



Pilot Evaluation of the POWER Program: Positive Outcomes with Emotion Regulation

Brittany Zakszeski¹ · Michelle Cain² · Katie Eklund³ · Lissy Heurich¹ · Reagan Friedman¹ · Ashleigh Ward¹ · Jingwen Zhou¹

Accepted: 7 February 2024
© The Author(s) 2024

Abstract

The Positive Outcomes With Emotion Regulation (POWER) Program is a transdiagnostic intervention for adolescents at risk of developing emotional disorders. The POWER Program was designed to be implemented in secondary schools, by school personnel with or without specialized mental health training, as a Tier 2 intervention. In this pilot study, the POWER Program was implemented by school psychologists and school psychologists-in-training and evaluated across four focal student participants using a multiple-baseline-across-participants single-case design. Program efficacy was assessed using systematic direct classroom observations of student negative affect and social engagement as well as student and caregiver ratings of emotional and behavioral symptoms. Program usability was assessed through rating scales completed by intervention facilitators and student participants. Overall, results provide evidence of the POWER Program's small- to large-sized effects on students' emotional and behavioral functioning as observed in the classroom and self-reported by students. In addition, results suggest implementation facilitators' and students' positive impressions of the program, evident in ratings of high understanding, feasibility, and acceptability across groups. Study limitations are highlighted with attention to opportunities to further refine and evaluate the POWER Program.

Keywords School mental health · Secondary students · Tier 2 · Emotion regulation

Even prior to the onset of the COVID-19 pandemic, the USA faced a public health crisis due to the number of youth with mental health needs and a limited availability of providers to address these needs (United States Department of Health & Human Services, 2020). Schools have continued to increase their provision of mental health assessment and intervention services across the twenty-first century (National Center for Education Statistics, 2022). Yet many schools grapple with how to deliver mental health services most efficiently and effectively, particularly in the context of provider shortages (National Association of School Psychologists [NASP], 2021).

Multi-tiered systems of support (MTSS) frameworks guide schools in allocating prevention and intervention practices of various intensities according to students' demonstrated needs (Eber et al., 2017). Within this framework, primary prevention practices are provided to all students at Tier 1, and universal screenings of students' social, emotional, and behavioral risk are conducted to identify students to receive targeted supports at Tier 2. Targeted interventions are delivered to some students, those identified as at risk in a domain, to ameliorate early-onset symptoms, mitigate risk factors, and boost protective factors (Mitchell et al., 2011). For example, Check-In/Check-Out (CICO) is a widely implemented Tier 2 intervention designed for students with externalizing behaviors that includes three primary components: (1) morning check-in meetings with an adult mentor to review self-management strategies and set a daily behavior goal; (2) adult attention, prompts, and behavioral feedback provided throughout the day using a daily progress report; and (3) afternoon check-out meetings with an adult mentor to review performance and discuss goal attainment (Crone et al., 2004; Todd et al., 2008). Historically, more attention has been placed and options have

✉ Brittany Zakszeski
bnz@udel.edu

¹ School of Education, University of Delaware, 16 W. Main St., Newark, DE 19716, USA

² Christina School District, 1899 S. College Ave., Newark, DE 19702, USA

³ School of Education, University of Wisconsin–Madison, 1025 W. Johnson St., Madison, WI 53706, USA

been provided for Tier 2 interventions addressing conduct-related concerns as opposed to emotion-related concerns (McIntosh et al., 2014).

Targeted School-Based Interventions for Emotional Risk

Recently, school-based Tier 2 programs targeting emotional risk have been developed and evaluated for elementary and middle school populations. The Calm Cat Program (Zakszeski et al., 2023) was designed for and evaluated with early elementary students. The Calm Cat Program includes small-group instruction sessions, which use a behavioral skills training approach (Dib & Sturmey, 2012) to teach relaxation strategies, and adult-mediated mentoring and self-management support, which follow an adapted CICO protocol. Both the Courage and Confidence Mentor Program (CCMP; Cook et al., 2015; Fiat et al., 2017) and the Resilience Education Program (REP; Allen et al., 2019; Eklund et al., 2021; Kilpatrick et al., 2021) were designed for and evaluated with upper elementary and middle school students. Both programs include cognitive-behavioral instruction (in CCMP, individual sessions delivered at intervention initiation; in REP, small-group sessions delivered across multiple weeks) as well as daily mentoring within an adapted CICO approach. All three programs have demonstrated efficacy in decreasing students' emotion-related concerns (Allen et al., 2019; Cook et al., 2015; Eklund et al., 2021; Fiat et al., 2017; Kilpatrick et al., 2021; Zakszeski et al., 2023).

By contrast, few intervention options with demonstrated efficacy and ecological validity for school-based, Tier 2 implementation are available for high school populations (Cilar et al., 2020; O'Reilly et al., 2018), in which mental health needs are both prevalent and increasing in prevalence (Centers for Disease Control & Prevention, 2023). To support schools in addressing this public health crisis, evidence-based intervention programs must be made available and accessible to all schools, including schools with limited resources (e.g., to purchase curricula, to fund specialized providers). These points suggest the need to develop and evaluate a widely accessible intervention program, one that includes the evidence-based techniques found to be effective with other populations but tailored to the unique adolescent developmental stage and high school implementation context.

Features of Targeted School-Based Interventions

For a targeted high school intervention to be adopted and its implementation sustained, it must evidence feasibility and contextual fit, particularly given prominent barriers to service delivery in high school settings. High schools commonly host barriers to mental health intervention implementation as a function of (a) their size (i.e., having many students and faculty can complicate communication and lead to diminished oversight and ownership of initiatives); (b) their organizational culture (e.g., emphasis placed on academic achievement can lead to limited buy-in for social, emotional, and behavioral well-being initiatives); and (c) the developmental level of their students (i.e., students generally value independence and peer relationships, such that adult-directed and -mediated techniques may flounder; Martinez et al., 2019). Other logistical constraints, such as those related to intervention scheduling and personnel availability, are common barriers to Tier 2 implementation across grade levels (Kern et al., 2017). Given national educational professional (United States Department of Education, 2022) and mental health service provider shortages (United States Department of Health & Human Services, 2020), it may be especially critical to identify and develop mental health services that can be delivered by personnel without specialized mental health training (Colizzi et al., 2020).

For such a program to be effective, it must meaningfully engage high school students (Fusar-Poli, 2019; Tylee et al., 2007), with whom other available programs may lack developmental appropriateness and relevance. "Engagement" is defined in the current study as attendance in scheduled intervention activities (e.g., small-group meetings, individual services) as well as sustained attention, on-task behavior, and participation in such activities. The World Health Organization (WHO, 2012) emphasizes the importance of designing and implementing "adolescent-friendly" health services, describing five dimensions of adolescent-friendly health services: (1) *equitable*, (2) *accessible*, (3) *acceptable*, (4) *appropriate*, and (5) *effective*. Specifically, the WHO (2012) clarifies that beyond making services available (*accessible*) to all (*equitable*), care must be taken to ensure services are *acceptable* to youth, such that youth are willing to participate in them; *appropriate* for youth, such that they meet the needs that youth prioritize; and *effective* for youth, such that they meet the needs prioritized by youth in a way that leads to positive improvements in health, well-being, and quality of life. These dimensions must guide the development, evaluation, and refinement efforts of any youth-facing health

programs, including school-based Tier 2 interventions for emotional risk.

For such a program to be efficient, it must be flexible enough to account for the range of emotional risk-related presentations evident in adolescent populations, which have been found to be highly variable across the high school years (Moore et al., 2019). Historically, mental health services have constituted time-limited, group-based interventions that are problem-specific—for example, comprising separate intervention protocols for needs associated with anxiety, depression, and traumatic stress, respectively (Clifford et al., 2020). Although this problem-specific (i.e., “focal treatment”) approach draws upon empirical foundations and has been found to produce positive outcomes, it is one that has been adapted from treatment settings with fewer barriers to service delivery, such as in relation to the synchronization of students’ schedules and the availability of qualified personnel (Eiraldi et al., 2015). In addition, although this problem-specific approach boasts face validity, it may be an inefficient model for service delivery, generally speaking, given the extent to which mental health disorders (a) evidence high rates of comorbidity across childhood and adolescence, (b) commonly share a small set of risk factors and underlying biopsychosocial processes, and (c) are commonly addressed through treatment approaches with overlapping elements and strategies (Ehrenreich-May & Chu, 2014; Kazdin, 1990; McMahon et al., 2003; Sherman & Ehrenreich-May, 2020; Weiss et al., 2003). Atop these important considerations are the well-documented implementation challenges associated with focal treatments, even by practitioners in clinical settings (Weisz et al., 2017). Thus, to maximize efficiency, school-based interventions might adopt a *transdiagnostic* approach by adopting techniques applicable to a range of emotional disorders (e.g., Martinsen et al., 2019).

