavior: $b = -2.43$, $p \leq .001$, $d = 0.1.51$). Between-condition differences were not found for other specific antecedent practices. Likewise, between-condition differences in the implementation of consequence practices were not found in the current study.

7.2. Effects of BSC-P on paraprofessional perceived supports, stress, and relationships with teachers

Baseline ratings of paraprofessional perceived supports and stress were comparable in both conditions at baseline (Table 4). Results from the two level multilevel models indicated BSC-P paraprofessionals rated greater improvement in emotional and instrumental support than WL paraprofessionals at post-assessment (Emotional Support: $b = -2.94$, $p = .023$, $d = 0.60$; Instrumental Support: $b = -5.56$, $p = .029$, $d = 0.63$). No between-condition differences were found for paraprofessional work-related stress and relationships with teachers.

7.3. Effects of BSC-P on observed student behavior and academic achievement

Table 5 presents descriptive statistics for primary student outcome variables. Baseline rates of student behavior and academic engagement were collected by independent observers and were comparable with those observed in prior research (e.g., DuPaul et al., 2004; Reddy, Shernoff, & Lekwa, 2021; Sheridan et al., 2012). Results from the two-level models indicated significant Timepoint X Condition effects for five out of the seven BOSS indices. Specifically, Total Engagement ($b = -0.14$, $p = .001$, $d = 0.87$), Active Engagement ($b = -0.07$, $p = .004$, $d = 0.54$), Passive Engagement ($b = -0.06$, $p = .03$, $d = 0.34$), Inappropriate Physical ($b = 0.10$, $p = .001$, $d = -0.68$), and Inappropriate Verbal ($b = 0.09$, $p = .001$, $d = -0.68$) behaviors. Results did not indicate significant Time X Condition effects for Noncompliance and Disruptive Academic behaviors at post-assessment.

Baseline student achievement for both conditions was comparable (Table 5). Significant Timepoint X Condition effects were not found for student performance on the WJ IV Brief Achievement ($b = -7.67$, $p = .06$).

7.4. Effects of BSC-P on teacher-rated student behavior and social skills

Baseline teacher ratings of student behavior and social skills were comparable in both conditions (Table 5). Two level multilevel models revealed that, based on teacher ratings, students supported by BSC-P paraprofessionals had greater reductions in school problems (BASC-3 School Problems composite: $b = 3.06$, $p = .02$, $d = -0.43$) and greater improvements in adaptive behavior (BASC-3 Adaptive Skills composite: $b = -3.42$, $p = .005$, $d = 0.44$) than those supported by WL control paraprofessionals at post-assessment. Results did not indicate between-condition effects for the BASC-3 composites of Behavior Symptoms, Externalizing Problems, and Internalizing Problems. Moreover, results indicated that, based on teacher ratings, students supported by BSC-P paraprofessionals were rated as having greater improvements in social skills (SSIS Social Skills Total: $b = -6.31$, $p = .01$, $d = 0.42$).

7.5. Paraprofessional and teachers rated satisfaction, usability, and feasibility with coaching

Paraprofessionals and teachers assigned to the BSC-P coaching condition completed the CES and URP-IR to assess their satisfaction, usability (i.e., understandability), and feasibility of BSC-P following the completion of the coaching and post-assessment observations. Results from descriptive analyses of CES and URP-IR data indicated that, overall, BSC-P paraprofessionals and teachers reported satisfaction with coaching (on a 7-point scale with higher ratings indicating more favorable perceptions, CES $M = 6.48$, 5.68; $SD = 0.96$, 1.50, respectively). Mean CES ratings for paraprofessionals did not fall below 6.25 (for "I feel better prepared to work with children with challenging behaviors"; $SD = 1.10$) and were as high as 6.77 (for "I would work with this coach again"; $SD = 1.00$). Teachers CES scores were slightly lower, ranging from as low as 5.12 (for "Demands placed on me by the assessment components were reasonable"; $SD = 1.80$) to as high as 6.26 (for "The professional I worked with showed a high level of concern for me and my classroom"; $SD = 1.27$).

Likewise, paraprofessionals and teachers indicated positive ratings of agreement for Acceptability, Understanding, and Feasibility factors with coaching via the URP-IR. Specifically, on a 6-point scale with higher ratings indicating more favorable perceptions, mean paraprofessionals' and teachers' acceptability ratings were 5.15 ($SD = 1.03$) and 4.66 ($SD = 1.28$), respectively, and mean feasibility

ratings were 4.96 ($SD = 1.03$) and 4.50 ($SD = 1.29$), respectively. Mean understanding factor ratings were 5.39 ($SD = 0.61$) and 5.26 ($SD = 0.86$), respectively. Collectively, paraprofessionals' and teachers' ratings provide evidence of acceptable satisfaction, usability, and feasibility for the coaching intervention.

