# An Evaluation of the Potential Efficacy and Feasibility of the Resilience Education Program: A Tier 2 Internalizing Intervention

Kayla D. Kilpatrick, Ph.D.
kilpatrickk@missouri.edu
University of Missouri

Stephen P. Kilgus, Ph.D.
skilgus@wisc.edu
University of Wisconsin-Madison

Katie Eklund, Ph.D.
katie.eklund@wisc.edu
University of Wisconsin-Madison

Keith C. Herman, Ph.D.
hermanke@missouri.edu
University of Missouri

*School Mental Health*



Note: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A190129 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence regarding this article should be addressed to Stephen Kilgus at the University of Wisconsin-Madison, Department of Educational Psychology, Madison, WI 53706; email: skilgus@wisc.edu

Abstract

This study examined the feasibility and potential efficacy of the Resilience Education Program (REP), a Tier 2 school-based internalizing intervention. REP represents a hybrid intervention approach, incorporating both small-group cognitive-behavioral instruction and a Check-In/Check-Out reinforcement-based mentorship program. A randomized controlled trial research design was employed, in which students (grades 4-7) were randomly assigned to treatment ($n$ = 21) or waitlist control ($n$ = 17) groups. Given the early phase of REP research, the trial was underpowered but capable of generating unbiased effect size estimates that could inform subsequent fully powered efficacy trials. Outcomes of interest included student internalizing concerns (as reported by teacher and self-report measures) and the change mechanisms by which REP was theorized to influence internalizing concerns (i.e., emotional control and social support). Primary MANCOVA findings indicated that although non-statistically significant, between-group differences in youth self-reported and teacher-reported internalizing concerns corresponded to large effect sizes ($\eta_p^2$= .15–.19). A follow-up MANCOVA inclusive of change mechanism variables was also non-statistically significant, but representative of a large effect ($\eta_p^2$= .38). Following trial completion, REP implementers positively rated the acceptability of the REP intervention as a Tier 2 intervention for addressing internalizing concerns in the school setting.

# An Evaluation of the Potential Efficacy and Feasibility of the Resilience Education Program: A Tier 2 Internalizing Intervention

An estimated 10-20% of youth experience a mental health disorder (World Health Organization, 2017). Approximately half of these individuals exhibit internalizing problems (Merikangas et al., 2010), defined as emotional or mood problems that may include social withdrawal, somatic and physical problems, rumination, difficulties regulating emotions, and avoidance of specific situations or stimuli (Levitt & Merrell, 2009). For those children who experience internalizing problems early in their school career, these challenges increase over time and are more likely to manifest as a full disorder. In fact, approximately 20% of adolescents who are affected by internalizing problems early in life will experience a depressive episode before reaching age 18 and nearly 32% will be affected by symptoms of an anxiety disorder (Merikangas et al., 2010). Youth internalizing problems present along a continuum, ranging from (a) *subthreshold symptoms* that do not meet criteria for a mental health disorder, but can nevertheless prove chronic and predictive of later more significant mental health concerns; to (b) *clinically significant symptoms* indicative of the presence of a mental health disorder (e.g., generalized anxiety disorder; Fergusson et al., 2005). Research suggests these varying degrees of internalizing problems have negative impacts on youth social and academic functioning across multiple settings (Mychailyszyn et al., 2010).

Research has established evidence-based interventions for internalizing problems, with cognitive-behavioral therapy being one of the leading treatments for youth ages 8 and up (Weisz, 2015). However, much of the research on internalizing interventions has taken place in outpatient clinical mental health settings, examining intensive interventions with youth exhibiting clinically significant internalizing symptoms (Sanchez et al., 2018). This is

unfortunate, as individuals exhibiting subthreshold internalizing problems are also at risk for long-term negative outcomes (Fergusson et al., 2005). Furthermore, many youths may have difficulty accessing clinical settings, either as a result of their limited availability or due to a number of access barriers (e.g., inflexible parent work schedules; Yoshikawa et al., 2012).