Positive Outcomes with Emotion Regulation: The POWER Program

A partnership between the Mental Health Technology Transfer Center Network and Devereux Center for Effective Schools led to the development of a manualized program designed to meet the following criteria: (a) suitable for school-based, Tier 2 implementation; (b) feasible for implementation by school-based providers without advanced mental health training (e.g., educators); (c) developmentally appropriate for high school students; and (d) targets emotion regulation as a common process underlying diverse mental health needs (including “internalizing” and “externalizing” concerns; Sloan et al., 2017) and supporting wellness within a dual-factor approach to mental health (Greenspoon & Saklofske, 2001). School psychologists and counselors with

experience working in and consulting with urban, suburban, and rural secondary schools comprised the development team. Through a series of meetings, the development team specified the overall objectives of the program; adopted a program format, scope and sequence, and session format; and identified empirically supported techniques to embed within group sessions. Development team members independently drafted session plans and program materials, which were then reviewed by additional members and revised by the team leader to ensure consistency and adherence to therapeutic principles. Program materials were pilot-tested by a small sample of school-based professionals and their middle and high school students. Feedback obtained on surveys and in focus groups was used to expand upon instructional content, revise instructional activities, and provide recommendations for differentiation.

The Positive Outcomes with Emotion Regulation (POWER) Program (Zakszeski et al., 2022) was designed to promote emotion regulation competencies for students who present with emotion-related symptoms and are identified as at risk for developing emotional disorders. The POWER Program combines two components shared by other school-based Tier 2 interventions targeting risk for mental health needs (albeit for alternative student age groups; e.g., Allen et al., 2019; Cook et al., 2015; Zakszeski et al., 2023): (1) group-based skill instruction and (2) staff mentoring.

Program Components

Skill instruction occurs across seven weekly 40-min core group meetings with three to six students per group following the scope and sequence as presented in Table 1. Group activities are facilitated using a scripted manual with audiovisual materials and student-facing worksheets and handouts. Intervention techniques embedded within the group meeting plans are consistent with a transdiagnostic approach and include motivational interviewing; behavioral skills training; and cognitive instruction, defusion, and restructuring (consistent with cognitive-behavioral therapeutic [CBT] and acceptance and commitment therapy [ACT] approaches). Specifically, the POWER Program teaches the following emotion-related strategies: emotion identification, relaxation strategies (i.e., deep breathing, muscle relaxation, guided imagery), taking time, behavioral activation, observing thoughts, and Catch–Check–Choose (an adaptation of the Catch–Check–Change technique [e.g., Creed et al., 2011] in which individuals choose to either change or let go of unrealistic or unhelpful thoughts). A standard sequence is used across group meetings: The facilitator directs students to complete their emotion ratings for the past week, reviews the group meeting expectations and agenda, guides practice of a relaxation strategy, leads discussion of the Weekly Challenge (i.e., the home practice assignment from the last group

Table 1 POWER program group meeting scope and sequence

	Group Meeting	Learning Objectives
	Core Meetings	Students will...
1	Introduction to the POWER Program	Describe the purpose of the POWER Program List norms for group meetings Explain how everyone encounters stress and negative emotions
2	Recognizing and Labeling Emotions	List and describe different types of emotions Develop an individual goal related to an emotion Describe external clues to use to determine when others are feeling an emotion Describe internal and external clues to use to determine when one is feeling an emotion
3	Relaxation and Behavioral Activation	Classify emotions as pleasant or unpleasant and activating or deactivating Explain why relaxation strategies are used to manage unpleasant, activating emotions and behavioral activation strategies are used to manage unpleasant, deactivating emotions List and independently demonstrate the steps to relaxation strategies List and independently demonstrate positive replacement behaviors for behavioral activation
4	Understanding and Observing Thoughts	Describe how emotions, thoughts, and behaviors influence one another Explain the rationale for using the Observing Thoughts strategy List and independently demonstrate the steps to using the Observing Thoughts strategy
5	Catch–Check–Choose, Part 1	Explain the rationale for using the Catch–Check–Choose strategy List the steps to the Catch–Check–Choose strategy
6	Catch–Check–Choose, Part 2	Independently demonstrate the steps to the Catch–Check–Choose strategy
7	Planning for Success	Independently demonstrate the steps to emotion identification, relaxation, behavioral activation, Observing Thoughts, and Catch–Check–Choose Develop a plan to support use of these strategies

meeting), introduces and models a new strategy, facilitates practice of and discussion around the new strategy, describes the new Weekly Challenge, and concludes the meeting.

Enhanced POWER Program supports are provided through booster group meetings, individual meetings using a motivational interviewing protocol, and staff mentoring through daily Check-Out meetings between the student and POWER Program facilitator or another adult mentor familiar with POWER Program strategies. Check-Out meetings are intended to build relationships between students and adult mentors through which mentors support students with processing emotions, practicing skills, and problem-solving challenges. In Check-Out meetings, mentors greet the student; direct the student to complete their daily emotion monitoring log; summarize the student's emotion ratings; support the student in processing the factors that impacted their day and planning for future challenges; review and practice strategies, as needed; and end the interaction on a positive, encouraging note.

Mechanisms of Change

Through direct instruction in emotion-related strategies and enhanced adult support in applying these strategies,

the POWER Program was designed to increase students' regulation (i.e., upregulation or downregulation) of emotional symptoms that present physiologically (e.g., fatigue, increased heart rate), cognitively (e.g., rumination, automatic thoughts or attributions), and behaviorally (e.g., negative affect, social withdrawal) across a range of mental health disorders and subthreshold risk conditions. Emotional symptom modulation, in turn, is expected to increase students' subjective well-being and support their success in various educational domains. This study represents the first attempt to rigorously evaluate the POWER Program's promise for demonstrating such effects.

Study Objectives

The present study evaluated the efficacy and usability of the core group meetings of the POWER Program, a transdiagnostic Tier 2 program for high school students. Employing a multiple-baseline-across-participants single-case design, which is commonly used in school-based intervention research (Ledford & Gast, 2018), this study was designed according to What Works Clearinghouse (2022) standards to provide high-quality evidence of the POWER Program's promise for improving student

outcomes. This study addresses the following research questions:

1. To what extent is high school students' participation in the POWER Program associated with decreases in negative affect as well as increases in social engagement, as measured through direct classroom observations?
2. To what extent is high school students' participation in the POWER Program associated with decreases in emotional symptoms, as measured through ratings provided by students and their parents?
3. What is the usability of the POWER Program, as rated by intervention facilitators and high school student participants?

Method

Setting and Participants

This study occurred in a public high school in the Mid-Atlantic region of the USA during the winter and spring of 2023. As a part of this school's Tier 2 system of social, emotional, and behavioral (SEB) support, the MTSS team adopted the POWER Program to implement in small groups of students identified as at risk for emotional concerns. One school psychologist, one school psychology intern, and one school psychology practicum student served as POWER Program facilitators. These facilitators were selected given their role in the school during the study and their availability to implement program activities. Although all three facilitators had some training in mental health, training and professional experiences varied (e.g., with years of school-based experience ranging from 1 to 10).

Student participants were identified by the MTSS team through review of Pupil Attitudes toward Self and School (PASS; GL Education, 2023) screening results, intervention referrals, attendance and academic records, and responses to previously implemented SEB interventions. Students were considered eligible for POWER Program participation if they (a) screened below the 20th standardized percentile on the PASS Self-Regard as a Learner scale or were referred by school personnel for intervention due to internalizing behaviors; (b) were classified by the MTSS team as demonstrating a profile most consistent with the POWER Program relative to other Tier 2 intervention offerings; (c) were not receiving supports through a behavior intervention plan or individual school-based counseling, given that it would be difficult to discern what effects are attributable to the POWER Program versus other SEB interventions; (d) did not have an educational classification of intellectual

disability or autism; and (e) provided informed assent as well as parental consent for study participation. The school psychologist contacted families via email or phone to discuss the opportunity in addition to meeting with prospective student participants individually to share information about the program. Four groups of three to four students were formed, with one student per group serving as the focal student, whose data were used in the single-case design to evaluate the impact of the intervention.

