Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2021, Vol. 29(1) 44–55

DOI: 10.1177/1063426620902721
jebd.sagepub.com
SAGE

Evaluation of a Comprehensive Assessment-Based Intervention for Secondary Students With Social, Emotional, and Behavioral Problems

Lee Kern, PhD[1], Steven W. Evans, PhD[2], Timothy J. Lewis, PhD[3], Talida M. State, PhD[4], Paras D. Mehta, PhD[5], Mark D. Weist, PhD[6], Howard P. Wills, PhD[7], and Nicholas A. Gage, PhD[8]

Abstract

Despite a high dropout rate and poor educational and posteducational outcomes, limited evidence-based practices have been validated with students at the high school level who have emotional and behavioral problems. The purpose of this study was to evaluate the impact of a multitreatment, assessment-based intervention package. Participants were 647 high-school-age students with both emotional/behavioral problems and impairment in school functioning. Using a randomized controlled trial, the intervention group showed significantly fewer disruptions, measured using direct observation. No significant differences were found in directly observed engagement or on standardized measures. However, in general, interventions were implemented with low frequency. Dosage effects were examined for classroom interventions and a social skills intervention (Interpersonal Skills Group [ISG]). Results indicated a dosage effect for ISG, reflected in standardized measures. Although implementation was low, teachers and staff rated practices that they implemented highly favorably.

Keywords

adolescent(s), disorders/disabilities, secondary education, school based, mental health services, adolescent, behavior(s), emotional disturbance

The academic and behavioral challenges of children and youth identified with "emotional disturbance" (hereafter referred to as emotional/behavioral Disorder [EBD]) under the Individuals with Disabilities Education Act (IDEA) have been well documented. Compared with peers, students with emotional and behavioral problems have poor academic performance, high levels of grade retention, low rates of graduation, high frequencies of arrest, and bleak postsecondary education and employment options (Wagner et al., 2005). Moreover, adolescence is a particularly vulnerable time given the increasing emergence of mental health concerns (Merikangas et al., 2010). Several system factors add to the complexity of the aforementioned issues, including reliance on restrictive educational placements, poorly trained or beginning teachers, and the significant underidentification of students as EBD (Forness et al., 2012). Forness et al. (2012) estimated a more accurate prevalence of between 12% and 25%. Regarding underidentification, general educators who receive limited professional development in managing emotional and behavioral concerns of students are left to address students with significant challenges (Wagner et al., 2005).

Despite system and other challenges, there are evidence-based practices that have demonstrated academic and social improvements, including mentoring (e.g., DuBois et al., 2002), classroom-based interventions (Simonsen et al., 2008), training on academic and interpersonal skills (Evans et al., 2016), and mental health interventions (e.g., Hilt-Panahon et al., 2007). Although the extant literature base supporting these interventions is promising, the majority of the work to date has been conducted at the elementary level (Kern et al., 2015). Furthermore, given students' diverse social, emotional, and behavioral needs, particularly at adolescence, research

[1]Lehigh University, Bethlehem, PA, USA
[2]Ohio University, Athens, USA
[3]University of Missouri, Columbia, USA
[4]Montclair State University, NJ, USA
[5]University of Houston, TX, USA
[6]University of South Carolina, Columbia, USA
[7]The University of Kansas, Kansas City, USA
[8]University of Florida, Gainesville, USA

Corresponding Author:
Lee Kern, Lehigh University, Bethlehem, PA 18015, USA.
Email: lek6@lehigh.edu

that delineates ways that multicomponent interventions can be selected and adapted to both target specific student concerns and contextually fit the unique milieu of high school settings is needed. Specifically, intervention packages must include efficient assessment strategies, data-based decision-making, resource planning, and on-going technical assistance across multiple classrooms and mental health providers (Evans & Weist, 2004).

The purpose of this study was to develop and evaluate the impact of a multitreatment intervention, with intervention components guided by a standard set of data-decision rules, on the social, emotional, and academic performance of students with disabilities and those at high risk due to significant behavioral challenges (see Kern et al., 2011). We aimed to ascertain lessons from a large-scale randomized controlled trial that might inform the future development of an individualized intervention package tailored to address both student and contextual variables. The following research questions guided this large-scale study:

Research Question 1: Does an individualized intervention package for secondary students with social, emotional, and behavioral problems improve distal and/or proximal outcomes?
Research Question 2: To what extent is intervention dosage related to student outcomes?
Research Question 3: Do educators who implement interventions rate them as acceptable?

Method

Participating Schools and Students

Schools. Fifty-four high schools within proximity to universities affiliated with the research grant participated in the study, distributed across five states. These included five in Kansas, seven in Missouri, 16 in Ohio, 10 in Pennsylvania, and 16 in South Carolina. The total number of students attending each high school ranged from 482 to 3,141 (M = 1,349; SD = 672). The size of the schools varied, with three smaller than 500 students, 16 with 501 to 1,000 students, 11 with 1,001 to 1,500 students, 16 with 1,501 to 2,000 students, three with 2,001 to 2,500 students, three with 2,501 to 3,000 students, and two with more than 3,000 students. Schools were fairly evenly distributed with respect to community location (as defined by Department of Education), with 21 (39%) suburban, 20 (37%) rural, and 13 (24%) urban. A mean of 31.66% (SD = 28.64%) of the total school population was minority (range = 1.56%–93.42% per school) and a mean of 38.54% (SD = 19.51%) was low socioeconomic status (SES; range = 7%–75% per school).