## School-based Internalizing Interventions

Given this state of affairs, scholars have called for increased research that examines interventions that are suitable for implementation within school settings (Weist et al., 2018). The focus on schools is founded in an emphasis on access, as education is compulsory in the United States and youths spend the majority of their time at school. These calls have also recommended that research increasingly examine interventions specific to subthreshold internalizing symptoms (Kilgus et al., 2015). In accordance with a multi-tiered system of support (MTSS) approach, such interventions would be applicable to "Tier 2," which is defined by brief and targeted intervention for youth exhibiting risk for mental health concerns.

**Check in/check out.** Research emerging in response to these calls has examined a range of Tier 2 interventions. Several of these studies have examined whether Check In/Check Out (CICO), a Tier 2 intervention commonly used in schools with youth exhibiting externalizing problems, can be adapted to address internalizing problems. Typical CICO implementation involves each target youth meeting with a mentor in the morning who assesses the youth's readiness for school and reminds the youth of their individualized goals for that day. Throughout the day, teachers provide the youth feedback relative to school-wide behavioral expectations (e.g., respect, responsibility, and safety) via ratings on a Daily Progress Report (DPR). At the end of the day, the mentor meets with the youth to review the DPR and determine if they met

their daily goal and thus earned a reward. The youth then takes the DPR home for their parents to review and provide a home-based reward if applicable.

In adapting CICO to address internalizing concerns, researchers have employed (1) alternative behavioral expectations, corresponding to positive replacement behaviors that are incompatible with each youth's internalizing symptoms (e.g., *making eye contact when in conversation*); and (2) alternative check in/out procedures, wherein mentors use cognitive-behavioral logic to problem-solve challenging situations youths are experiencing. Multiple single-case experimental design (SCED) studies have examined the efficacy of these adapted CICO interventions with elementary school students (e.g., Dart et al., 2015). Visual and statistical analyses suggested CICO was associated with increases in academic engagement and reductions in internalizing symptoms per multiple outcomes.

**Integrated interventions**. Though findings from these studies were promising, potential limitations to these CICO adaptations have been noted (Allen et al., 2019). Specifically, though CICO might be capable of promoting a youth's use of cognitive-behavioral coping skills they already possess, the intervention cannot necessarily promote a youth's learning of such skills if they have yet to acquire them. That is, CICO lacks an active instructional component. This is unfortunate, as research suggests coping skills (e.g., cognitive restructuring, behavioral activation) represent a primary means by which individuals can combat and overcome internalizing problems (Webb et al., 2019). Scholars have thus begun to examine integrated interventions that combined CICO with abbreviated cognitive-behavioral instruction (CBI) to support both skill learning and use. This approach has been described as aligning with social-ecological theory, as caretakers (e.g., educators and parents) are structuring youth environments to not only instruct key skills, but also prompt and reinforce the skills (Allen et al., 2019).

**Resilience Education Program**

An example of a social-ecological approach to integrated internalizing intervention is found in the *Resilience Education Program* (REP; Kilgus & Eklund, 2017). REP combines abbreviated CBI with CICO to support youth learning and use of key coping skills. In accordance with a transdiagnostic treatment approach (Barlow et al., 2011), REP CBI consists of five lessons that are intended to promote youth acquisition of cognitive-behavioral skills applicable to depression or anxiety (e.g., cognitive restructuring and positive imagery). REP CICO procedures are mostly consistent with standard implementation, including daily check-ins, teacher feedback throughout the day, and check-outs with mentors in which the youth receives praise and a reward if they met their daily goal. A small number of adaptations have also been included to align REP CICO with internalizing problems. First, rather than rating school-wide behavioral expectations, teachers rate the extent to which youths have *controlled their emotions* and *made good choices*, two likely outcomes of youths effectively using skills instructed through CBI. A student's educators and parents can also select to rate one additional behavior relevant to each particular student's needs. Second, using the REP DPR, youths rate their emotional state throughout the day, each time the teacher rates their behavior. These ratings are meant to call a youth's attention to their emotions and their potential variance throughout the day, while also demonstrating the correspondence between one's emotions and behavior. Third, the REP CICO facilitator with whom the student checks in and out each day is also versed in the CBI lessons. Accordingly, this individual is able to use the CBI logic in problem-solving with students around challenging situations encountered throughout the school day.