Student A was a 15-year-old ninth-grader who identified as male. At baseline, Student A and his mother provided ratings on the Behavior Assessment System for Children, 3rd Edition (BASC-3; Reynolds & Kamphaus, 2015; see “Measures”). On the Self-Report of Personality (SRP) form, Student A's ratings resulted in composite scores in the Typical range for School Problems ($T=58$); in the At-Risk range for Inattention/Hyperactivity ($T=63$) and Personal Adjustment ($T=34$); and in the Clinically Significant range for Internalizing Problems ($T=71$) and the Emotional Symptoms Index ($T=71$). On the Parent Rating Scales–Adolescent (PRS-A) form, his mother's ratings resulted in composite scores in the Typical range for Adaptive Skills ($T=46$), Internalizing Problems ($T=58$), and the Behavioral Symptoms Index ($T=59$) and in the At-Risk range for Externalizing Problems ($T=61$). Student A's mother described his strengths as being strong-willed and open-minded and noted a concern regarding his regulation of anger.

Student B was a 14-year-old ninth-grader who identified as female. Student B noted concerns regarding symptoms of anxiety, though her ratings on the SRP resulted in composite scores all in the Typical range (Personal Adjustment, $T=45$; School Problems, $T=48$; Inattention/Hyperactivity, $T=51$; Internalizing Problems, $T=51$; Emotional Symptoms Index, $T=57$). Likewise, Student B's father noted concern for her developing an anxiety disorder and provided ratings on the PRS-A resulting in composite scores in the Typical range (Adaptive Skills, $T=49$; Behavioral Symptoms Index, $T=50$; Externalizing Problems, $T=51$; Internalizing Problems, $T=55$). The only T -score in the At-Risk range on either form was for the Anxiety subscale on the PRS-A ($T=62$). Student B's father noted her strengths as including being level-headed and emotionally intelligent.

Student C was a 16-year-old tenth-grader who identified as male. Student C's SRP composite scores were in the Typical range for School Problems ($T=58$); in the At-Risk range for Inattention/Hyperactivity ($T=65$); and in the Clinically Significant range for Personal Adjustment ($T=24$), Internalizing Problems ($T=79$), and the Emotional Symptoms Index ($T=79$). His mother provided PRS-A ratings in the Typical range for all composite scores (Adaptive Skills, $T=56$; Internalizing Problems, $T=40$;

Behavioral Symptoms Index, $T=43$; Externalizing Problems, $T=44$), describing Student C as having strengths related to identifying and discussing emotions but struggling academically and in relation to motivation at school.

Student D was a 15-year-old tenth-grader who identified as male. Student D's SRP scores must be evaluated with caution due to an elevated F index, which suggests either very high levels of distress and impairment or an overly negative depiction of symptoms. His composite scores were in the Typical range for School Problems ($T=46$) and in the Clinically Significant range for Personal Adjustment ($T=24$), Inattention/Hyperactivity ($T=80$), the Emotional Symptoms Index ($T=85$), and Internalizing Problems ($T=92$). By contrast, the PRS-A ratings provided by Student D's mother resulted in composite scores in the Typical range (Adaptive Skills, $T=56$; Externalizing Problems, $T=45$; Internalizing Problems, $T=48$; Behavioral Symptoms Index, $T=50$). Student D's mother described him as a loving teen who had difficulty adjusting to changes and stressors.

Measures

Progress Monitoring

Observed behaviors Momentary time sampling using the Internalizing Behavior Observation Protocol (IBOP; Kilgus et al., 2017) measured two target behaviors: *negative affect* (NA; “facial expressions, nonverbal body language, or verbal statements that signal the individual is feeling unhappy, annoyed, or disinterested;” Allen et al., 2019, p. 169) and *social engagement* (SE; “socially appropriate interaction or communication with peers or adults [including] nonverbal and verbal behaviors, both positive or neutral in nature;” Allen et al., 2019, p. 169). In this study, NA was conceptualized as overt (externalized) displays of emotional symptoms; by contrast, SE was conceptualized as behavior incompatible with social withdrawal, one manifestation of internalized emotional symptoms. Each target behavior was scored as present or absent in the final moment of 10-s intervals in 20-min observations.

Data collectors, trained school psychology graduate students who were blind to study phase and intervention timing, observed the focal students two to three times per week in a class period the MTSS team identified as challenging for the student, according to teacher reports of student classroom behavior as well as student grades: Earth/physical science for Student A, Advanced Placement (AP) human geography for Student B, integrated mathematics for Student C, and AP biology for Student D. Data collected using the IBOP in prior research evidenced high interobserver agreement and sensitivity to Tier 2 intervention (Allen et al., 2019). In the present study, interval-by-interval agreement was calculated for 29.50%

of observations (26.47–31.43% of observations across participants) and found to be high (for NA, 100%, $\kappa=1.00$; for SE, 91.30%, $\kappa=0.78$).

Self-reported emotions Student participants provided emotion monitoring ratings as a secondary dependent variable. During POWER Program Group Meeting 2, students selected from a menu one emotion they would like to increase (experience more) and one they would like to decrease (experience less) to reflect their priorities related to their well-being. Subsequently, at the start of weekly group meetings, students evaluated “how much of the week” they “felt” each emotion on a scale of 0 (0%, *never*) to 10 (100%, *always*), a format like that of direct behavior ratings (Kilgus et al., 2019). The primary purpose of this practice was to prompt students to reflect on their emotions as a component of intervention; however, this practice also offered the benefit of incorporating student perceptions of internal states in progress monitoring.

Outcome Evaluation

In addition to describing student participants' pre-intervention functioning, the BASC-3 SRP and PRS-A (Reynolds & Kamphaus, 2015) were used to assess, respectively, students' and parents' perspectives of changes in students' functioning following participation in the POWER Program. Although teacher informants would also offer a unique perspective, teachers' completion of the BASC-3 was not pursued given logistical barriers associated with administration timing: Namely, baseline data collection commenced several weeks after the beginning of a new semester, in which students began new classes with new teachers, and teachers did not perceive knowing their students well enough to complete the BASC prior to the POWER Program's phased rollout. Pre- and post-program implementation, students completed paper forms that were entered and scored into the *Q-global* online system, and parents submitted ratings using the *Q-global* system. Both students and parents were prompted to consider the student's functioning across the “last several weeks” in providing their ratings. Evidence of the BASC-3's reliability and validity is available in the administration manual (Reynolds & Kamphaus, 2015).

Implementation Adherence and Dosage

Implementation adherence was measured using group meeting checklists completed by intervention facilitators. These checklists identified the instructional steps to be led in each group meeting. Facilitators marked whether (*yes*, *no*) they completed each step as described and, in cases they did not, noted the reason(s) or adaptation(s) made. Although direct observations of implementation fidelity and,

moreover, multiple methods of measuring implementation fidelity are considered best practice (e.g., Collier-Meek et al., 2020), in vivo observations and group meeting recordings were determined to be non-viable in this study because of district policy and student/family discomfort. Across POWER groups and group meetings, implementation adherence averaged 93.79% of steps (Student A's group: 97.59%; Student B's group: 96.61%; Student C's group: 99.11%; and Student D's group: 84.78%). The most common reasons for non-adherence included running out of time (e.g., discussion took longer than expected so an activity was moved to the next week; the Weekly Challenge was posted on Schoology), daily schedule deviations (e.g., a fire drill interrupted the period, so a discussion was skipped), and student absences (e.g., activities were modified due to few students in attendance or moved to a subsequent session to ensure all students received the content).

Dosage was measured using group meeting attendance logs maintained by intervention facilitators. Student A was absent for one group meeting (Session 3), the content of which was subsequently reviewed with the student prior to and at the beginning of the next session. Students B, C, and D attended all other scheduled meetings.