8. Discussion

Paraprofessionals serve as vital support staff for classroom teachers and students eligible for special education services. As school personnel, they have historically received limited supervision and training to meet the complex academic, behavioral, and social needs of the students they serve (e.g., Ashbaker & Morgan, 2006; Giangreco et al., 2013). Job-embedded professional development models, such as data-driven coaching, are a promising avenue to enhance paraprofessionals' use of research-based behavior practices and improve student behavior, academic engagement, and learning in authentic classroom settings (e.g., Kraft et al., 2018; Reddy, Lekwa, & Shernoff, 2021; Reddy, Shernoff, & Lekwa, 2021). Research on effective coaching approaches for paraprofessionals is very limited, but essential for promoting paraprofessional knowledge, practices, and wellness. This investigation tested and demonstrated the effectiveness of an innovative coaching model that iteratively used observational data to guide coaching decisions and actions for enhancing paraprofessional behavior practices and student outcomes in elementary schools.

Overall, outcomes from the current investigation were promising and offer evidence supporting the use of school-based coaching with paraprofessionals. Specifically, the BSC-P model, which emphasizes key coaching actions (e.g., use of ongoing observational data, modeling and facilitating opportunities to practice strategy implementation, providing performance feedback) and social support mechanisms (e.g., focusing on professional strengths, providing encouragement) appeared to enhance paraprofessional skills and wellbeing, as well as improve student behavioral outcomes.

In this study, we found significant between-condition differences with respect to improvements in paraprofessionals' behavioral strategy implementation (Table 4). Relative to WL control participants, BSC-P paraprofessionals reported overall greater increases in antecedent practices ($d = 0.17$) as measured by a paraprofessional log. This is a positive finding given that antecedent practices are essential for paraprofessionals to proactively intervene and cultivate a supportive relationship with students with disruptive behaviors who are often accustomed to negative adult interactions (O'Connor et al., 2011; Sutherland et al., 2020). Specifically, BSC-P paraprofessionals reported greater use of rewarding appropriate behavior than WL control participants ($d = 0.67$). These encouraging findings are aligned with previous results indicating that coaching with performance feedback enhances teachers' use of behavior praise and overall quality of praise delivered in learning opportunities (e.g., Coffee & Kratochwill, 2013; Fabiano et al., 2018; Reddy, Shernoff, & Lekwa, 2021; Sutherland et al., 2020). Between-condition effects were not found for consequent strategies (i.e., practices used to respond to behavior concerns). In the present study, BSC-P paraprofessionals may have become more attuned to students' behavioral triggers, thus increasing their reliance on proactive strategies. In addition, consistent with prior coaching research (Stormont et al., 2015), BSC-P was favorably rated by participating paraprofessionals and teachers.

In this investigation we assessed the impact of BSC-P on paraprofessionals' perceived instructional and emotional support. Emotional support for paraprofessionals was defined as feelings of acknowledgement, respect, and encouragement, whereas instructional support was defined as practical guidance received from classroom teachers and other school personnel to inform changes in practices with students. Relative to WL control participants, BSC-P paraprofessionals reported a greater increase in emotional ($d = .60$) and instructional supports ($d = 0.63$), but no improvement in work-related stress. Given concerns about paraprofessional stress and burnout when managing student behavior (e.g., Bettini et al., 2016; Garwood & Vernon-Feagans, 2017; Tillery et al., 2003), these findings are important and suggest that the BSC-P Model, like other coaching approaches (e.g., Eddy & Thomas, 2019; Reddy, Shernoff, & Lekwa, 2021; Reddy, Lekwa, & Shernoff, 2021), may bolster aspects of perceived support and wellbeing by the recipient. It is possible that BSC-P's solution-oriented framework assisted paraprofessionals in developing adaptive coping mechanisms and perceptions of self-efficacy to meet the demands of their classrooms, which then reduced their stress. Additionally, findings from this study also indicate that our support scales (PSSA) for paraprofessionals were sensitive to change following coaching.

This study utilized both direct observations and teacher ratings to assess the effect of BSC-P on student outcomes (Table 5). Multilevel analyses revealed that relative to WL controls, students from BSC-P classrooms had greater improvements in verbal and physical aggression ($ds = -0.68$), academic engagement ($d = 0.87$), and teacher-rated student school symptomology (BASC-3 $d = -0.43$), adaptive skills (BASC-3 $d = 0.44$), and social skills (SSIS-RS $d = 0.42$). Our findings are important due to the lack of existing empirical investigations demonstrating effective paraprofessional professional development approaches on student behavior. Although results of this investigation add to this limited research base, our findings are consistent with the results of prior teacher coaching studies that demonstrated an experimental effect on reducing student disruptive behavior and increasing student adaptive and social skills (e.g., Colton & Sheridan, 1998; Mautone et al., 2012; Owens et al., 2012; Sheridan et al., 2012, 2017; Sutherland et al., 2020; Wilkinson, 2005). Given the focus of the BSC-P model on specific student behaviors, this may be one potential explanation as to why coaching did not significantly impact student achievement ($p = .06$). Sutherland et al. (2020) teacher coaching investigation reported similar findings with respect to student academic achievement. Alternatively, with the short-term nature and foci of BSC-P, it is possible that more time may be needed to see improvements in academic performance. Similar to prior intervention research (e.g., Kellam et al., 2014), BSC-P may lead to better long term academic outcomes through helping students be more engaged and less disruptive. Collectively, direct observation and teacher ratings converged on the potential benefit of the BSC-P in impacting critical student outcomes.