Student recruitment and eligibility. After schools agreed to participate, school personnel were asked to compile a list of at least 25 students who (a) would attend ninth to 11th grade during Year 1 of the study, and (b) exhibited serious social, emotional, and/or behavioral problems. A total of 852 English-speaking families/students agreed to screening. To assess social/emotional/behavioral functioning, standardized assessments were completed by a parent/guardian, a school staff member most familiar with the student, and the student. Students met criteria for social/emotional/behavioral problems if they had a t score of 60 or higher on either the internalizing or externalizing composites of the Behavior Assessment System for Children–Teacher or Parent Version (BASC; C. R. Reynolds & Kamphaus, 2004), a t score of 60 or higher on the Multidimensional Anxiety Scale for Children (MASC; March et al., 1997), or a t score of 60 or higher on the Reynolds Adolescent Depression Scale 2 (RADS-2; W. M. Reynolds, 2002).

Students were subsequently evaluated to determine whether they met the criteria for impairment in school functioning. Criteria in this area could be met by exhibiting any two of the following: (a) four or more office referrals/behavioral infractions across the semester prior to enrollment or five or more in any month of the current semester, (b) five or more absences (other than illness) or tardiness to class in any month of the current or previous semester, (c) two or more in-school or out-of-school suspensions in the current academic year, or (d) one or more Fs or two or more Ds in any core academic subject in one of the two most recent grading periods.

Students were excluded if diagnosed with autism spectrum disorder or if they had an IQ below 70. This criterion was used because some of the interventions required understanding of concepts that could challenge students with low IQs. If IQ testing had not been conducted within 3 years, the Wechsler Abbreviated Scale of Intelligence (WASI; Stano, 2004) was administered.

Participating students. A total of 647 students met eligibility criteria (see Table 1). The sample consisted of 66.5% male (n = 430) and 33.5% female (n = 217) students. Approximately half (n = 314; 48.5%) had special education labels, including the primary category of learning disability (n = 156; 24% of total sample), emotional disturbance (n = 64; 9.9%), other health impairment (n = 55; 8.5%), and other/not available (n = 15; 2.3%). The majority (52.1%) were White/Caucasian, followed by Black/African American (38.6%). Among the sample, 68.5% (n = 429) received free/reduced lunch, and data were not available for 3.7% (n = 23).

Participating school staff. A range of school staff implemented interventions based on student need. Teachers implemented all classroom interventions, individual student interventions, and some Interpersonal Skills Group (ISG;

Table 1. Student Demographics.

Category	Missing	*n* (*N* = 647)	%
Gender	0		
Male		430	66.5
Female		217	33.5
Ethnicity	0		
White/Caucasian		337	52.1
Black/African American		250	38.6
Hispanic/Latino		34	5.3
Black/White		8	1.2
Biracial		6	0.9
Other		12	1.9
Grade			
8	5; 0.77%	21	3.25
9		295	45.60
10		236	36.48
11		90	13.91
Special education classification			
No	8; 1.2%	325	50.2
Yes		314	48.5
Student has a 504 plan			
No	2; 0.3%	621	96.3
Yes		24	3.7
Neither special education classification nor 504 plan			
None	2; 0.3%	342	52.9
Either/Or		303	46.8
Primary special education classification			
None	0	373	57.7
LD		156	29.0
ED		64	9.9
OHI		55	8.5
Other		15	2.3
Total household yearly income			
Category 1 (US$0–US$20,000)	30; 4.6%	227	35.1
Category 2 (US$20,000–US$40,000)		200	30.9
Category 3 (US$40,000–US$$60,000)		95	14.7
Category 4 (US$60,000–US$80,000)		48	7.4
Category 5 (US$80,000–US$100,000)		21	3.2
Category 6 (US$100,000–US$120,000)		16	2.5
Category 7 (US$120,000–US$140,000)		3	0.5
Category 8 (US$140,000+)		7	1.1
Site			
Lehigh University	0	121	18.7
Ohio University		219	33.8
University of Kansas		66	10.2
University of Missouri		75	11.6
University of South Carolina		166	25.7

Note. LD = learning disabilities; ED = emotional disabilities; OHI = other health impairment.

Evans et al., 2014). School mental health professionals (SMHPs) led mental health interventions and ISG.

Check & connect (C&C) mentors. Mentors were 234 adults: 153 in Year 1 and 156 in Year 2 (75 served both years). The majority (76%) were teachers (20% also served as case managers), 4.26% were counselors/school psychologists, 2.65% were social workers, 2.12% were administrators, and 10.58% were other school-based professionals (e.g., coaches, security officers), with 4.39% unspecified.

Teachers. Teachers who implemented classroom interventions were mostly female (63.1%) and White/Caucasian (90.5%). The grade taught was evenly distributed, with many teaching multiple grades and multiple subjects, the majority of the general education teachers teaching English or Math. Most teachers had taught <10 years (67.4%) and had experience with students with EBD <10 years (44.3%). The majority of classroom interventions was delivered in general education classrooms (n = 240; 71.43%), with fewer in special education (n = 65; 19.35%).

School mental health providers. SMHPs were (a) a school employee with a graduate degree in school counseling, community counseling, school psychology, or social work or (b) a local university graduate student completing a practicum or field placement as part of one of the above degrees. ISG leaders were 21 to 73 years (M = 39.52, SD = 10.30) with 0 to 35 years (M = 6.7, SD = 6.31) in the profession. Most were female (66.1%) and identified as White (78.9%), followed by African American/Black (12.8%), Hispanic/Latino (1.7%), and Asian (1.1%).

Distal Measures

BASC, Second Edition (*BASC-2;* C. R. Reynolds & Kamphaus, 2004) is a norm-referenced behavior rating scale measuring internalizing and externalizing behaviors. The BASC-2 has strong psychometric properties (C. R. Reynolds & Kamphaus, 2004). Teacher Standard scores of Externalizing and Internalizing Problems composites were used as measures of student behavior outcomes, as reported by teacher and parent. The Emotional Symptoms Index composite was used as a measure of students' self-rated behavior. *T* scores of 60 or above generally indicate students are "at risk" of developing clinically significant problems.