Intervention Usability

Following POWER Program implementation, intervention facilitators completed the Usage Rating Profile–Intervention, Revised (URP-IR; Briesch et al., 2013), and student participants completed the Children's Usage Rating Profile (CURP; Briesch & Chafouleas, 2009). The URP-IR is a 29-item questionnaire on which implementation facilitators rate the extent to which they agree with a statement regarding an intervention on a 6-point Likert-type scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). The CURP is a similar 21-item questionnaire that uses a 4-point Likert-type scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). Both scales evidence adequate internal consistency (Briesch & Chafouleas, 2009; Briesch et al., 2013).

Procedures

The first author trained (a) high school interventionists on POWER Program procedures and (b) graduate assistants on data collection procedures. The POWER Program training included an overview of the intervention, including a review of the implementation manual and other program materials; it ended in a meeting to plan the rollout of the intervention and the study procedures. The data collection trainings occurred over four meetings, in which interval recording procedures were taught and practiced; target behavior definitions, examples, and non-examples were reviewed and applied; and setting-specific observation routines were

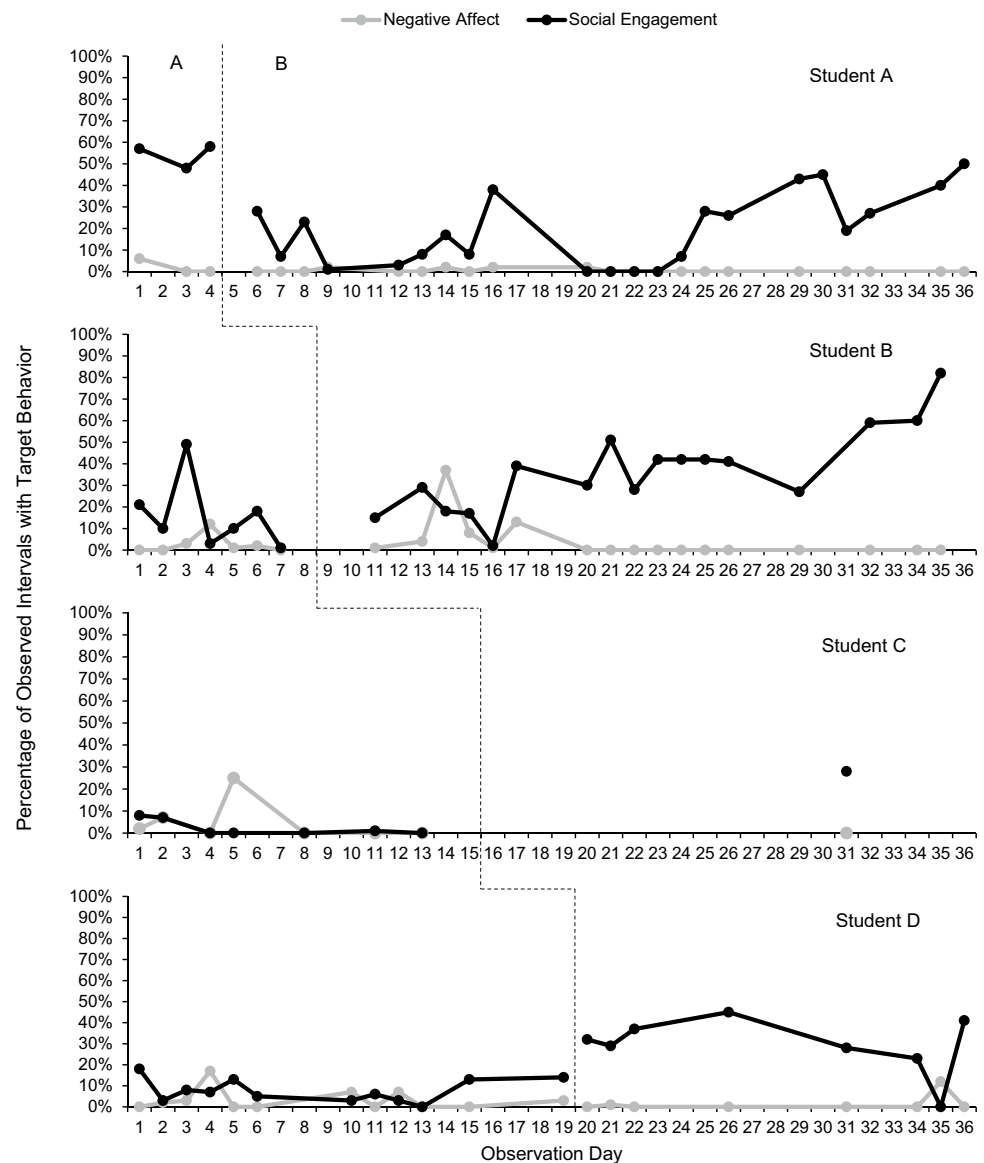
developed and discussed. Data collectors conducted practice observations and together reviewed results until achieving two consecutive observations with at least 90% interval-by-interval agreement on each IBOP target behavior. Practice observations were conducted in target classrooms to provide naturalistic practice and desensitize students to the observers' presence. Baseline data collection began in February 2023, providing student participants three weeks to adjust to their new semester's schedule. Classroom observations were scheduled to occur three times per week for each student participant across the baseline and intervention phases; however, observations were canceled in the following circumstances: (a) the student was absent; (b) the class period was not held (e.g., assembly or field day held in place of class); and (c) limited opportunities for social engagement were planned for the class period (e.g., test or study hall scheduled). Following baseline data collection, facilitators started POWER Program implementation according to phase change decision rules. POWER Program group meetings were held weekly during a study hall or academic period according to (a) facilitators' availability and (b) students' ability to independently make up missed academic work, consistent with expectations for SEB intervention scheduling in this school context.

Research Design and Analyses

A concurrent multiple-baseline (A–B) design (Ledford & Gast, 2018) across four participants was used to examine the effects of the POWER Program on students' NA and SE. The baseline (A phase) consisted of business-as-usual Tier 1 emotional supports within the high school. The POWER Program (B phase) was rolled out in a staggered manner across students. Allowing for at least three non-overlapping data points between students in the intervention phase, the POWER Program was initiated for each student (a) when their baseline data demonstrated stability in NA, primarily, and SE, secondarily, and (b) the prior student's intervention data demonstrated general stability.

POWER Program effects for focal student participants were assessed through visual analysis of graphed data's level, trend, and variability (Ledford & Gast, 2018). Single-case effect sizes were computed using the standardized mean difference (SMD) method (Busk & Serlin, 1992) and percentage of all non-overlapping data (PAND) approach (Parker et al., 2007). Specifically, SMD effect sizes were calculated by subtracting the baseline mean from the intervention mean for each target behavior and dividing that difference by the baseline standard deviation. PAND effect sizes and aligned *phi* coefficients were calculated using the procedure described in Parker and colleagues' (2007) article (see p. 204). PAND values exceeding 0.9 were considered to

Fig. 1 Direct Observation Results for Negative Affect and Social Engagement



signify strong effects, between 0.7 and 0.9 were considered to signify moderate effects, and between 0.5 and 0.7 were considered to signify questionable effects (Scruggs & Mastropieri, 1998). *Phi* values greater than 0.5 were considered

“large,” between 0.3 and 0.5 were considered “moderate,” and between 0.1 and 0.3 were considered “small” (Cohen, 1988). BASC-3 scores, emotion ratings, and usability ratings were descriptively analyzed.

Table 2 Descriptive Statistics and Effect Sizes for Directly Observed Classroom Behavior

Participant	Target behavior	Baseline <i>M</i>	Intervention <i>M</i>	Baseline <i>SD</i>	SMD	PAND	ϕ	Sig
Student A	Negative affect	.02	.00	.03	-.67	83%	.24	.24
	Social engagement	.54	.18	.06	-6.00	75%	-.14	.48
Student B	Negative affect	.03	.04	.04	.25	67%	.19	.34
	Social engagement	.16	.37	.16	1.31	75%	.40	.05
Student D	Negative affect	.03	.02	.05	-.20	70%	.38	.09
	Social engagement	.08	.29	.06	3.50	90%	.79	<.001

Effect sizes for Student C were not calculated given the limited number of intervention phase observations. PAND = percentage of all non-overlapping data. SMD = standardized mean difference

Results

Directly Observed Outcomes

Students' directly observed NA and SE results are graphed in Fig. 1; note the low levels (< 10% of observed intervals) of NA observed in the baseline condition for each student, which suggest the likelihood of floor effects. Descriptive statistics and effect sizes are summarized in Table 2. Results are described below for each student participant.