**Theory of change**. Figure 1 depicts the theory of change upon which REP is founded. Collectively, REP CBI and CICO are presumed to influence two change mechanisms: (1)

emotional control, as a result of cognitive-behavioral skill use, and (2) social support for youth use of these skills and other positive behaviors. It is expected these change mechanisms then mediate reductions in internalizing problem behaviors. The goal for this reduction is to remove barriers to youth participation in academic and social activities (Fanti & Henrich, 2010), resulting in increased academic and social engagement. Reductions in internalizing problems and improvements in engagement should then reduce the likelihood of youths going on to receive mental health diagnoses or referrals for more intensive treatment.

**Initial evidence**. Allen et al. (2019) conducted an initial pilot SCED trial of REP efficacy within an urban elementary school setting. Participants included three students in 4-5$^{th}$ grade who demonstrated subthreshold internalizing symptoms. REP was evaluated via a multiple baseline design, wherein each student proceeded through baseline and intervention phases. Systematic direct observation was used as a primary outcome measure of three behaviors. *Social engagement* and *negative affect* were evaluated for all students. An additional internalizing problem behavior specific to each student's internalizing concerns was also observed (e.g., worrying or withdrawal). Teachers also completed the *Behavioral Assessment Scale for Children, Third Edition – Teacher Rating Scale* (BASC-3) at both pre and posttest. Visual analyses of systematic direct observation data suggested REP was effective for two students and potentially effective for the third. This conclusion was further support by effect size statistics, which revealed moderate-to-large effects for these two students. Improvements in BASC-3 Internalizing Problems scale scores were also noted for two participants, with scores reducing by approximately one standard deviation. Educator ratings on the *Usage Rating Profile-Intervention, Revised* (UPR-IR; Briesch et al., 2013) supported REP social validity, with findings indicating REP implementers found the intervention acceptable, understandable, and feasible.

**Purpose of the Study**

With these initial promising findings, there is now a need to evaluate REP with larger samples to derive more generalizable estimates of the intervention's efficacy. Additionally, there is a need to examine a broader range of the components present within the REP theory of change. Such research would inform an improved understanding of (1) the change mechanisms through which REP operates and (2) the outcomes it can be expected to influence. The purpose of this study was to address these research needs through a randomized controlled trial (RCT), wherein students in grades 4-7 were assigned to two groups: REP and waitlist control. Given the early phase of REP research, the study was underpowered. Though such an investigation is less likely to yield statistically significant findings, it can nevertheless yield unbiased effect size estimates that can be used to inform subsequent larger-scale trials. The research questions (RQs) examined through this study were as follows:

1. What is the effect of REP on students' internalizing concerns as reported by student self-report and teacher report, compared to a waitlist control group? Based on prior SCED findings, it was hypothesized that the REP intervention would decrease internalizing concerns for participants in the intervention group, whereas those in the control group would demonstrate stable or increasing concerns.
2. What is the effect of REP on change mechanisms specified in the theory of change, including (a) emotional control and (b) social support? It was expected REP would be associated with increases in these variables for participants in the intervention group whereas those in the control group would remain stable.

3. To what extent do educators find REP a socially valid approach to Tier 2 intervention? Based upon results from Allen et al. (2019), it was expected educators would rate REP as highly acceptable, usable, and understandable.

## Method

### Power Analysis

An *a priori* power analysis was conducted using the G*Power software. Given the underpowered nature of this pilot study, we chose to solve for a power level (1- β) of .60, rather than the traditional .80 level. The power analysis was specific to a repeated measures MANOVA, with two between-factor levels (REP and waitlist control) and two within-factor levels (i.e., pre and posttest). Also assumed was an expected moderate effect size of $f(V) = .35$. It was determined that a sample of 42 students was needed to achieve the desired statistical power. A post hoc power analysis was also conducted to determine what effect size could be detected as statistically significant given our particular sample size with power equal to the traditional .80 threshold. Results suggested the current RCT would need to yield a large effect size of $f(V) = .47$ for omnibus hypothesis tests to achieve statistical significance.