Impairment Rating Scale (*IRS*; Fabiano et al., 2006) is a multidimensional scale that assesses student impairment in academic, social, and overall functioning (Evans et al., 2013) and has adequate temporal stability (Fabiano et al., 2006). The rater identifies the student's primary problem behavior and the degree to which it affects the student's performance and rates using a 7-point scale. Parent and teacher versions were used for the study. The parent version has seven items, four items that measure social relationships with peers, siblings, parents, and family and one item each for academic functioning, self-esteem, and overall problem severity. The teacher version has six items, two that measure relationships with peers and teacher, two for academic and classroom performance, one for self-esteem, and one describing overall severity. Higher item scores indicate more impairment.

Disruptive Behavior Disorders Scale (*DBD;* Pelham et al., 1992) measures symptoms associated with Attention-Deficit/Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), or Conduct Disorder (CD). The psychometrics are adequate in adolescent samples (Evans et al., 2013). Both teacher and parent versions were used, which consist of 45 items rated on a 4-point Likert-type scale. The scale can be interpreted using symptom counts or factor scores.

RADS-2 (W. M. Reynolds, 2002) is a 30-item self-report measure designed to assess symptoms associated with depression in adolescents, ranging in age from 11 to 20 years. The overall psychometric properties are strong (W. M. Reynolds, 2002). RADS-2 measures four basic dimensions of depression: Dysphoric Mood, Anhedonia/Negative Affect, Negative Self-Evaluation, and Somatic Complaints. Total *t* scores of 60 or above indicate level of symptoms associated with clinical depression.

MASC (March et al., 1997) is a 39-item self-report assessment of anxiety symptoms for children aged 8 to 18 years with good psychometrics (March et al., 1997; March et al., 1999). The scale provides four main index scores for Social Anxiety, Separation Anxiety, Harm Avoidance, and Physical Symptoms and a total score. *T* scores of 65 or above generally indicate level of symptoms associated with clinical anxiety.

Woodcock Johnson Tests of Achievement, Third Edition (*WJ- III;* Woodcock et al., 2001), a battery of subtests, assesses student achievement in reading, writing, and mathematics. The WJ-III has strong psychometric properties (Woodcock et al., 2001). The Broad Reading standard score (Letter–Word Identification, Reading Fluency, Passage Comprehension subtests) and the Broad Math standard score (Calculation, Math Fluency, and Applied Problems subtests) were tracked at the study start and end.

Proximal Measures

Direct observation data were collected for student active and passive engagement and disruptive behavior. Active engagement was defined as being engaged with instructional content, demonstrated via choral response, raising hand, responding to teacher instruction, writing, reading, or otherwise actively completing an assigned task (e.g., manipulating assigned materials). Passive engagement was defined as passively attending to instruction (i.e., listening or observing) by orienting to teacher, peer, or materials as appropriate. Disruptive behavior was defined as displaying behavior that interrupts or could interrupt the lesson in such a way that it distracts the teacher and/or other students (e.g., out of seat, talking to peer, making derogatory comments) or causes physical disruption (e.g., throwing things, pushing chair or desk).

Direct observations were conducted using the Multiple Option Observation System for Experimental Studies (MOOSES; Tapp et al., 1995) via handheld devices. Prior to

observing, all data collectors (graduate students blind to the purposes of the study) were trained in data collection until they reached minimum 80% agreement across two observations on all codes using videos and in vivo observations. Interobserver agreement was assessed during 25% of observations. If agreement fell below 80%, brief retraining occurred through in vivo observations. Data collectors gathered duration or frequency data on instructional context (e.g., whole group, small group), teacher behavior (e.g., opportunities to respond [OTR], positive feedback, corrections), and target student behavior (e.g., active engagement, downtime, disruptions).

For all eligible and active students in the treatment condition, two to three data points were collected prior to intervention implementation in one targeted classroom. Following the start of intervention, four observations were collected at 4- to 6-week intervals across the school year. One third of students in control schools were randomly selected for comparison of direct observation. Baseline and intervention data were collected on the same schedule as treatment students.

Each observation was 15 min in length and took place in a core academic classroom that teachers identified as problematic. If the student had problems in multiple classrooms, a class that spanned both fall and spring was selected. Otherwise, the class was randomly selected.

Other Measures

School Intervention Rating Form (*SIRF*; Kern & Gresham, 2002) is a 21-item intervention rating scale adapted from the Treatment Acceptability Rating Form–Revised (TARF-R; Reimers et al., 1991). The wording was changed to reflect school-based interventions, and item consistency reflected critical social validity domains (Harrison et al., 2016). Teachers and SMHPs who implemented interventions rated items on a 7-point Likert-type scale (0–6), resulting in a total score ranging from 0 to 126.

Procedures

Students in the intervention group received one or more interventions (individually or class-wide) as indicated by data described below. School staff implemented all interventions. Graduate students in an education-related field served as project facilitators (one in each high school) and facilitated intervention implementation. Facilitators were trained in study (Center for Adolescent Research in the Schools [CARS]) interventions and the coaching process and were regularly supervised by investigators.

Facilitators used a coaching model (Kretlow & Bartholomew, 2010) to instruct and support staff implementing interventions in the following manner. First, interventionists were given a handout with a description of and rationale for the intervention, materials needed, steps for implementation, sample worksheets, helpful tips, and examples. Facilitators then held a brief (e.g., 15 min) meeting with the interventionist to provide an overview of the intervention, review steps, and model or role-play implementation. In some cases, this meeting was conducted with groups (e.g., all C&C mentors in a school). After initial training, procedures described under Intervention Fidelity were followed. The following interventions were used.