In the baseline phase, *Student A* had low levels of NA (range, 0–6% of observed intervals) and moderate levels of SE (range, 48–58% of observed intervals). In transitioning to the intervention phase, *Student A*'s NA further decreased and stabilized (range, 0–2% of observed intervals) and SE immediately decreased (initial intervention point: 28% of observed intervals) with fluctuations across the intervention phase (range, 0–50% of observed intervals). Effect size statistics suggest the POWER Program's moderate effectiveness in decreasing *Student A*'s NA ($SMD = -0.67$; $PAND = 83\%$, $\phi = 0.24$) but lack of effectiveness in increasing *Student A*'s SE ($SMD = -6.00$; $PAND = 75\%$, $\phi = -0.14$). Given (a) *Student A*'s individual goal related to increasing “productivity;” (b) the operational definition for SE not accounting for behaviors' on- or off-task nature; (c) variability in engagement expectations and opportunities across observation periods (e.g., for this Earth/physical science class, some class periods required partner/group work throughout their duration, and others included more direct instruction or independent work time); and (d) positive student-reported outcomes (see “Informant-Rated Outcomes”), it is unlikely the observed decreases in SE for *Student A* reflect worsened student outcomes.

Student B demonstrated low levels of NA (range, 0–12% of observed intervals) and moderate but variable levels of SE (range, 1–49% of observed intervals) in the baseline phase. In the intervention phase, *Student B* demonstrated initially variable levels of NA that later decreased and stabilized (range, 0–37% of observed intervals; final 11 intervention data points equal to 0%). *Student B*'s SE data across the intervention phase evidenced a positive trend ($m = 0.02$). Overall, for *Student B*, statistical estimates suggested small effects on NA ($SMD = 0.25$; $PAND = 67\%$, $\phi = 0.19$), primarily realized later in program implementation, and moderate effects on SE ($SMD = 1.31$; $PAND = 75\%$, $\phi = 0.40$).

Except for an outlier data point for NA (25% of observed intervals), *Student C* exhibited stable and low levels of NA (range without outlier, 0–7% of observed intervals) and SE (range, 0–8% of observed intervals) across the baseline phase. Only one data point was available for *Student C* in the intervention phase given the student's frequent

absences from school and limited attendance of the targeted class period. Accordingly, results from the single observation available for the intervention phase cannot be used to draw conclusions about the POWER Program's effectiveness for this student, although this data point reflects sustained low levels of NA (0% of observed intervals; overlapping with three of seven baseline points but lower than the baseline mean of 5%) and a positive increase in SE from baseline (28% of observed intervals; $\Delta = +20\%$ from the most extreme baseline data point).

Student D's NA ($M = 3\%$ of observed intervals, $SD = 5\%$; range, 0–17%) and SE ($M = 8\%$ of observed intervals, $SD = 6\%$; range, 0–18%) were of a similar level across the baseline phase. Upon POWER Program initiation, NA decreased and stabilized at a low level (six of seven intervention data points at 0–1% of observed intervals). By contrast, SE immediately increased (29% of observed intervals; $\Delta = +9\%$ from the most extreme baseline data point) and sustained at a level distinct from that observed in the baseline phase (range without outlier, 23–45% of observed intervals). These patterns held true for all observation days except Day 35, on which NA spiked and SE dipped. For *Student D*, estimates indicated moderate effects on NA ($SMD = -0.20$; $PAND = 70\%$, $\phi = 0.38$) and large effects on SE ($SMD = 3.50$; $PAND = 90\%$, $\phi = 0.79$).

Informant-Rated Outcomes

Pre- and post-intervention BASC-3 SRP and PRS-A T scores (in a normative sample, $M = 50$, $SD = 10$) for composites and Internalizing Problems subscales are reported in Table 3. On the SRP, post-intervention ratings led to decreased T scores for all four students on the Emotional Symptoms Index ($\Delta_M = -9.75$) and for two of the four students on the Internalizing Problems composite ($\Delta_M = -9.00$). SRP subscale analysis suggested improvements were made primarily in the domains of Anxiety ($\Delta = -8$), Sense of Inadequacy ($\Delta = -4$), and Social Stress ($\Delta = -2$) for *Student A*; Somatization ($\Delta = -12$), Sense of Inadequacy ($\Delta = -10$), Anxiety ($\Delta = -10$), and Depression ($\Delta = -3$) for *Student B*; Social Stress ($\Delta = -7$) and Depression ($\Delta = -6$) for *Student C*; and across all Internalizing Problems domains for *Student D* (Atypicality $\Delta = -55$; Somatization $\Delta = -39$; Depression $\Delta = -33$; Sense of Inadequacy $\Delta = -27$; Social Stress $\Delta = -26$; Anxiety $\Delta = -14$; Locus of Control $\Delta = -3$). As previously referenced, in contrast with pre-intervention SRP T scores, most pre-intervention PRS-A T scores were in the Typical range (i.e., < 60), and post-intervention PRS-A T scores reflected some change in composite and subscale scores but primarily in the direction of increased risk (see Table 3).

Student emotion ratings are graphed in Fig. 2. Across POWER Program group meetings, emotion ratings trended

F Index

Ratings	Self-Report of Personality		Parent Rating Scale—Adolescent										
	Pre	Post	Pre	Post									
Student A	Acceptable	Acceptable	Acceptable	Acceptable									
Student B	Acceptable	Acceptable	Acceptable	Acceptable									
Student C	Acceptable	Acceptable	Acceptable	Acceptable									
Student D	Caution	Acceptable	Acceptable	—									
Self-Report of Personality													
Composite Emotional Symptoms Index													
	Pre	Post	Δ										
Student A	71	68	−3	71									
Student B	57	50	−7	51									
Student C	79	75	−4	79									
Student D	85	60	−25	92									
M (SD)	73.00 (12.11)	63.25 (10.75)	−9.75 (10.31)	73.25 (17.17)									
			64.25 (14.66)	−9.00 (16.83)									
Parent Rating Scale—Adolescent													
Composite Behavioral Symptoms Index													
	Pre	Post	Δ										
Student A	59	60	1	58									
Student B	50	50	0	55									
Student C	43	49	6	40									
Student D	50			48									
M (SD)	50.50 (6.56)	53.00 (6.08)	2.33 (3.21)	50.25 (8.02)									
			56.33 (4.16)	5.33 (8.74)									
Self-Report of Personality													
Subscale	Social Stress		Anxiety		Depression		Sense of Inadequacy		Somatization				
	Pre	Post	Δ	Pre	Post	Δ	Pre	Post	Δ	Pre	Post	Δ	
Student A	66	64	−2	77	69	−8	71	62	58	−4	63	66	3
Student B	45	44	−1	59	49	−10	51	58	48	−10	57	45	−12
Student C	73	66	−7	70	71	1	71	80	85	5	81	81	0
Student D	78	52	−26	79	65	−14	49	91	64	−27	84	45	−39
M (SD)	65.50 (14.53)	56.50 (8.99)	−9.00 (11.63)	71.25 (7.82)	63.50 (9.98)	−7.75 (5.49)	60.50 (12.15)	72.75 (13.58)	63.75 (15.48)	−9.00 (13.49)	71.25 (17.56)	59.25 (19.13)	−12.00 (19.13)

Table 3 (continued)

Parent Rating Scale—Adolescent						
Subscale	Anxiety		Depression		Somatization	
	Pre	Post	Δ	Pre	Post	Δ
Student A	64	61	−3	54	59	5
Student B	62	56	−6	49	48	−1
Student C	38	62	24	43	45	2
Student D	45			54		
<i>M (SD)</i>	52.25 (12.76)	59.67 (3.21)	5.00 (16.52)	50.00 (5.23)	50.67 (7.37)	2.00 (3.00)
				48.50 (5.32)	56.33 (2.52)	6.67 (5.69)

BASC-3 Parent Rating Scale—Adolescent post-intervention ratings for Student D were not available

in the desired directions (i.e., based on the emotion goal; increasing productivity and decreasing anxiety) for Student A. For Student B, ratings for the emotion to increase (content) demonstrated a negative trend, and ratings for the emotion to decrease (anxious) demonstrated a positive trend. For Student C, ratings for the emotion to increase (relaxed) were largely stable, and ratings for the emotion to decrease (stressed) trended downward. Finally, for Student D, ratings for both the emotion to increase (calm) and the emotion to decrease (angry) trended downward.