8.1. Study limitations

Findings from this investigation should be considered in light of limitations regarding setting, sample, and data collection. First, the observed outcomes may not generalize to all paraprofessionals or teachers. Participating paraprofessionals were predominately Black, White, and Hispanic females and detailed information on quality of prior education and professional development training were not collected. Likewise, comparable to demographics at the state level, participating elementary school teachers were predominately White females. Thus, our results may not be generalizable to other groups and grade levels (i.e., pre, middle, or high school) as we were unable to assess the influence of paraprofessional and teacher characteristics or other services received on behavior practices or coaching outcomes. Differences in race/ethnic composition between paraprofessionals and teachers in this study were not examined and may have implications for cultural competence or humility in supervising and coaching paraprofessionals. Second, BSC-P was delivered primarily in elementary general education classrooms in a northeastern state. Thus, results may not generalize to other learning contexts (e.g., more restrictive classrooms settings and alternative education programs), schools (e.g., middle schools), states, or regions of the country. Third, school administrators may not have allowed certain teachers and paraprofessionals to participate based on concerns regarding their classroom practices. It remains unclear if coaching supports would produce similar findings with such teachers and paraprofessionals. Fourth, this study utilized BSC-P and WL paraprofessional logs to assess changes in behavior practices during the study. Self-report may be susceptible to social desirability and not sensitive enough to capture practice nuances and changes over time. Fifth, the present investigation did not assess the long-term effects of BSC-P. Thus, it is unclear the degree to which paraprofessional research-based behavior practices implemented as a result of coaching and/or observed student outcomes would be sustained over time. Sixth, it is possible that the presence of independent observers in the classroom may have influenced teacher classroom management, paraprofessional practices, and teacher or paraprofessional-student interactions. Seventh, we were not able to collect an expanded range of student characteristics (e.g., Title 1 status, special education classifications) due to school district procedures. Eighth, due to sample size, we were unable to explore possible factors that may mediate or moderate the relationship between BSC-P on paraprofessional practices or student outcomes. Finally, this study used modified BOSS behavior categories for meeting project and student behavior needs. The nuances of determining noncompliant and disruptive academic behaviors are a limitation and warrants future investigations.

8.2. Directions for research

Findings from this study offer several directions for bridging the research to practice gap. First, replication of these findings with a larger number of paraprofessionals is needed. Such replication will allow for greater statistical power and potential identification of small differences in effects of coaching between paraprofessionals. Furthermore, a larger sample size will enable analyses of the extent to which the qualities of the classroom environment may influence paraprofessionals' behavior practices and interactions for students with or at risk for EBDs. Second, findings in this study underscore the need for replication with diverse classroom teachers, paraprofessionals, and student populations. For example, studies of the efficacy of BSC-P could take place in urban or rural high-poverty settings, private schools, middle school, preschool, high school, and with paraprofessionals who are bilingual. Third, future studies incorporating observational and self-report assessment of paraprofessional behavior practices are needed (Lekwa & Reddy, 2021). The development and validation of assessments of paraprofessional practices would enable researchers and practitioners to analyze baseline/post-intervention differences in use of research-based behavior practices in relation to coaching supports. Fourth, this study utilized behavior coaches that were external to participating schools. The ability to scale use of the BSC-P with school-based coaches (e.g., specialists, school psychologists) would be beneficial. It is possible that BSC-P adoption by schools may result in varied implementations, which in turn may inform future coaching model development efforts. Fifth, as the BSC-P model included multiple components and processes, a more-fine grained analysis of the influence of specific components may offer further insight into the active elements that lead to paraprofessional and student outcomes for this population. Finally, future research should examine the influence of possible moderators or mediators on paraprofessional practices and student outcomes. For example, an investigation examining how paraprofessional perceived supports moderate their behavior strategy implementation or student behavior outcomes would be beneficial.