C&C. All students in the treatment condition received C&C (Anderson et al., 2004). C&C mentors met weekly with students for approximately 10 to 15 min; reviewed school engagement using a C&C monitoring sheet to assess risk indicators such as absences, behavior infractions, grades, or missing assignments; and implemented a collaborative problem-solving process by selecting from a continuum of strategies to address the problem. Project facilitators trained C&C mentors for approximately 1.5 hr in small group or 1:1 contexts.

ISG. All treatment group participants were enrolled in a weekly ISG (Evans et al., 2014). ISG incorporates developmental skills with training in the cognitive processes thought to contribute to social dysfunction (Sadler et al., 2011). Adolescents are taught to establish goals for how they wish to be perceived by others, identify behaviors likely to promote those perceptions, consider verbal and nonverbal feedback from others, recognize goals may vary by context, and modify their behavior relative to their goals and feedback of others.

Classroom-based interventions. When C&C data indicated a student met a preestablished risk level in a core academic class (e.g., student had four or more behavior referrals, completed <90% of assignments), a facilitator contacted the classroom teacher to assess his or her willingness to implement supports for the student. After securing teacher permission, project facilitators conducted a Classroom Assessment consisting of teacher and student interviews followed by a class-wide observation and one to three observations of target student. Facilitators then used decision rules to select specific class-wide or individual interventions (see below). Class-wide interventions were implemented if a problem pertained to multiple students (e.g., no class-wide expectations with multiple students engaging in problem behavior).

Classroom expectations. This intervention involved establishing three to five positively stated classroom expectations, providing reinforcement to students following expectations, and predetermined consequences for students not following expectations. Classroom structure was implemented in 20 classrooms.

Classroom routines. Routines included one or more of the following: developing a daily agenda, entering the classroom, turning in assignments, and assisting student to track assignments. This intervention was implemented in five classrooms.

Improving student–teacher interactions. This intervention was recommended when more negative than positive interactions were observed and involved increasing the number of praise statements issued class-wide or to individual students, assuring the ratio of positive to negative statements was at least 4:1. This intervention was implemented in 86 classrooms.

Increasing OTR. Providing OTR was recommended when they were not routinely used by the classroom teacher. This intervention consisted of using one of the following with all students in the classroom: guided notes, response cards, computer-assisted instruction, and peer tutoring. OTR was implemented in 43 classrooms.

Individual student classroom interventions. When the classroom assessment indicated few class-wide problems, one of the following individual student interventions was implemented with the CARS study participant, which was prescribed by the assessment.

Organizational skills and self-monitoring. When students completed fewer than 90% of assignments, organizational skills training was recommended. This intervention consisted of one or more activity(ies) to facilitate organization and assignment tracking using a daily planner, missing assignment tracking form, and organizational checklist regarding the student's book bag, binder, or locker (Evans et al., 2016). This intervention was implemented with 109 students.

Study skills. When student's core academic class grade fell below a C over the previous 6 weeks and/or the student or teacher reported lack of study skills or test anxiety, instruction in study skills was recommended. Specific intervention included developing strategic study skills (i.e., use of acrostics, acronyms, and imagery), using flash cards to study, and developing effective test-taking skills. Study skills intervention was implemented with 18 students.

Accommodations. Accommodations were recommended when data demonstrated that contextual features were in place (e.g., expectations, positive teacher–student interactions), but students were struggling academically (e.g., poor grades). Accommodation selection was facilitated by CARS Accommodations Guide that facilitated the selection of accommodations matched to student need. This intervention was implemented with 17 study participants.

Mental health interventions. Selection of school-based mental health interventions relied on an assessment protocol and decision-making algorithm implemented by SMHP (with support from CARS staff). Required elements were student and parent/care-provider interviews with 14 open-ended questions pertaining to core aspects of common adolescent disorders (depression, anxiety, ADHD, CD, and ODD). SMHPs used data from interviews and optional assessments to identify a domain of problem. The algorithm linked the domain of problem to a specific mental health intervention (e.g., depression-related problems linked to Cognitive Behavioral Therapy for Depression [CBT-D]). Decision rules based on the results of interviews and optional assessments and SMHP ratings of feasibility and acceptability guided intervention selection.

CBT. The CBT-manualized programs included many of the core elements of well-studied CBT protocols (e.g., cognitive restructuring, exposure), which were modified to include additional implementation detail and be implemented in high schools. Two CBT programs were implemented, one for anxiety and the other for depression. CBT interventions were implemented with 19 students.

Comparison group intervention. Newsletters from a prior study focusing on wellness (e.g., nutrition, exercise) were refined and disseminated to all comparison schools monthly during the school year throughout the 2 years of the study. Three different versions were disseminated to teachers, SMHPs, and parents. Each newsletter included a brief overview of the monthly topic and everyday life strategies to enhance a particular skill or set of skills.

Intervention Fidelity

Intervention fidelity data were regularly collected throughout the intervention implementation process and across sites using forms describing steps of each intervention (see Kern et al., 2015). Classroom fidelity was first assessed on or near the first day of implementation, with two additional checks conducted within the first 10 days of implementation. Facilitators delivered brief performance feedback (5–10 min) to the teachers after each check. A booster session consisting of detailed feedback (e.g., problem-solving process, retraining of critical steps) was conducted if teachers failed to meet the preestablished fidelity criterion of 80% or above. Fidelity was assessed monthly for the duration of the implementation using direct observation and permanent product data. Mean intervention fidelity for classroom-related interventions (C&C, classroom-based interventions, individual student interventions) was 81.15% (range = 67.46%–95.00% for C&C and classroom interventions; range = 80.05%–99.86% for individual student interventions). Integrity data were collected for 604 of 1,080

(55.93%) ISG sessions. Sessions were considered adherent if designated activities took place during 90% or more of each session. This was met for 64.4% of sessions, with integrity best in later sessions (Phase 1 = 55%; Phase 2 = 75%; Phase 3 = 100%).