### Participants

Participants were selected from two school districts in the Midwest, one of which was rural (Site 1) while the other was urban (Site 2). The total number of student participants ($n = 39$) closely approximated the target sample size per the power analysis. Constrained randomization was used to ensure an equal number of students were assigned to the intervention and waitlist control groups across schools. Student participants ranged in age from 9 to 13 years and were recruited from grades 4-7. Inclusionary criteria for students to participate in the intervention included (1) participation in the general education setting and (2) demonstration of emotional or

internalizing risk per the universal screening tool the school used as part of their normal educational practice. Exclusionary criteria for student participants included receiving special education supports, current mental health diagnoses, and current receipt of other school- or community-based mental health behavioral supports as reported by parents or school personnel.

Participants were enrolled in the study between December 2017 and March 2018, and data collection took place between January 2018 and May 2018. At Site 1, 35 middle school students met criteria and were recruited, with 13 students receiving parental consent. In addition, 24 elementary students were recruited at Site 1, with 13 students receiving parental consent. One Site 1 participant withdrew from the study prior to beginning the intervention due to their parent revoking their initial consent. At Site 2, 30 elementary students were recruited, with 14 students receiving parental consent. The total sample obtained after removing one participant, who became home-schooled during the study, was 38 participants, with 21 assigned to the intervention group and 17 assigned to the waitlist control group. Student demographic information was collected for these participants, including information regarding age, grade, race/ethnicity, gender, and socio-economic status. See Table 1 for a summary of student demographic information.

In addition, educators in the schools were recruited to serve as CICO mentors. Educators interested in serving as CICO mentors for the study could be teachers, administrators, school psychologists, or school counselors. The goal was to have one CICO mentor for every three students in the intervention; therefore, eight CICO mentors were recruited. Teachers of participating students were asked to support CICO delivery by providing performance feedback to students throughout each school day.

**Measures**

**Universal screeners**. Both research sites used universal screening tools to identify students at-risk for internalizing concerns. These measures were already in place as universal school practices at each site prior to the current study. Site 1 used the *Social, Academic, and Emotional Behavior Risk Screener* (SAEBRS) teacher rating scale and the *Strengths and Difficulties Questionnaire* (SDQ) student self-report rating scale. Site 2 used the *Early Identification System* (EIS) teacher report and self-report rating scales. Research has supported the psychometric defensibility of these screeners (e.g., Huang et al., 2019; Kilgus et al., 2018).

**Teacher interview**. A brief teacher interview was conducted to identify internalizing symptoms and target replacement behaviors for each student. Teachers were first asked to provide a description of student strengths and challenges. Then, 15 specific internalizing problems were presented to the teacher (e.g., negative affect, crying, irritability); the teacher was to identify the specific types of internalizing symptoms the student exhibited. Each student's teacher (or teachers in the case of middle school students) selected the top internalizing problem behavior for each student. A desired replacement behavior was then identified for this behavior. Each student was provided feedback on their desired behavioral replacement behavior through their DPR form on a daily basis as part of the REP CICO process.

**Internalizing concerns**. Multiple forms from the *Achenbach System of Empirically Based Assessment* (ASEBA) School-Age Scales were used to assess student internalizing concerns. The ASEBA scales consist of three separate forms that measure student behavior across a variety of subscales (Achenbach & Rescorla, 2001). For the purposes of this study, each student's teacher completed the *Teacher's Report Form* for Ages 6-18 (TRF) and each student completed the *Youth Self-Report* for Ages 11-18 (YSR). All students were given the options of having self-report items read aloud to them. Studies have shown that the ASEBA School-Age

Scale scores demonstrate internal consistency, test-retest, and inter-rater reliability (Achenbach & Rescorla, 2001). ASEBA School-Age Scales have also demonstrated adequate criterion-related and construct validity (Achenbach & Rescorla, 2001). Five ASEBA scales were of interest in this study given their relation to student internalizing concerns. These included the Internalizing Problems scale and four subscales subsumed under this scale: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, and Social Problems. All scales yielded *T* scores (*M* = 50, *SD* = 10) derived in consideration of normative ASEBA data. Across the TRF and YSR measures and these five scales, alpha coefficients ranged from .65–.92 at pretest and .76–.94 at posttest. To note, all ASEBA reliability coefficients were acceptable (>.70) except two YSR subscales at pretest (Withdrawn/Depressed $\alpha = .65$ and Social Problems $\alpha = .69$). Of note, a limitation of the current study is that several participants were younger than age 11 and outside the scope of the YSR validation age range; student grade level was controlled for as a covariate in the analyses to address this limitation.