8.3. Implications for school practice

Findings from this investigation offer a solution for addressing longstanding concerns from the education community regarding paraprofessional qualifications and skills in working with students with severe academic, behavior, and social impairments (e.g., Broer et al., 2005; Carter et al., 2009; Giangreco et al., 2013; Lochman et al., 2010). This study demonstrated the successful adaptation of a job-embedded coaching model based on teacher instructional coaching to paraprofessionals in elementary school settings. Despite reporting moderate levels of knowledge and training in behavior management, paraprofessionals in this study were able to implement research-based behavior interventions with high levels of fidelity in authentic classroom environments as the result of working with a behavior support coach. Many school districts already employ instructional coaches or behavior interventionists in some capacity to support educator and student needs. Thus, many districts may possess the internal capacity to support coaching models such as the BSC-P for their paraprofessionals.

School psychologists are uniquely positioned to serve their districts as BSC-P coaches given their knowledge and training in behavior management, consultation, and research-based interventions. School psychologists regularly consult and support classroom teachers and paraprofessionals regarding student academic and behavioral concerns (Bramlett et al., 2002; Shernoff et al., 2017). As a

manualized model, BSC-P may provide school psychologists and teachers with the structure and guidance needed to supervise and coach paraprofessionals in the classroom contexts. School psychologists also have the background and knowledge necessary to train other school personnel (e.g., interventionists, classroom teachers) to serve as school-based coaches using BSC-P.

In addition to building paraprofessional skills, BSC-P was found to enhance paraprofessionals' wellbeing via their perceived instrumental and emotional supports, as well as job satisfaction. This is noteworthy as increased feelings of support and job satisfaction have been found to decrease school personnel burnout and turnover (e.g., Ghere & York-Barr, 2007; Tillery et al., 2003). Research has found that paraprofessionals often resign from their positions due to on-the-job stress and conflicts with educators (Ghere & York-Barr, 2007) and students are less likely to meet their Individualized Education Plan goals when working with paraprofessionals who are experiencing burnout (Garwood et al., 2018). Thus, these findings underscore the need for supporting paraprofessionals in their work with challenging students in ways that are similar to and equivalent of the supports provided to classroom teachers and other educators. Specifically, research-based approaches are need that improve paraprofessional skills, wellbeing, and retention in schools, which in turn can improve student outcomes.

Finally, BSC-P may be a useful approach for school systems that have adopted multi-tiered system of supports (MTSS). MTSS frameworks, including positive behavior interventions and supports programs, utilize data to guide decisions for selecting and implementing research-based practices to meet the continuum of student needs in classroom settings (Gamm et al., 2012; Sugai, 2009). Thus, BSC-P may be a promising professional development intervention for paraprofessionals supporting Tier 2 and Tier 3 student behavior needs in schools.

9. Conclusion

Findings from the present investigation advance knowledge regarding the implementation of behavior support coaching for paraprofessionals supporting students with or at risk for DBDs in elementary school classrooms. The current RCT marks the first to test the efficacy of the BSC-P model specifically for supporting paraprofessionals in their use of data-driven, research-based behavior interventions and promoting positive student outcomes in elementary school settings. Overall, findings from this study were promising and spotlight the need for further validation with larger paraprofessional samples and in alternate contexts. This research offers possible new avenues for school personnel preparation and professional development.

References

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2001). *Ratings of relations between DSM-IV diagnostic categories and items*. University of Vermont, Research Center for Children, Youth, & Families. https://aseba.org/wp-content/uploads/2019/02/DSM6-18ratings.pdf.

Allen, C. (2016). Externalizing disorders: Assessment, treatment, and school-based interventions. In R. Flanagan, & C. Allen (Eds.), *Cognitive behavioral therapies for children in schools* (pp. 161–180). Springer Publishing.

Alperin, A., Reddy, L. A., Glover, T. A., Breeden, N., Dudek, C., & Regan, P. (2020). Behavior support coaching for a paraprofessional working with first-grade students exhibiting disruptive behavior problems in an urban high-poverty elementary school. *Clinical Case Studies, 19*(5), 303–320. https://doi.org/10.1177/1534650120935197

Ashbaker, B. Y., & Morgan, J. (2006). The role of administrators in paraprofessional supervision to support ethnic minority students with special needs. *Educational Considerations, 34*(1), 17–21. https://doi.org/10.4148/0146-9282.1196

Barker, E. D., Oliver, B. R., & Maughan, B. (2010). Co-occurring problems of early onset persistent, childhood limited, and adolescent onset conduct problem youth. *Journal of Child Psychology and Psychiatry, 51*(11), 1217–1226. https://doi.org/10.1111/j.1469-7610.2010.02240.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy: Applied clinical psychology*. Plenum Press.

Bettini, E. A., Crockett, J. B., Brownell, M. T., & Merrill, K. L. (2016). Relationships between working conditions and special educators' instruction. *Journal of Special Education, 50*(3), 178–190. https://doi.org/10.1177/0022466916644425

Bottiani, J. H., Duran, C. A., Pas, E. T., & Bradshaw, C. P. (2019). Teacher stress and burnout in urban middle schools: Associations with job demands, resources, and effective classroom practices. *Journal of School Psychology, 77*, 36–51.