Program Usability

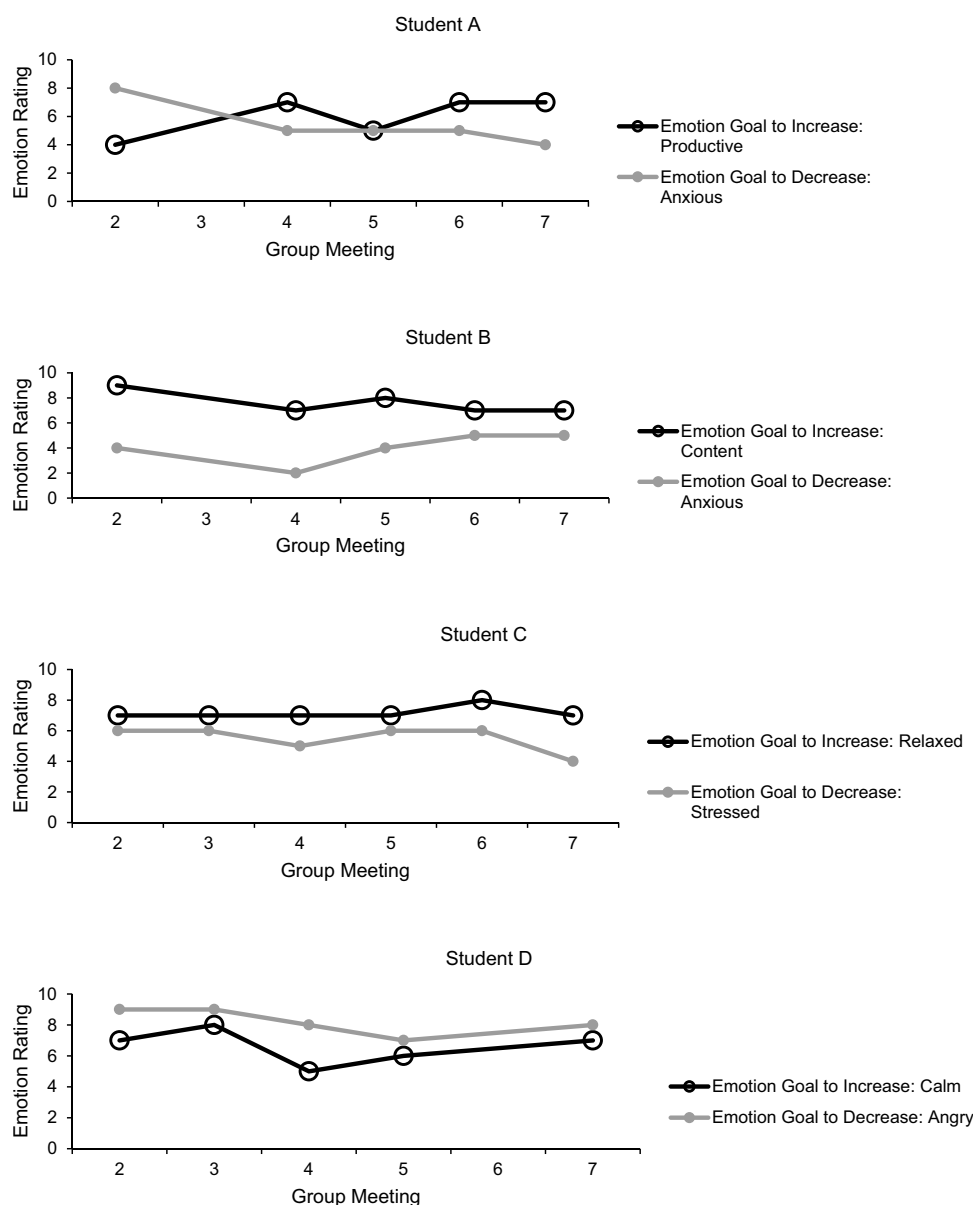
URP-IR and CURP ratings are summarized in Table 4. On the URP-IR, overall, POWER Program facilitators ($n=2$) *strongly agreed* they understood how to implement the program; *agreed* the program was acceptable, feasible, and contextually appropriate; *somewhat disagreed* that strong home–school collaboration was necessary to implement the program; and *disagreed* that they would benefit from additional support in implementing the program. In open-ended comments, facilitators noted strengths of the program as including varied and inclusive meeting activities, frequent opportunities for student sharing and discussion, the scripted nature of the program as increasing access and reducing facilitator preparation time, and specific lessons that worked well with their group. With regard to improvement, facilitators recommended (a) shortening select group meeting plans to better fit within an instructional period and (b) reducing scripting and breaking up facilitator-led instruction to engage students more effectively throughout group meetings. On the CURP, overall, POWER Program focal student participants ($n=4$) *strongly agreed* the program was feasible, *agreed* they understood the program, and *agreed* the program was personally desirable.

Discussion

The POWER Program was designed to be suitable for Tier 2 implementation in schools, feasible for implementation by school personnel without specialized training, appropriate and engaging for high school students, and effective at promoting emotion regulation as a core process underlying emotional disorder risk. The present pilot study evaluated the efficacy and usability of the POWER Program. Results from a sample of four high school students and three facilitators provide initial evidence of (a) positive effects on students' emotional and behavioral symptoms and (b) students' and facilitators' positive impressions of the program's usability.

In this single-case design study, POWER Program participation was associated with decreases in directly observed NA for all three students with sufficient

Fig. 2 Emotion Ratings Across POWER Group Meetings



intervention data (Students A, B, and D). Participation was also associated with increases in directly observed SE for two of these three students (Students B and D). Across students and target behaviors, effects ranged from small to large in magnitude, and approximated those reported in other evaluations of Tier 2 interventions for emotional risk (e.g., for REP, mean PAND for NA was 74% compared to 73.33% in the current study, and mean PAND for SE was 72.33% compared to 80% in the current study; Allen et al., 2019).

In addition, all four students' self-reported data on the BASC-3 reflected reductions in emotional symptoms and/or internalizing problems. Prior Tier 2 research has similarly reported students' self-rated improvements in emotional functioning, for upper elementary students and REP using the BASC-3 (Allen et al., 2019) and Achenbach System

of Empirically Based Assessment (ASEBA) School-Age Scale, Youth Self-Report Form (Achenbach & Rescorla, 2001) as well as for younger elementary students and the Calm Cat Program (Zakszeski et al., 2023) using the Screen for Child Anxiety and Related Disorders (Birmaher et al., 1999). Although parents' post-intervention BASC-3 ratings did not suggest favorable intervention effects, parents' pre-intervention BASC-3 ratings varied significantly from students' and were primarily in the Typical range across composites, which may reflect (a) the distinct information to which each rater (student and parent) has access and (b) increasing rates of emotional symptoms' internalization and somatization during adolescence (e.g., De Los Reyes et al., 2019). Moreover, students' self-ratings of their emotion goals across Group Meetings 2–7 demonstrated mixed

Table 4 POWER Program Usability ResultsUsage Rating Profile—Intervention, Revised (URP-IR)^a

Subscale	<i>M</i>	<i>SD</i>
Acceptability	5.22	.16
Understanding	5.50	.24
Feasibility	4.92	.35
System climate	4.60	.00
Home–school collaboration	2.50	.71
System support	1.67	.47

Children's Usage Rating Profile (CURP)^b

Subscale	<i>M</i>	<i>SD</i>
Personal Desirability	3.00	.32
Feasibility	3.54	.40
Understanding	3.39	.42

Results are presented as subscale item averages (subscale scores divided by the number of subscale items)

^aRating options ranged from 1 (*strongly disagree*) to 6 (*strongly agree*)^bRating options ranged from 1 (*strongly disagree*) to 4 (*strongly agree*)

evidence for improvements across the program; however, baseline ratings were not available to serve as comparisons, and confounds related to the timing of the group meetings (later meetings nearing the end of the school year) may explain some patterns observed.

Altogether, multimethod results provide a comprehensive picture of program effects, offering evidence of efficacy as measured by direct observations, self-reported ratings (Student C), or both methods (Students A, B, D). Yet, effects were less clear for Student C, whose focal class attendance pattern precluded planned classroom observations. In fact, Student C's attendance pattern may be interpreted as providing evidence for the program not meaningfully impacting behavioral functioning, although factors outside the purview of this study further explain and contextualize this pattern.