**Emotional control.** Two measures were used as indicators of emotional control. The first was the *Self-Report Coping Scale* (SRCS; Causey & Dubow, 1992), a 34-item rating scale. Students were provided with two vignettes describing challenging situations (i.e., disagreement with a friend and getting a bad grade) and used a 5-point Likert scale to indicate how often they typically engage in both Approach and Avoidance coping skills (1 = *None of the time* to 5 = *All of the time*). The Approach subscales include Seeking Social Support and Problem-Solving, whereas the Avoidance subscales include Distancing, Internalizing, and Externalizing. The overall Approach and Avoidance scales along with each subscale were used as outcome measures in this study. The SRCS has been shown to demonstrate reliability and validity with students in elementary and middle school grades (Causey & Dubow, 1992). Within this study,

alpha coefficients ranged from .51–.88 for SRCS subscales at pretest and .56–.94 at posttest, respetively. To note, all SRCS reliabilities were acceptable (>.70) except the Approach: Self-reliance/Problem-Solving pretest and Avoidance: Externalizing subscales pre and posttest (Approach: Self-reliance/Problem-Solving pretest $\alpha = .64$, Avoidance: Externalizing prestest $\alpha = .51$, and Avoidance: Externalizing posttest $\alpha = .56$).

The second indicator of emotional control was the *Perceived Control of Internal States Scale* (PCOISS), an 18-item rating scale (Pallant, 2000). Using a 5-point Likert scale (1 = *Strongly agree* to 5 = *Strongly disagree*), individuals rate items such as "if my stress levels get too high, I know there are things I can do to help myself." The PCOISS has been shown to yield reliable and valid scores (Pallant, 2000). The PCOISS total scale yielded an alpha coefficient equal to .72 and .84 at pre and post-test, respectively.

**Social support**. Student perceptions of their social support was assessed using the *Child and Adolescent Social Support Scale* (CASSS; Malecki et al., 2000), a 60-item self-report rating scale. Using a 6-point Likert scale (1 = *Never* to 6 = *Always*), students rated the frequency with which they received various support from multiple sources, including Parents, Teachers, Classmates, Close Friends, and School. In addition, students rated the importance of each support using a 3-point Likert scale (1 = *Not important* to 3 = *Very important*). CASSS scales of interest in this study were Teacher Support, Importance of Teacher Support, School Support, and Importance of School Support. Past research has supported the reliability and validity of CASSS scores (Malecki et al., 2000). Within this study, CASSS alpha coefficients ranged from .94–.97 at pretest and .89–.97 at posttest.

**Social validity**. CICO mentors and participating teachers completed the *Usage Rating Profile-Intervention, Revised* (URP-IR) to provide their perceptions of REP social validity. The

URP-IR includes 29 items that users rate on a 6-point Likert scale (1 = *Strongly Disagree* to 6 = *Strongly Agree*). A total of 6 subscales are produced, including Acceptability, Understanding, Feasibility, Family-School Collaboration, System Climate, and System Support. The URP-IR has been found to have acceptable reliability and validity (Briesch et al., 2013).

**Intervention**

The REP intervention consists of two components: (1) CBI instruction delivered across five weeks and (b) a modified CICO procedure delivered daily over the course of 10 weeks. Additional information related to REP implementation within this study is provided below.