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39*(3), 327–335. https://doi.org/10.1002/pits.10022

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology, 51*(1), 81–96.

Broer, S. M., Doyle, M. B., & Giangreco, M. F. (2005). Perspectives of students with intellectual disabilities about their experiences with paraprofessional support. *Exceptional Children, 71*(4), 415–430.

Bruhn, A., Mcdaniel, S., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. *Behavioral Disorders, 40*(2), 102–121. https://doi.org/10.17988/BD-13-45.1

Carter, E., O'Rourke, L., Sisco, L. G., & Pelsue, D. (2009). Knowledge, responsibilities, and training needs of paraprofessionals in elementary and secondary schools. *Remedial and Special Education, 30*(6), 344–359.

Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct behavior rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention, 34*(4), 195–200. https://doi.org/10.1177/1534508409340391

Chopra, R. V., Sandoval-Lucero, E., & French, N. K. (2011). Effective supervision of paraeducators: Multiple benefits and outcomes. *National Teacher Education Journal, 4*(2), 15–26. https://paracenter.org/files/Effective%20Supervision%20of%20Paraeducators%20-%20Multiple%20Benefits%20and%20Outcomes.pdf.

Coffee, G., & Kratochwill, T. R. (2013). Examining teacher use of praise taught during behavioral consultation: Implementation and generalization considerations. *Journal of Educational and Psychological Consultation, 23*(1), 1–35. https://doi.org/10.1080/10474412.2013.757150

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Academic Press.

Colton, D. L., & Sheridan, S. M. (1998). Conjoint behavioral consultation and social skills training: Enhancing the play behaviors of boys with attention deficit hyperactivity disorder. *Journal of Educational and Psychological Consultation, 9*(1), 3–28. https://doi.org/10.1207/s1532768xjepc0901_1

Conners, C. K. (2008). *Conners* (3rd ed.). Multi-Health Systems.

Delligatti, N., Akin-Little, A. N., & Little, S. G. (2003). Conduct disorder in girls: Diagnostic and intervention issues. *Psychology in the Schools, 40*(2), 183–192. https://doi.org/10.1002/pits.10080

Denton, C. A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation, 19*, 150–175.

DuPaul, G. J., Volpe, R. J., Jitendra, A. K., Lutz, J. G., Lorah, K. S., & Gruber, R. (2004). Elementary school students with AD/HD: Predictors of academic achievement. *Journal of School Psychology, 42*(4), 285–301. https://doi.org/10.1016/j.jsp.2004.05.001

Eddy, K. T., & Thomas, J. J. (2019). Introduction to a special issue on child and adolescent feeding and eating disorders and avoidant/restrictive food intake disorder. *The International Journal of Eating Disorders, 52*(4), 327–330. https://doi.org/10.1002/eat.23052

Education Advisory Board. (2019). Breaking bad behavior: The rise of classroom disruptions in early grades and how districts are responding. District Leadership Forum http://pages.eab.com/rs/732-GKV-655/images/BreakingBadBehaviorStudy.pdf.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Erchul, W. P. (2015). Put me in, coach: Observations on selected studies implementing supportive interventions to teachers. *School Mental Health, 7*(1), 74–79.

Evers, W. J. G., Tomic, W., & Brouwers, A. (2004). Burnout among teachers: Students' and teachers' perceptions compared. *School Psychology International, 25*(2), 131–148. https://doi.org/10.1177/0143034304043670

Every Student Succeeds Act. (2015). Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016). Retrieved from: https://www.ed.gov/essa?src=rn.

Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child and Adolescent Psychology, 37*(1), 215–237. https://doi.org/10.1080/15374410701820117

Fabiano, G., Reddy, L. A., & Dudek, C. M. (2018). Use of teacher coaching and formative assessment for improving classroom practices. *School Psychology Quarterly, 33*(2), 293–304. https://doi.org/10.1037/spq0000223

Fisher, M., & Pleasants, S. L. (2012). Roles, responsibilities, and concerns of paraeducators: Findings from a statewide survey. *Remedial and Special Education, 33*(5), 287–297. https://doi.org/10.1177/0741932510397762

Gamm, S., Elliott, J., Halbert, J. W., Price-Baugh, R., Hall, R., Walstron, D., Uro, G., & Casserly, M. (2012). *Common core state standards and diverse urban students: Using multi-tiered systems of support.* Council of the Great City Schools.