Design

The study was designed as a blocked cluster randomized trial (e.g., Murray, 1998). Participating schools were randomly assigned, within state, to either the intervention (n = 27) or comparison (n = 27) condition. For all but one outcome (BASC internalizing), the mean differences between the two conditions were not significant. The magnitude of the mean difference was small, suggesting that the randomization was successful.

Data Analysis

Research Question 1, distal measures. Linear mixed-effects modeling with restricted maximum likelihood (REML) was used to estimate model parameters. Significant proportion of subjects (0.44–0.61) had missing data at posttest. Intention-to-treat (ITT) approach was used for the analysis of distal measures to account for missing data. Multiple imputed data sets were generated using Blimp software (Keller & Enders, 2017). Multiple imputation (MI) was performed via fully conditional specification (FCS; Enders et al., 2018). Mixed-effects models were estimated using SAS Proc Mixed. Parameter estimates were aggregated using SAS Proc MI. For each outcome, treatment effects were estimated after adjusting for student-level pretest score, gender, minority status, and income, along with their corresponding school-level means. Four site dummy variables were also used as covariates to account for within-site randomization to intervention condition. School-level random intercepts were estimated to account for school-level dependency.

Research Question 1, proximal measures. Direct observation of student disruptive behavior collected at two time points (i.e., baseline, intervention) was compared across 137 students in the treatment group and 52 in the control group. We estimated a series of multilevel models (MLMs; Raudenbush & Bryk, 2002) to account for the nesting of students in schools. Due to the small number of sites, we used dummy codes for each site to account for nesting of students and schools within sites instead of a three-level model. The two-level MLMs were fit via R3.2.3 using the lme4 (Bates, 2005) package. We modeled the relationship between the treatment conditions and student behavior at the end of the first school year. The primary predictors in the models included (a) assignment to either treatment or control condition and (b) student behavior during baseline observations. In addition to the dummy codes for sites to control for site-level differences, we modeled student characteristics as covariates to increase the accuracy of the models, including gender, ethnicity, SES, and special education. We estimated models as follows: (a) Model 1: variance components model (null model); (b) Model 2: random intercept model with sites; (c) Model 3: random intercept model with sites and the treatment condition; and (d) Model 4: random intercept model with sites, the treatment condition, baseline disruptive behavior, and all other covariates. Baseline disruptive behavior was grand-mean centered for ease of interpretation; all other predictors were dichotomous.

Research Question 2. Dosage for each student and each intervention was assessed using several sources of data. For C&C, monthly fidelity sheets were used to document when mentoring started, the frequency of weekly meetings, and whether mentoring ended (e.g., student refused to participate). For classroom-based interventions, the start and end date of each intervention was coded using a TeleForm. These data were converted to months of intervention for each student, which served as a proxy for dosage. For ISG, the number of sessions attended was used to measure dosage. Similar linear modeling analysis used in Research Question 1 was used to examine the impact of dosage. For each of the intervention dose variable, the model included the student-level dosage and school-level mean to account for potential contextual effect of dosage. Models included school-level random intercepts to account for clustering.

Research Question 3. Descriptive analyses were conducted in the form of mean rating across teachers or SMHPs on SIRF for each intervention following implementation.

Results

Research Question 1: Distal Measures

Table 2 presents intervention and control group means and variances at pretest and posttest for the sample with data. Intraclass correlations (ICCs) at posttest ranged from .02 to .08. These values are consistent with ICCs reported by Hedges and Hedberg (2007) for educational and mental health outcomes. Unadjusted mean differences between treatment and control condition at posttest for cases with data were small and not significantly different from zero. Adjusted treatment effects based on the ITT analyses were not significant either. Results indicated that the treatment did not have positive change for any of the distal outcomes.

Research Question 1: Proximal Measures

Descriptive statistics and baseline equivalence. The majority of students were male, low SES, in ninth and 10th grades,

Table 2. Treatment Effects for Distal Outcomes.

Outcome	Treatment effect (*SE*)
Student self-report	
BASC: Inattention/Hyperactivity	−0.96 (1.32)
BASC: Internalizing problems	0.46 (1.22)
MASC	0.49 (1.30)
RADS-2	0.19 (1.26)
Parent report	
BASC: Externalizing problems	1.82 (1.17)
BASC: Internalizing problems	0.15 (1.16)
DBD: Impulsivity/Overactivity	−0.04 (0.08)
DBD: Inattention	0.04 (0.10)
DBD: Oppositional/Defiant	0.01 (0.09)
Teacher report	
BASC: Externalizing problems	0.05 (2.07)
BASC: Internalizing problems	−0.80 (1.34)

Note. BASC = Behavior Assessment System for Children–Teacher or Parent Version; MASC = Multidimensional Anxiety Scale for Children; RADS-2 = Reynolds Adolescent Depression Scale 2; DBD = Disruptive Behavior Disorders Scale.

and not receiving special education services. The only characteristic that was statistically significantly different between the groups was ethnicity, with more White students in the treatment group ($\chi^2 = 13.2$, $p = .010$). Therefore, the models included student ethnicity, dummy coded as White and not White, in the models.

We examined the average rate of disruptive behavior and percentage of time students were actively and passively engaged during baseline to establish equivalence between the groups. The average per-minute rate of disruptive behavior during the baseline observation for the treatment group was 0.21 disruptions per minute ($SD = 0.26$) and for the control group was 0.18 disruptions per minute ($SD = 0.26$). The difference between the groups during baseline was $g = .09$ ($p > .05$); therefore, equivalence was established.