**CBI**. REP includes five CBI lessons that are delivered via small-group instruction. Each CBI lesson lasts approximately 30-45 minutes. All lesson plans are scripted and supported by instructional materials, including skill cards outlining the steps of cognitive-behavioral skills, graphic cards illustrating situations or skill use, and homework and activity sheets. Within this study, each CBI lesson was co-led by two individuals. Primary instructors were responsible for leading the lesson and guiding students through all portions of instruction. Co-facilitators then supported lessons by participating in role plays and managing student behavior. All primary instructors and co-facilitators were school psychology doctoral students who had completed coursework in psychotherapy and had experience delivering small-group instruction.

There was a total of six small groups in the intervention group; each small group included three to four students. CBI lessons for elementary students generally occurred during scheduled independent work time agreed upon by the teachers of students in each small group. Middle school students received CBI instruction during an elective class of their choice which was also agreed upon by their elective teachers. Elementary students were dismissed for CBI lessons by

their classroom teachers during whole-class transitions; middle school students independently transitioned from their prior class to small group CBI lessons.

Lesson 1 includes team building opportunities, as well as an introduction to REP and expectations for student behavior and participation. Lesson 2 then supports student identification of strong negative emotions (e.g., worry and sadness). REP instructors emphasize that these emotions are normal and expected, while also under our control when we use effective strategies. In Lesson 3, students are taught how negative emotions (e.g., worry) can be the result of negative thoughts (e.g., "I'm going to embarrass myself") that emerge in response to triggering scenarios (e.g., requests to speak in front of class). They are then taught basic cognitive restructuring skills, which involve the use of positive self-talk to challenge and replace negative thoughts. In Lesson 4, students are taught to use deep breathing and positive imagery to manage negative emotions. These skills are taught in the form of "STU" skills, which call for students to: See the triggers of their negative emotions (S), Take deep belly breaths (T), and Use their imagination (U). Lesson 5 then orients students to the "Think and Act" steps to problem-solving difficult situations. These steps involve the consideration of plausible solutions to these situations and then picking one that is likely to be mutually beneficial for the student and others.

**CICO**. Students in the REP group also participated in CICO, which began within a few days of the first CBI lesson. CICO lasted for the duration of CBI and for multiple weeks following to encourage generalization and maintenance of the skills taught through CBI. CICO is founded upon the DPR form, which follows the students throughout each school day and serves as a medium through which (a) students can evaluate their emotional functioning across multiple activities/periods of interest, and (b) educators can provide students feedback on their behavior during these activities/periods. Also foundational to CICO is each student's mentor: an adult

within the school building who conducts daily morning check-ins and afternoon check-outs, while also affirming their efforts to manage their emotions and behavior.

***DPR***. On a daily basis, participating teachers completed the DPR for each student in the intervention at the end of 3-4 designated class periods. At the end of each period, the teacher and student convened to evaluate how well the student "controlled emotions" and "made good choices." Each teacher and student also evaluated to what extent the student engaged in an additional positive behavior (e.g., "participated in class activities"), which had been specifically selected as a replacement behavior for their most problematic internalizing symptom specific to each student. Once the teacher and student discussed each behavior, the teacher used a 3-point Likert scale to rate how frequently the student exhibited each behavior (0 = *Never* to 2 = *Often*). Next, the student rated their emotions during the previous period. Specifically, students were prompted to rate "How I Felt" using a 7-point Likert scale, with each point including a corresponding descriptor and emoticon (1 = *Terrible!* to 7 = *Fantastic!*). Teachers then worked with students to consider their emotion ratings in relation to behavior ratings. For instance, if the student exhibited positive behavior but felt poorly, teachers could praise students for having a good period despite not feeling their best. Of possible concern was the potential for students to feel stigmatized due to the use of the DPR forms and CICO procedures. Likely due to the familiar practice of CICO as a common Tier 2 intervention within these schools, stigmatization was not a concern that arose.

***Check-Ins and Check-Outs.*** At the start of each day, REP group participants completed a daily check-in with their CICO mentor. Check-ins began with the mentor identifying the student's personalized goal for that day, which corresponded to a percentage of possible points earned on teacher ratings of student behavior across class periods. Goals were determined on a

weekly basis, corresponding to the average of the student's performance in the prior week. Students were informed that if they met their daily goal, they would earn a reward from a menu of items (e.g., snack or small toy) and activities (e.g., 5 minutes of computer time or drawing time). The student then selected the reward they would receive that day contingent upon goal attainment. Mentors then concluded check-ins by encouraging students to practice CBI skills throughout the day.