Both facilitators and students provided generally positive feedback on the POWER Program core group meetings' usability, with ratings in the desired direction on each URP and CURP subscale, indicating that drastic changes to the program's format, foci, and training system are not needed. Indeed, intervention facilitators noted several strengths of the program. Facilitators suggested two improvements, which included further abbreviating two meetings' plans and breaking up facilitator-led direct instruction. These changes may lead to improved feasibility and student engagement within the program. These refinements should be made and pilot-tested prior to future POWER Program research. It must also be noted, however, that the facilitators were school psychologists and school psychologists-in-training, selected to implement the POWER Program given their SEB intervention responsibilities in the participating school. Usability

results for these facilitators may not generalize to facilitators with different training and experiences. With the POWER Program designed to be implemented by a broad range of school-based professionals, additional research is needed to explore usability perceptions among facilitators who are not trained as school psychologists.

Limitations and Future Directions

Limitations of the present study highlight opportunities for future research on the POWER Program. Perhaps most notably, this study's single-case design included only four focal student participants in four POWER groups, three POWER Program facilitators, and one high school. Accordingly, this study focused on evaluating program effects for few students and implementation facilitators in a single setting, prioritizing the evaluation of individual-level effects rather than producing evidence that can generalize to other populations and contexts. Given this study's promising results, future studies should evaluate the POWER Program using larger and more diverse student and implementer samples, continuing to examine usability dimensions such as feasibility and acceptability as well as using experimental designs such as randomized controlled trials (RCTs) and cluster RCTs. Additionally, although school personnel initially identified both NA and SE as concerns for the student participants in their identified class periods, NA was consistently observed in a small percentage (i.e., <10%) of intervals across students during the baseline phase, such that floor effects likely influenced study findings. An alternative approach to progress monitoring could use classroom

behavioral targets aligned with students' selected POWER Program goals, idiosyncratically defined and confirmed to be present in a considerable percentage of intervals during baseline observations.

Future studies might also evaluate other types of outcomes associated with POWER Program participation, such as outcomes commonly collected in schools' information systems (e.g., attendance, grades), teacher-rated behavioral and emotional symptoms, and even peer-rated behavioral and emotional symptoms (considering the close connections between students and their friends during the adolescent developmental stage). Multimethod, multisource assessment of intervention effects—with regard to both progress monitoring and outcome evaluation—is likely to provide the most nuanced characterization of students' SEB functioning, as exemplified in the present study's variable patterns of results. Practitioners may feasibly embed this type of evaluation procedure in their Tier 2 systems by drawing upon data already collected in their schools, administering brief measures of intervention targets during time allocated for intervention activities (e.g., students' emotion goal monitoring during POWER group meetings and Check-Outs), and collecting data from other informants on fewer (e.g., pre/post) occasions.

Another category of limitations relates to the restricted ways in which the present study can inform the application of the POWER Program in schools. As a pilot study, this investigation concerned *whether* the program works, not *how, for whom, and under what conditions* it works. For example, this study evaluated only the core group meeting sequence of the POWER Program, not evaluating program components such as booster group meetings, individual meetings, and the Check-Out process. Future research should test the feasibility of these intervention components as well as the incremental benefits of including these components, directly testing, for example, the initial effects, generalizability, and maintenance of program outcomes for students who did and did not participate in POWER Check-Outs. Moreover, although focal student participants with somewhat varied strengths and concerns were included in this preliminary study, a direct evaluation of differential intervention effectiveness was not conducted, such that it remains unclear what student profiles are best served by this program. It also remains unclear what levels of implementation fidelity dimensions (e.g., adherence, dosage) are necessary for positive student outcomes, and future research might include and compare multiple methods of assessing program fidelity (see Collier-Meek et al., 2020). It would be beneficial for future studies to assess (a) what organizational conditions (e.g., building and district policies and infrastructure) and (b) what training and implementation support conditions are

most closely associated with high-quality implementation and strong effects on student outcomes.

Conclusions

This pilot study provides evidence of the POWER Program's efficacy and usability as a Tier 2 intervention with secondary students at risk of developing emotional disorders. With results limited to a small participant sample and single implementation context, future research is needed to evaluate effectiveness and appropriateness in larger, more diverse samples and varied contexts. In addition, research is needed to specify optimal implementation parameters to guide potential applications in schools. Continued refinement and evaluation of the POWER Program may result in an intervention that can be adopted by secondary settings to support students' well-being and school success.

Funding This research was funded through [grant information redacted for review].

Declarations

Conflict of interest The authors declare no conflicts of interest.

Research Involving Human Participants All study procedures were approved by the authors' institutional review board and conducted in accordance with its ethical standards.

Informed Consent All human participants provided informed consent and assent to participate in study procedures.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, and Families.
- Allen, A. N., Kilgus, S. P., & Eklund, K. (2019). An initial investigation of the efficacy of the resilience education program (REP). *School Mental Health, 11*, 163–178. <https://doi.org/10.1007/s12310-018-9276-1>

- Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the screen for child anxiety related emotional disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 1230–1236. <https://doi.org/10.1097/00004583-199910000-00011>
- Briesch, A. M., & Chafouleas, S. M. (2009). Exploring student buy-in: Initial development of an instrument to measure likelihood of children's intervention usage. *Journal of Educational and Psychological Consultation*. <https://doi.org/10.1080/10474410903408885>
- 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, 81–96. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Lawrence Erlbaum.
- Centers for Disease Control and Prevention. (2023). *Youth Risk Behavior Survey: Data summary and trends report, 2011–2021*. https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBS_Data-Summary-Trends_Report2023_508.pdf
- Cilar, L., Štiglic, G., Kmetec, S., Barr, O., & Pajnkihar, M. (2020). Effectiveness of school-based mental well-being interventions among adolescents: A systematic review. *Journal of Advanced Nursing*, 76(8), 2023–2045. <https://doi.org/10.1111/jan.14408>
- Clifford, M. E., Nguyen, A. J., & Bradshaw, C. P. (2020). Both/and: Tier 2 interventions with transdiagnostic utility in addressing emotional and behavioral disorders in youth. *Journal of Applied School Psychology*, 36, 173–197. <https://doi.org/10.1080/15377903.2020.1714859>
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Colizzi, M., Lasalvia, A., & Ruggieri, M. (2020). Prevention and early intervention in youth mental health: Is it time for a multidisciplinary and transdiagnostic model for care? *International Journal of Mental Health Systems*, 14(23). <https://doi.org/10.1186/s13033-020-00356-9>
- Collier-Meek, M. A., Sanetti, L. M., Fallon, L., & Chafouleas, S. (2020). Exploring the influences of assessment method, intervention steps, intervention sessions, and observation timing on treatment fidelity estimates. *Assessment for Effective Intervention*, 46, 3–13. <https://doi.org/10.1177/1534508419857228>
- Cook, C. R., Xie, S. R., Earl, R. K., Lyon, A. R., Dart, E., & Zhang, Y. (2015). Evaluation of the courage and confidence mentor program as a Tier 2 intervention for middle school students with internalizing problems. *School Mental Health*, 7, 132–146. <https://doi.org/10.1007/s12310-014-9137-5>
- Creed, T. A., Reisweber, J., & Beck, A. T. (2011). *Cognitive therapy for adolescents in school settings*. Guilford Press.
- Crone, D. A., Horner, R. H., & Hawken, L. S. (2004). *Responding to problem behavior in schools: The Behavior Education Program*. Guilford.
- De Los Reyes, A., Cook, C. R., Gresham, F. M., Makol, B. A., & Wang, M. (2019). Informant discrepancies in assessments of psychosocial functioning in school-based services and research: Review and directions for future research. *Journal of School Psychology*, 74, 74–89. <https://doi.org/10.1016/j.jsp.2019.05.005>
- Dib, N., & Sturmey, P. (2012). Behavioral skills training and skill learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 437–438). Springer.
- Eber, L., Weist, M., & Barrett, S. (2017). An introduction to the interconnected systems framework. In S. Barrett, L. Eber, & M. Weist (Eds.), *Advancing education effectiveness: Interconnecting school mental health and school-wide positive behavior support* [Monograph; pp. 3–17]. <https://www.pbis.org/resource/advancing-education-effectiveness-interconnecting-school-mental-health-and-school-wide-positive-behavior-support>
- GL Education. (2023). *PASS (Pupil Attitudes Toward Self and School)* [Assessment instrument]
- Ehrenreich-May, J., & Chu, B. C. (2014). *Transdiagnostic treatments for children and adolescents: Principles and practice*. Guilford.
- Eiraldi, R., Wolk, C. B., Locke, J., & Beidas, R. (2015). Clearing hurdles: The challenges of implementation of mental health evidence-based practices in under-resourced schools. *Advances in School Mental Health Promotion*, 8, 124–140. <https://doi.org/10.1080/1754730X.2015.1037848>
- Eklund, K., Kilgus, S. P., Izumi, J., DeMarchena, S., & McCollom, E. P. (2021). The Resilience Education Program: Examining the efficacy of a tier 2 internalizing intervention. *Psychology in the Schools*, 58, 2114–2129. <https://doi.org/10.1002/pits.22580>
- Fiat, A. E., Cook, C. R., Zhang, Y., Renshaw, T. L., DeCano, P., & Merrick, J. S. (2017). Mentoring to promote course and confidence among elementary school students with internalizing problems: A single-case design pilot study. *Journal of Applied School Psychology*, 33, 261–287. <https://doi.org/10.1080/15377903.2017.1292975>
- Fusar-Poli, P. (2019). Integrated mental health services for the developmental period (0 to 25 years): A critical review of the evidence. *Frontiers in Psychiatry*, 10(355). <https://doi.org/10.3389/fpsy.2019.00355>
- Greenspoon, P. J., & Saklofske, D. H. (2001). Toward an integration of subjective well-being and psychopathology. *Social Indicators Research*, 54, 81–108. <https://doi.org/10.1023/A:1007219227883>
- Kazdin, A. E. (1990). Psychotherapy for children and adolescents. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 41, pp. 21–54). Annual Reviews.
- Kern, L., Mathur, S. R., Albrecht, S. F., Poland, S., Rozalski, M., & Skiba, R. J. (2017). The need for school-based mental health services and recommendations for implementation. *School Mental Health*, 9, 205–217. <https://doi.org/10.1007/s12310-017-9216-5>
- Kilgus, S. P., Dart, E., von der Embse, N. P., & Collins, T. (2017). *Development and validation of the Internalizing Behavior Observation Protocol (IBOP)*. Unpublished measure.
- Kilgus, S. P., Van Wie, M. P., Sinclair, J. S., Riley-Tillman, T. C., & Herman, K. C. (2019). Developing a direct behavior rating scale for depression in middle school students. *School Psychology*, 34, 86–95.
- Kilpatrick, K., Kilgus, S. P., Eklund, K., & Herman, K. C. (2021). An evaluation of the potential efficacy and feasibility of the Resilience Education Program: A Tier 2 internalizing intervention. *School Mental Health*, 13, 376–391. <https://doi.org/10.1007/s12310-021-09428-8>
- Ledford, J. R., & Gast, D. L. (2018). *SinOgle case research methodology: Applications in special education and behavioral sciences* (3rd ed.). Routledge.
- Martinez, S., Kern, L., Flannery, B., White, A., Freeman, J., & George, H. P. (2019). *High school PBIS implementation: Staff buy-in*. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org
- Martinsen, K. D., Rasmussen, L. M. P., Wentzel-Larsen, T., Holen, S., Sund, A. M., Løvaas, M. E. S., Patras, J., Kendall, P. C., Waaktaar, T., & Neumer, S.-P. (2019). Prevention of anxiety and depression in school children: Effectiveness of the transdiagnostic EMOTION program. *Journal of Consulting and Clinical Psychology*, 37, 212–219. <https://doi.org/10.1037/ccp0000360>
- McIntosh, K., Ty, S. V., & Miller, L. D. (2014). Effects of school-wide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions*, 16, 209–218. <https://doi.org/10.1177/1098300713491980>

- McMahon, S. D., Grant, K. E., Compas, B. E., Thurm, A. E., & Ey, S. (2003). Stress and psychopathology in children and adolescents: Is there evidence of specificity? *Journal of Child Psychology and Psychiatry*, 44, 107–133. <https://doi.org/10.1111/1469-7610.00105>
- Mitchell, B. S., Stormont, M., & Gage, N. (2011). Tier two interventions implemented within the context of a tiered prevention framework. *Behavioral Disorders*, 36, 241–261. <https://doi.org/10.1177/019874291103600404>
- Moore, S. A., Dowdy, E., Nylund-Gibson, K., & Furlong, M. (2019). A latent transition analysis of the longitudinal stability of dual-factor mental health in adolescence. *Journal of School Psychology*, 73, 56–73. <https://doi.org/10.1016/j.jsp.2019.03.003>
- National Association of School Psychologists. (2021). *Shortages in school psychology: Challenges to meeting the growing needs of U.S. students and schools* [Research summary].
- National Center for Education Statistics. (2022). Public schools and limitations in schools' efforts to provide mental health services. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. <https://nces.ed.gov/programs/coe/indicator/a23>.
- O'Reilly, M., Sviryzdenka, N., Adams, S., & Dogra, N. (2018). Review of mental health promotion interventions in schools. *Social Psychiatry & Psychiatric Epidemiology*, 53(7), 647–662. <https://doi.org/10.1007/s00127-018-1530-1>
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND) an alternative to PND. *Journal of Special Education*, 40, 194–204. <https://doi.org/10.1177/00224669070400040101>
- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children (3rd ed.)*. Pearson.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single subject research: Issues and applications. *Behavior Modification*, 22, 221–242. <https://doi.org/10.1177/01454455980223001>
- Sherman, J. A., & Ehrenreich-May, J. (2020). Changes in risk factors during the unified protocol transdiagnostic treatment of emotional disorders in adolescents. *Behavior Therapy*, 51, 869–881. <https://doi.org/10.1016/j.beth.2019.12.002>
- Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., & Staiger, P. K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, 57, 141–163. <https://doi.org/10.1016/j.cpr.2017.09.002>
- Todd, A., Campbell, A., Meyer, G., & Horner, R. (2008). Evaluation of a targeted group intervention in elementary students: The check-in check-out program. *Journal of Positive Behavior Interventions*, 10, 46–55. <https://doi.org/10.1177/1098300707311369>
- Tylee, A., Haller, D. M., Graham, T., Churchill, R., & Sanci, L. A. (2007). Youth-friendly primary-care services: How are we doing, and what more needs to be done? *Lancet*, 369(9572), 1565–1573. [https://doi.org/10.1016/S0140-6736\(07\)60371-7](https://doi.org/10.1016/S0140-6736(07)60371-7)
- United States Department of Health and Human Services. (2020). *Designated health professional shortage areas: Statistics (December 31, 2019)* [Research summary].
- United States Department of Education. (2022). *The U.S. Department of Education announces partnerships across states, school districts, and colleges of education to meet Secretary Cardona's call to action to address the teacher shortage* [Fact sheet]. <https://www.ed.gov/coronavirus/factsheets/teacher-shortage>.
- Weiss, B., Harris, V., Catron, T., & Han, S. S. (2003). Efficacy of the RECAP intervention program for children with concurrent internalizing and externalizing problems. *Journal of Consulting and Clinical Psychology*, 71, 364–374. <https://doi.org/10.1037/0022-006X.71.2.364>
- Weisz, J., Bearman, S. K., Santucci, L. C., & Jensen-Doss, A. (2017). Initial test of a principle-guided approach to transdiagnostic psychotherapy with children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 46, 44–58. <https://doi.org/10.1080/15374416.2016.1163708>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). <https://ies.ed.gov/ncee/wwc/Handbooks>.
- World Health Organization. (2012). *Making health services adolescent friendly: Developing national quality standards for adolescent-friendly health services*. Author. https://apps.who.int/iris/bitstream/handle/10665/75217/9789241503594_eng.pdf.
- Zakszeski, B. N., Banks, E., & Parks, T. (2023). Targeted intervention for elementary students with internalizing behaviors: A pilot evaluation. *School Psychology Review*. <https://doi.org/10.1080/2372966X.2023.2195806>
- Zakszeski, B. N., Rutherford, L., Francisco, J., & Sanders, J. (2022). Positive Outcomes with Emotion Regulation: The POWER Program [Intervention manual]. Mental Health Technology Transfer Center Network and Devereux Center for Effective Schools.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.