The average percentage of time actively and passively engaged during baseline was 32.7% ($SD = 24.4\%$) and 33.6% ($SD = 25.5\%$), respectively, for the treatment group and 52.1% ($SD = 24.1\%$) and 27.2% ($SD = 22.8\%$), respectively, for the control group. The differences between the groups during the winter were $g = -.79$ ($p < .05$) and $g = .26$ ($p < .05$); therefore, equivalence was not established for the two engagement measures.

Treatment effect model results. First, we estimated the null model to calculate the ICC. Results indicate that ~66% of the variance in student disruptive behavior was attributable to school-level differences. Next, we modeled the effect of the research site and found that students in Kansas were significantly more disruptive than at any other site. Subsequently, we modeled the treatment effect controlling for school-level and site-level differences and found a statistically significant treatment effect, with students in the treatment group demonstrating significantly fewer disruptions per minute than students in the control condition (Table 3). Last, we replicated the third model, but included all relevant student-level characteristics. Results remained the same, with students in the treatment schools demonstrating significantly fewer disruptions per minute, even when controlling for disruptions measured during baseline and differences in student ethnicity (Table 4).

We calculated the treatment effect on rates of disruptive behavior using the treatment group standard deviation (0.171), the treatment coefficient (−.132), and the control group mean (0.224) and standard deviation (0.369). Based on this approach, the overall effect of treatment on student disruptive behavior was $g = -0.55$ ($p = .001$), indicating lower rates of behavior were observed during intervention.

Passive and active engagement models. Using the same modeling procedures as used for disruptive behavior, we found that schools accounted for only 3% of the variance in active engagement and 16% of the variance in passive engagement, suggesting minimal differences between schools. Next, we modeled research site and found no significant differences between sites for active or passive engagement. Similarly, we found no differences in active or passive engagement by treatment group. The limited findings precluded the calculation of effect sizes.

Research Question 2: Dose–Response Relationship

There was significant variability in the "dose" of various classroom and school mental health interventions both within and across schools. ICCs for the four interventions ranged from .08 to .26. There were significant site differences in dose for school mental health intervention.

Correlations among the intervention variables ranged from .18 to .51, suggesting that students who received one type of intervention also were likely to be exposed to other types of interventions. Effect of dosage on outcomes was examined using mixed-effects models. School-level random intercepts were estimated to account for clustering. Outcome at posttest was regressed on pretest, adjusting for individual gender, minority status, and yearly family income. Corresponding school-level aggregates (mean pretest, proportion of girls, proportion of minority, and mean income) were used as covariates to account for school-level variability in outcomes. In addition, four fixed dummy covariates were included to account for site-level variability in the outcome. Grand-mean centered dosage variables (ISG number of sessions, C&C, student-focused classroom intervention, and teacher-focused classroom intervention)

Table 3. Multilevel Regression Estimates of Rate of Disruptive Behavior (Equivalence Sample).

	Model 1		Model 2		Model 3		Model 4	
Effects	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed effects								
Intercept	**0.154*****	0.039	0.018	0.104	0.128	0.107	0.138	0.108
Site 1			0.088	0.123	0.082	0.114	0.084	0.113
Site 2			0.103	0.155	0.067	0.146	0.071	0.140
Site 3			**0.444****	0.125	**0.429****	0.115	**0.348****	0.116
Site 4			0.049	0.114	0.027	0.106	0.008	0.104
Treatment					**−0.140***	0.059	**−0.132***	0.061
Prior disruptions							**0.183*****	0.047
Gender							0.010	0.022
Special education							0.009	0.025
White							0.002	0.025
SES							0.000	0.024
Random effects								
Level 1 error	0.045		0.024		0.019		0.020	
Between school variance	0.023		0.023		0.024		0.016	
Fit								
ICC	.662		.511		.442		.556	
AIC	−81.933		−95.693		−99.774		−142.160	
BIC	−72.354		−73.342		−74.230		−101.550	
Deviance	−87.933		−109.693		−115.77		−168.160	

Note. 180 students, 33 schools. SES = socioeconomic status; ICC = intraclass correlation; AIC = Akaike information criterion; BIC = Bayesian information criterion.
Significant estimates are in bold with *$p < .05$, **$p < .01$, and ***$p < .001$.

Table 4. Multilevel Regression Estimates of Rate of Disruptive Behavior (Full Available Sample).

	Model 1		Model 2		Model 3	
Effects	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed effects						
Intercept	**0.143*****	0.024	0.045	0.065	0.098	0.066
Site 1			0.070	0.080	0.062	0.076
Site 2			0.066	0.075	0.054	0.071
Site 3			**0.382*****	0.085	**0.038*****	0.080
Site 4			0.048	0.074	0.039	0.071
Treatment					**−0.087***	0.037
Random effects						
Level 1 error	0.027		0.016		0.014	
Between school variance	0.031		0.031		0.031	
Fit						
ICC	.469		.340		.311	
AIC	−126.75		−146.65		−150.68	
BIC	−115.02		−119.28		−119.39	
Deviance	−132.75		−160.65		−166.68	

Note. 369 students, 54 schools. ICC = intraclass correlation; AIC = Akaike information criterion; BIC = Bayesian information criterion.
Significant estimates are in bold with *$p < .05$, **$p < .01$, and ***$p < .001$.

Table 5. Results of School Staff Ratings on School Intervention Rating Form.