Students met with their mentors for check-out at the end of each school day. The mentor began by calculating the percentage of points the student earned throughout the day. The student then received the pre-determined reward if they met their goal. If the student had not met their goal, they were encouraged to try hard again the next day and use their CBI skills to manage their emotions and navigate difficult situations. To note, standard CICO protocol includes a home component, involving parents signing daily DPR forms and providing home-based rewards contingent upon goal attainment. However, participating schools elected to not include this component given concerns related to parent participation.

**Implementation Fidelity**

Co-facilitators monitored REP implementation fidelity via direct observation using adherence checklists inclusive of key intervention steps. On a small number of occasions, primary instructors completed fidelity checklists in a self-report manner when a co-facilitator was unavailable. CBI fidelity was evaluated for 28 of the 30 group lessons. Fidelity results indicated that the CBI lessons were implemented with 97.78% fidelity and a range from 88.89% to 100% fidelity. The most commonly missed item was reviewing homework, which CBI instructors failed to do on four of the observed occasions. Student attendance was also monitored; a total of four student absences across groups were documented. CICO fidelity was

assessed through direct observation by the first author with occasional assistance from trained graduate research assistants (also school psychology doctoral students). A total of 70 fidelity checks were completed, including 48 checks of the morning check-in procedures and 22 checks of the afternoon check-out procedures. CICO fidelity checks were conducted for 35% of the implementation days at Site 1 and 25% of the implementation days at Site 2. Results indicated CICO was completed with an average of 88.10% fidelity and a range from 42.86% to 100% fidelity. Lower fidelity ratings were during check-outs in which the student forgot to bring their DPR sheet with them to the check-out at the end of the day. The most commonly missed item was the mentor reminding students to use their CBI skills during the day.

## Procedure

**Recruitment**. Following receipt of Institutional Review Board and district-level approval, elementary and middle schools that were interested in participating were approached and recruited. To participate, schools had to be engaging in universal screening for internalizing or emotional risk. Once a school had been recruited, the research team approached classroom teachers to identify those interested in participation. Schools then provided the research team with universal screening data for all students in grades 4-7 who were enrolled in the classrooms of consented teachers. All data had been collected in the fall or winter of the 2017-2018 school year. Parental consent was then sought for all students identified via screening as exhibiting emotional or internalizing risk. The first author made initial phone calls to these parents, inviting their child to participate and providing basic information about the study. Follow-up phone calls and additional consent forms were provided as needed. Informed assent was then sought for students whose parents provided their consent. Following receipt of informed consent and assent, students were randomized into the two groups: (1) REP intervention or (2) waitlist control group.

**Training**. A total of eight CICO mentors were recruited for the study across participating schools. Mentors were school counselors, electives teachers, or educational support staff that had time available to check-in and check-out with students each day. A staff member was considered an acceptable mentor if they had a history of positive interactions with the student(s) in question, and thus had the potential to be viewed by the student as supportive and reinforcing. The first author trained CICO mentors through a single 30-minute session that included explicit instruction, modeling, and practice implementing all CICO procedures. All CBI primary instructors and co-facilitators received a two-hour training. The first author conducted this training in concert with the second and third authors. The principles of CBI were reviewed, as were the specific REP lessons, scripts, and instructional materials. Training consisted of explicit instruction relative to these topics along with modeling of implementation and opportunities to practice CBI delivery with performance feedback. Instructors and co-facilitators were also trained to evaluate the fidelity of CICO and CBI using the previously described checklists.

**Assessment and intervention**. Pretest measures were administered after randomization was complete. Students completed the YSR, CASSS, SRCS, and PCOISS during the school day, with the first author present to address student questions. For students in middle school, one core academic teacher completed the TRF for each student; for elementary school students, each student's primary homeroom teacher completed the TRF. The first author also conducted brief teacher interviews following TRF administration to identify replacement behaviors relevant to each student's internalizing symptoms. The replacement behaviors then became additional targets on each student's DPR. After pretest, students assigned to the REP group began intervention. REP was intended to be implemented over 10 weeks. However, due to time

constraints as a result of delayed school recruitment, inclement weather days, and state testing, REP was implemented over only seven weeks.