Garwood, J. D., Van Loan, C. L., & Werts, M. G. (2018). Mindset of paraprofessionals serving students with emotional and behavioral disorders. *Intervention in School and Clinic, 53*(4), 206–211. https://doi.org/10.1177/1053451217712958

Garwood, J. D., & Vernon-Feagans, L. (2017). Classroom management affects literacy development of students with emotional and behavioral disorders. *Exceptional Children, 83*(2), 123–142. https://doi.org/10.1177/0014402916651846

Ghere, G., & York-Barr, J. (2007). Paraprofessional turnover and retention in inclusive programs: Hidden costs and promising practices. *Remedial and Special Education, 28*(1), 21–32. https://doi.org/10.1177/07419325070280010301

Giangreco, M. F., & Broer, S. M. (2005). Questionable utilization of paraprofessionals in inclusive schools: Are we addressing symptoms or causes? *Focus on Autism and Other Developmental Disabilities, 20*(1), 10–26. https://doi.org/10.1177/10883576050200010201

Giangreco, M. F., Suter, J. C., & Hurley, S. M. (2013). Revisiting personnel utilization in inclusion-oriented schools. *Journal of Special Education, 47*(2), 121–132. https://doi.org/10.1177/0022466911419015

Glover, T.A., Reddy, L.A., & Crouse, K., (n.d). Instructional coaching actions that predict improved teacher classroom practices and student achievement. Journal of School Psychology (under review).

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales.* Minneapolis, MN: NCE Pearson.

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the social skills rating system to the social skills improvement system: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*(1), 27–44. https://doi.org/10.1037/a0022662.

Individuals with Disabilities Education Act. (2004). 20 U.S.C. § 1400. https://sites.ed.gov/idea/statute-chapter-33/subchapter-ii/1412.

Kamphaus, R. W., & Reynolds, C. R. (2007). *The behavioral and emotional screening system.* Pearson Assessment.

Kamphaus, R. W., & Reynolds, C. R. (2015). *BASC-3 behavioral and emotional screening system.* Bloomington, MN: Pearson.

Kellam, S. G., Wang, W., Mackenzie, A. C., Brown, C. H., Ompad, D. C., Or, F., & Windham, A. (2014). The impact of the good behavior game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science, 15*(1), 6–18. https://doi-org.proxy.libraries.rutgers.edu/10.1007/s11121-012-0296-z.

Kern, L, Delaney, B, Clarke, S, Dunlap, G, & Childs, K (2001). Improving the Classroom Behavior of Students with Emotional and Behavioral Disorders Using Individualized Curricular Modifications. *Journal of Emotional and Behavioral Disorders, 9*(4), 239–247. https://doi.org/10.1177/106342660100900404

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. https://doi.org/10.3102/0034654318759268.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. https://ies.ed.gov/ncee/wwc/Document/229.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education, 33*(4), 279–299. https://doi.org/10.1177/0888406410371643.

Kurz, A., Reddy, L. A., & Glover, T. A. (2017). A multidisciplinary framework of instructional coaching. *Theory Into Practice, 56*(1), 66–77.

Lekwa, A., & Reddy, L. A. (2021). Assessment of paraprofessional practices: Current status and future measurement directions. *Psychology in the Schools*. https://doi.org/10.1002/pits.22480

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.2307/2290157.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons, Inc.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of Pediatric Psychology, 39*(2), 151–162. https://doi.org/10.1093/jpepsy/jst048.

Lochman, J. E., Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., McMahon, R. J., & Pinderhughes, E. E. (2010). The difficulty of maintaining positive intervention effects: A look at disruptive behavior, deviant peer relations, and social skills during the middle school years. *Journal of Early Adolescence, 30*(4), 593–624. https://doi.org/10.1177/0272431609340513.

Mautone, J. A., Marshall, S. A., Sharman, J., Eiraldi, R. B., Jawad, A. F., & Power, T. J. (2012). Development of a family–school intervention for young children with attention deficit hyperactivity disorder. *School Psychology Review, 41*(4), 447–466. https://doi.org/10.1080/02796015.2012.12087499.

Miller, J. D., McCain, J., Lynam, D. R., Few, L. R., Gentile, B., MacKillop, J., & Campbell, W. K. (2014). A comparison of the criterion validity of popular measures of narcissism and narcissistic personality disorder via the use of expert ratings. *Psychological Assessment, 26*(3), 958–969. https://doi.org/10.1037/a0036613.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364–386. https://doi.org/10.1177/1094428106291059.

Mrachko, A. A., & Kaczmarek, L. A. (2017). Examining paraprofessional interventions to increase social communication for young children with ASD. *Topics in Early Childhood Special Education, 37*(1), 4–15. https://doi.org/10.1177/0271121416662870.

O'Connor, E. E., Dearing, E., & Collins, B. A. (2011). Teacher-child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal, 48*(1), 120–162. https://doi.org/10.3102/0002831210365008.

Ota, K. R., & DuPaul, G. J. (2002). Task engagement and mathematics performance in children with attention-deficit hyperactivity disorder: Effects of supplemental computer instruction. *School Psychology Quarterly, 17*(3), 242–257. https://doi.org/10.1521/scpq.17.3.242.20881.

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International, 33*(4), 433–449. https://doi.org/10.1177/0143034311427433.