Intervention	Number	*M*	*SD*	Item mean
Accommodations	18	79.77	12.02	4.43
CBT Anxiety	2	90.00	1.41	5.0
CBT Depression	2	67.00	12.73	3.72
CBT Mood	9	82.00	13.09	4.55
Check & Connect	402	80.87	12.95	4.49
De-escalation	1	—	—	
Expectations	20	83.4	9.69	4.63
ISG group	64	70.95	13.20	3.94
OTR	23	85.69	12.97	4.76
Organizational skills	98	82.61	11.83	4.59
Positive student/teacher interaction	62	86.61	11.83	4.80
Routines	4	83.00	7.26	4.61
Self-management	7	87.57	15.39	4.86
Study skills	14	80.86	15.33	4.49

Note. CBT = cognitive behavioral therapy; ISG = Interpersonal Skills Group; OTR = opportunities to respond.

along with corresponding school-level aggregates were used as predictors of the outcome.

For student report of personality (SRP) on the BASC (inattention/hyperactivity subscale [IHS]), the school mean of the number of ISG sessions was significant, $b = -1.28, t(9.75) = -2.33, p < .05$. As the number of ISG sessions was grand-mean centered, the effect of the school mean represents the contextual effect of ISG. This means that for every unit increase in the school mean of ISG, BASC-SRP-IHS decreases by −1.28 units over and above the student-level effect of number of ISG sessions. Student-level effect of the number of ISG sessions was positive, small (.05), and not significantly different from zero. This suggests that schools that implement more ISG sessions have a greater reduction in student-reported inattention/hyperactivity. For student report of anxiety (MASC), the within-school effect of the number of ISG sessions approached significance, $b = -0.31, t(86.84) = -1.77, p < .08$. This means that each additional ISG session that a student attended resulted in a 0.3 unit decrease in student-reported anxiety after controlling for the effect of the mean level of ISG sessions. For student report of depression (RADS), the within-school effect of the number of ISG sessions approached significance, $b = -0.24, t(144.91) = -1.77, p < .08$. This means that each additional ISG session that a student attended resulted in a 0.24 unit decrease in student-reported depression.

Research Question 3

Table 5 presents a summary of educator-reported acceptability of interventions. Mean ratings were generally high, with most interventions rated above 4 on a 0–6 scale. All mean item ratings across the interventions fell between ratings of "liked some" to "liked a lot."

Discussion

Depending on how behavioral change was measured, differential intervention effects were found. Specifically, using a proximal measure (direct observation), significant group differences resulted, with the subsample (29%) demonstrating less classroom disruptive behavior than those in the control group with a moderate effect size (g= −.55). Baseline rates indicated one act of disruptive behavior approximately every 5 min, and the effect size suggests that the average response to the intervention reduced this to approximately one disruptive behavior every 15 min. This difference is both statistically significant and meaningful to teachers, although continued occurrence of disruptive behavior remains problematic. There were some demographic differences between the subsample and entire sample as those in the subsample were more likely to come from the lowest income group and less likely to be in special education. Nevertheless, moderate reductions in disruptive behavior represent an advance in the current literature as there is very little research in classroom management in high schools.

Significant group differences did not emerge using distal measures with the entire sample. These outcomes underscore the importance of direct observation, despite the required labor intensity. Such data are particularly important when developing multicomponent interventions and when adapting or modifying interventions for specific students, contexts, and teachers. For instance, direct observations in the current study dictated whether intervention would be implemented at the class-wide or individual level. Furthermore, directly observed outcomes for students in this subsample were limited to those students for whom intervention was planned, whereas many students in the treatment group did not receive intervention. These reasons

may explain the differences between findings using proximal and distal measures.

A significant issue in the current study pertained to intervention implementation. Despite agreeing to do so, many high school teachers and staff were reluctant to implement interventions, even when data indicated need, resulting in large variability across treatment students. This occurred even though the development phase involved extensive vetting of interventions reported as feasible and acceptable, as well as the provision of extensive training, coaching, and intervention materials during the randomized controlled trial. Furthermore, teachers and SMHPs who implemented interventions overwhelmingly rated them favorably. Limited implementation suggests the need to better educate high school teachers about students with diverse needs and effective interventions. This outcome also may portent the need for a greater focus on matching interventions to the skills, beliefs, and capacity of each provider. For example, one of the principal features in promoting high implementation fidelity at the secondary level is the simplicity and contextual fit of the intervention (Bauman et al., 1991). Despite the clear call for multicomponent interventions that involve several implementers, components themselves may need to vary in response to critical implementation features present in each high school context. Additional research is needed to fully understand resistance to implementation.

Related to the aforementioned concern, the finding that there was a dosage effect for ISG has implications for the potency of that treatment and as a potential marker for the level of interventions provided at each school. Previous research has revealed benefits for ISG as part of a comprehensive treatment for middle school students with ADHD (Schultz et al., 2017). These ISG dosage findings add evidence of the potential efficacy of this group intervention and suggest the need to identify contextual variables that predict sustained implementation such that students receive a sufficient dosage to occasional meaningful effects. The dosage effect for ISG was at the school level and not at the student level possibly because delivery of ISG was consistent within school, so there was little variance between students within schools. Because ISG was provided to all students, a prescribing process was not used, which might have accounted for greater implementation of ISG compared with other mental health interventions. Staff may have been apprehensive about prescribing services at the individual level, resulting in these differences.

There are important limitations that warrant attention. First, the lack of intervention implementation limits conclusions that can be drawn from this study. Despite extensive support provided to school staff, there was a resistance to (a) initiating interventions that differed from those previously used, even though recommendations were aligned with the students' Individualized Education Programs (IEPs); (b) providing support for students with emotional and behavioral problems due to expectations of independence for this age group; and (c) altering administrative task responsibilities (e.g., scheduling, college advising) of school mental health staff to provide interventions to this group of high-risk students. As noted above, contextual variables at the practitioner and school level need to be studied to a far greater extent.