Penno, D. A., Frank, A. R., & Wacker, D. P. (2000). Instructional accommodations for adolescent students with severe emotional or behavioral disorders. *Behavioral Disorders, 25*(4), 325–343. https://doi.org/10.1177/019874290002500407.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85–112. https://doi.org/10.1016/j.jsp.2009.09.002.

Pianta, R. C., Lipscomb, D., & Ruzek, E. (2022). Indirect effects of coaching on pre-K students' engagement and literacy skill as a function of improved teacher–student interaction. *Journal of School Psychology, 91*, 65–80.

Pikard, J., Roberts, N., & Groll, D. (2018). Pediatric referrals for urgent psychiatric consultation: Clinical characteristics, diagnoses and outcome of 4 to 12 year old children. *Journal of Canadian Academy of Child and Adolescent Psychiatry, 27*(4), 245–251. https://www.cacap-acpea.org/wp-content/uploads/Pediatric-Referrals-for-Urgent-Psychiatric-Consultation.pdf.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Reddy, L. A., Alperin, A., & Glover, T. A. (2021). A critical review of the professional development literature for paraprofessionals supporting students with externalizing behavior disorders. *Psychology in the Schools, 58*(4), 742–763. https://doi.org/10.1002/pits.22381.

Reddy, L. A., Espelage, D. L., Anderman, E. M., Kanrich, J. B., & McMahon, S. D. (2018). Addressing violence against educators through measurement and research. *Aggression and Violent Behavior, 42*, 9–28. https://doi.org/10.1016/j.avb.2018.06.006.

Reddy, L. A., Glover, T. A., Alperin, A., Grant, C., & Regan, P. (2017). *Behavioral interventions toolkit for elementary school students*. Unpublished manuscript. Rutgers University.

Reddy, L. A., Lekwa, A., & Shernoff, E. (2021). Comparison of the effects of coaching for general and special education teachers in high-poverty urban elementary schools. *Journal of Learning Disabilities, 54*(1), 36–53. https://doi.org/10.1177/0022219420970194

Reddy, L. A., Newman, E., DeThomas, C., & Chun, V. (2009). Effectiveness of school-based prevention and intervention programs for children and adolescent with emotional disturbance: A meta-analysis. *Journal of School Psychology, 47*, 77–99. https://doi.org/10.1016/j.jsp.2008.11.001

Reddy, L. A., Shernoff, E., & Lekwa, A. (2021). A randomized controlled trial of instructional coaching in high-poverty schools: Examining teacher practices and student outcomes. *Journal of School Psychology, 86*, 151–168. https://doi.org/10.1016/j.jsp.2021.04.001

Reddy, L. A., Weissman, A. S., & Hale, J. B. (2013). *Neuropsychological assessment and intervention for youth: An evidence-based approach to emotional and behavioral disorders*. American Psychological Association.

Renshaw, T. L., Eklund, K., Dowdy, E., Jimerson, S. R., Hart, S. R., Earhart, J., & Jones, C. N. (2009). Examining the relationship between scores on the behavioral and emotional screening system and student academic, behavioral, and engagement outcomes: An investigation of concurrent validity in elementary school. *The California School Psychologist, 14*(1), 81–88. https://doiorg.proxy.libraries.rutgers.edu/10.1007/BF03340953.

Riggs, C. G., & Mueller, P. H. (2001). Employment and utilization of paraeducators in inclusive settings. *Journal of Special Education, 35*(1), 54–62. https://doi.org/10.1177/002246690103500106

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Riverside.

Senn, S. (2006). Change from baseline and analysis of covariance revisited. *Statistics in Medicine, 25*(24), 4334–4344. https://doi.org/10.1002/sim.2682

Shapiro, E. S. (1996). *Direct observation manual for the behavioral observation of students in school (BOSS) in academic skills problem workbook*. Guilford.

Shapiro, E. S. (2011). *Academic skills problems* (4th ed.). Guilford: Workbook.

Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools, 41*(5), 551–561. https://doi.org/10.1002/pits.10176

Sheridan, S. M., Bovaird, J. A., Glover, T. A., Garbacz, S. A., Witte, A., & Kwon, K. (2012). A randomized trial examining the effects of conjoint behavioral consultation and the mediating role of the parent–teacher relationship. *School Psychology Review, 41*(1), 23–46. https://doi.org/10.1080/02796015.2012.12087374

Sheridan, S. M., Witte, A. L., Holmes, S. R., Coutts, M. J., Dent, A. L., Kunz, G. M., & Wu, C. (2017). A randomized trial examining the effects of conjoint behavioral consultation in rural schools: Student outcomes and the mediating role of the teacher–parent relationship. *Journal of School Psychology, 61*, 33–53. https://doi.org/10.1016/j.jsp.2016.12.002

Shernoff, E. S., Bearman, S. K., & Kratochwill, T. R. (2017). Training the next generation of school psychologists to deliver evidence-based mental health practices: Current challenges and future directions. *School Psychology Review, 46*(2), 219–232. https://doi.org/10.17105/SPR-2015-0118

Shogren, K. A., Faggella-Luby, M. N., Jik, S., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions, 6*(4), 228–237. https://doi.org/10.1177/10983007040060040401

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*(3), 351–380. https://www.jstor.org/stable/42899983.