A second major limitation involves the relatively small subsample that was considered for the proximal variables analyses. Due to feasibility, it was not possible to directly observe all participants. Nevertheless, the findings provide a starting point for further research on intervention in high schools and point to the need for feasible and effective data collection strategies because they are critical for developing and adapting intervention packages (Lewis et al., 2014). Despite the limitations and barriers, this study offered some evidence for the effectiveness of classroom interventions and ISG. Among the interventions available to the educators and SMHPs, these were implemented with the greatest frequency. Continued research on both multicomponent interventions and the context in which they are implemented is critically important.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324C080006 awarded to Lehigh University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Anderson, A., Christenson, S., Sinclair, M., & Lehr, C. (2004). Check and connect: The importance of relationships for promoting engagement with school. *Journal of School Psychology*, *42*, 95–113. https://doi.org/10.1016/j.jsp.2004.01.002

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27–30.

Bauman, L. J., Stein, R. E. K., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, *19*, 619–639.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, *30*, 157–197.

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, *23*, 298–317.

Evans, S. W., Brady, C. E., Harrison, J. R., Bunford, N., Kern, L., State, T. & Andrews, C. (2013). Measuring ADHD and ODD symptoms and impairment using high school teachers' ratings. *Journal of Clinical Child and Adolescent Psychology*, *42*(2), 197–207.

Evans, S. W., Langberg, J. M., Schultz, B. K., Vaughn, A., Altaye, M., Marshall, S. A., & Zoromski, A. K. (2016). Evaluation of a school-based treatment program for young adolescents with ADHD. *Journal of Consulting and Clinical Psychology*, *84*, 15–30.

Evans, S. W., Schultz, B. K., & DeMars, C. E. (2014). High school based treatment for adolescents with ADHD: Results from a pilot study examining outcomes and dosage. *School Psychology Review*, *43*, 185–202.

Evans, S. W., & Weist, M. (2004). Implementing empirically supported treatments in schools: What are we asking? *Clinical Child and Family Psychology Review*, *7*, 263–267.

Fabiano, G., Pelham, W., Jr., Waschbusch, D., Gnagy, E., Lahey, B., Chronis, A., . . . Burrows-MacLean, L. (2006). A practical measure of impairment: Psychometric properties of the Impairment Rating Scale in samples of children with ADHD and two school-based samples. *Journal of Clinical Child & Adolescent Psychology*, *35*, 369–385.

Forness, S. R., Kim, J., & Walker, H. M. (2012). Prevalence of students with EBD: Impact on general education. *Beyond Behavior*, *21*, 3–10.

Harrison, J. R., State, T. M., Evans, S. W., & Schamberg, T. (2016). Construct and predictive validity of social acceptability: Scores from high school teacher ratings on the School Intervention Rating Form. *Journal of Positive Behavior Interventions*, *18*, 11–123.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87.

Hilt-Panahon, A., Kern, L., & Divatia, A. (2007). School-based interventions for students with or at-risk for depression: A review of the literature. *Advances in School Based Mental Health Promotion*, *1*, 32–41.

Keller, B. T., & Enders, C. (2017). *Blimp software manual* (Version Beta 6.7). Los Angeles, CA.

Kern, L., Evans, S. W., & Lewis, T. J. (2011). Description of an iterative process for intervention development. *Education and Treatment of Children*, *34*, 593–617.

Kern, L., Evans, S. W., Lewis, T. J., State, T. M., Weist, M. D., & Wills, H. P. (2015). CARS comprehensive intervention for secondary students with emotional and behavioral problems: Conceptualization and development. *Journal of Emotional and Behavioral Disorders*, *23*, 195–205. https://doi.org/10.1177/1063426615578173

Kern, L., & Gresham, F. (2002-2007). *Research exploring alternatives for children in schools (REACH)*. Office of Special Education Programs, National Center Grant funded by the Department of Education.

Kretlow, A., & Bartholomew, C. (2010). Using coaching to improve fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, *33*, 279–299.

Lewis, T. J., Scott, T. M., Wehby, J. H., & Wills, H. P. (2014). Direct observation of teacher and student behavior in school settings: Trends, issues, and future directions. *Behavioral Disorders*, *39*, 190–200.

March, J. S., Parker, J., Sullivan, K., Stallings, P., & Conners, C. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 554–565.

March, J. S., Sullivan, K., & Parker, J. D. (1999). Test-retest reliability of the Multidimensional Anxiety Scale for Children. *Journal of Anxiety Disorders*, *13*, 349–358.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 980–989.

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

Pelham, W. E., Gnagy, E., Greenslade, K., & Milich, R. (1992). Teacher ratings of DSM III-R symptoms for the Disruptive Behavior Disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 210–218.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE.

Reimers, T. M., Wacker, D. P., & Cooper, L. J. (1991). Evaluation of the acceptability of treatments for their children's behavioral difficulties: Ratings by parents receiving services in an outpatient clinic. *Child & Family Behavior Therapy*, *13*, 53–71.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children, second edition (BASC-2)*. AGS Publishing.

Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale: Professional manual* (2nd ed.). Psychological Assessment Resources.

Sadler, J. M., Evans, S. W., Schultz, B. K., & Zoromski, A. K. (2011). Potential mechanisms of action in the treatment of social impairment and disorganization with adolescents with ADHD. *School Mental Health*, *3*, 156–168.

Schultz, B. K., Evans, S. W., Langberg, J. M., & Schoemann, A. M. (2017). Outcomes for adolescents who comply with long-term psychosocial treatment for ADHD. *Journal of Consulting and Clinical Psychology*, *85*, 250–261.

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

Stano, J. F. (2004). Wechsler abbreviated scale of intelligence. *Rehabilitation Counseling Bulletin*, *48*, 56.

Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments and Computers*, *27*, 25–31.

Wagner, M., Kutash, K., Duchnowski, A., Epstein, M., & Sumi, W. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavior Disorders*, *13*, 79–96.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Riverside Publishing.