Snijders, T. A. B., Bosker, R. J., & Guldemond, H. (1999). PinT (Power in Two-level designs).

Sobeck, E. E., Robertson, R., & Smith, J. (2020). The effects of didactic instruction and performance feedback on paraeducator implementation of behavior support strategies in inclusive settings. *Journal of Special Education, 53*(4), 245–255. https://doi.org/10.1177/0022466919858989

Sobek, E. (2016). *The effects of didactic instruction and performance feedback on paraeducators' use of positive behavior support strategies in inclusive settings (Publication No. 10183750)*. Doctoral dissertation. University of Pittsburgh (ProQuest Dissertations & Theses Global).

Stormont, M., Reinke, W. M., Newcomer, L., Marchese, D., & Lewis, C. (2015). Coaching teachers' use of social behavior interventions to improve children's outcomes: A review of the literature. *Journal of Positive Behavior Interventions, 17*(2), 69–82. https://doi.org/10.1177/1098300714550657

Sugai, G. (2009). Reaching all students: RTI & SWPBS. [PowerPoint slides]. Retrieved from https://slideplayer.com/slide/15018406/.

Sutherland, K. S., Conroy, M. A., McLeod, B. D., Granger, K., Broda, M., & Kunemund, R. (2020). Preliminary study of the effects of best in class–elementary on outcomes of elementary students with problem behavior. *Journal of Positive Behavior Interventions, 22*(4), 220–233. https://doi.org/10.1177/1098300719900318

Tillery, C. Y., Werts, M. G., Roark, R., & Harris, S. (2003). Perceptions of paraeducators on job retention. *Teacher Education and Special Education, 26*(2), 118–127. https://doi.org/10.1177/088840640302600205

U.S. Department of Education, National Center for Education Statistics, Statistics of State School Systems. (2017). Table 213.10: Staff employed in public elementary and secondary school systems, by type of assignment: Selected years, 1949–50 through fall 2015. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of education statistics (2016 ed.)*. http://nces.ed.gov/programs/digest/current_tables.asp.

Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology, 59*(9), 920–925. https://doi.org/10.1016/j.jclinepi.2006.02.007

Vickers, H. S., & Minke, K. M. (1995). Exploring parent-teacher relationships: Joining and communication to others. *School Psychology Quarterly, 10*(2), 133–150. https://doi.org/10.1037/h0088300

Villarreal, V. (2015). *Test review: Schrank, F.A., Mather, N., & McGrew, K.S* (Vol. 2014).

Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*(4), 454–474. https://doi.org/10.1080/02796015.2005.12088009

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). How disruptive students escalate hostility and disorder-and how teachers can avoid it. *American Education, 27*(4), 22–27. https://www.aft.org/periodical/american-educator/winter-2003-2004/how-disruptive-students-escalate-hostility.

Walker, V. L., & Smith, C. G. (2015). Training paraprofessionals to support students with disabilities: A literature review. *Exceptionality, 23*(3), 170–191. https://doi.org/10.1080/09362835.2014.986606

Wiggs, N. B., Reddy, L. A., Bronstein, B., Glover, T. A., Dudek, C. M., & Alperin, A. (2021). A mixed-method study of paraprofessional roles, professional development, and needs for training in elementary schools. *Psychology in the Schools, 58*(11), 2238–2254.

Wiggs, N. B., Reddy, L. A., Glover, T. A., Dudek, C. M., Alperin, A., & Regan, P. (2020). Behavior support coaching for paraprofessionals and students with disruptive behavior disorders: A case study in a high-poverty elementary school. *Journal of Applied School Psychology, 19*(5), 1–25. https://doi.org/10.1080/15377903.2020.1821272

Wilkinson, L. A. (2005). Supporting the inclusion of students with emotional and behavioural disorders: Examples using conjoint behavioural consultation and self-management. *International Journal of Special Education, 20*(2), 73–84.

Woodcock-Johnson, I. V. (2022). Tests of achievement. *Journal of Psychoeducational Assessment, 33*(4), 391–398. https://doi.org/10.1177/0734282915569447

Bosker, R. J., Snijders, T. A., & Guldemond, H. (1999). *PINT (Power IN two-level designs): Estimating standard errors of regression coefficients in hierarchical linear models for power calculations*. User's manual. RU Groningen/UT-NWO Project 500-278-203.

Reynolds, C. R., Kamphaus, R. W., & Vannest, K. J. (2015). *BASC3: Behavior assessment system for children*. PsychCorp.