Following REP implementation, posttest measures were collected for participants in both groups. These measures were collected in the same fashion as at pretest. After completion of posttest measures at each school, students assigned to the waitlist control group received the REP intervention including the 5 CBI lessons along with daily CICO across a timespan of 4 to 6 weeks, depending on the amount of time left in the school year at each school. Finally, at the end of the school year and following all intervention implementation, CICO mentors and participating classroom teachers completed the URP-IR measure. A total of 13 educator participants completed the URP-IR, which equaled a completion rate of approximately 35%.

**Data Analysis Plan**

**RQ 1**. To address RQ 1 regarding the effect of REP on students' internalizing concerns compared to a waitlist control group, results were analyzed using multivariate analysis of covariance (MANCOVA) in IBM SPSS Statistics Version 25. Two MANCOVA models were tested, with one test specific to YSR scores and the other specific to TRF scores. Student intervention status served as the fixed factor independent variable in both models. REP effectiveness was then examined via the main between-group treatment effect. Student race/ethnicity, grade, and gender were also controlled for as covariates. YSR and TRF posttest scores on the internalizing-related scales noted above were included as dependent variables. Pretest scores on each of these scales were also included as covariates to control for student baseline functioning. In addition, ANCOVA tests were conducted with the YSR and TRF broad Internalizing Problems scale serving as the dependent variable. These follow-up analyses were used to further examine the overall effectiveness of the REP intervention while controlling for

pretest scores on the Internalizing Problems subscale. Internalizing Problems posttest scores were not included in either MANCOVA model due to concerns related to multicollinearity.

Additional follow-up ANCOVAs were conducted to determine if any of the individual outcome variables differed between the intervention and waitlist control groups. A Bonferroni adjustment was used to control for inflated type I error due to multiple univariate analyses. The critical p-value of .05 was divided by 8, due to eight comparisons that were conducted across outcome variables. Therefore, an adjusted critical p-value used for follow-up ANCOVA analyses equaled .006. Partial-eta squared ($\eta_p^2$) effect sizes were calculated for each main effect. In accordance with prior research, $\eta_p^2$ statistics were compared to standard interpretive criteria for judging the magnitude of effects. Statistics were considered small when greater than .01, medium when greater than .06, and large when greater than .14 (Richardson, 2011). The medium effect criterion was considered the threshold for practical significance within the context of this investigation. To note, we considered evaluating effects relative to interpretive criteria that were specific to the school-based targeted internalizing intervention literature. Unfortunately, recent meta-analytic research revealed only a small number of group-design studies that are specific to such interventions ($k = 4$; Sanchez et al., 2018). This suggested that any guidelines that could be derived from these studies would not necessarily be generalizable or applicable to this investigation.

**RQ 2**. An additional MANCOVA test was conducted to examine the effect of REP on the hypothesized change mechanisms. SRCS, PCOISS, and CASSS posttest scores served as dependent variables, while pretest scores were controlled for as covariates along with demographic factors. Again, student intervention status served as the fixed factor independent variable. REP efficacy was examined via the main between-group treatment effect. Follow-up

univariate ANCOVA analyses were then conducted, with each test considering a different posttest scale score and controlling for pretest scores. A Bonferroni correction was once again applied (adjusted critical $p$-value = .004). Effect sizes were calculated using $\eta_p^2$ values.

**RQ 3**. Descriptive statistics, including means and standard deviations, were calculated for each of the URP-IR subscales. Findings were then compared back to the original 6-point Likert scale in evaluating to what extent educator participants agreed with statements within each subscale. For the acceptability, understanding, feasibility, and system climate subscales, a higher score is considered more favorable, with mean item scores equal to or greater than 4 indicating teacher endorsement of REP as acceptable, understandable, etc. For the Home-School Collaboration and System Support subscales, a higher score would be considered less favorable, as it suggested participants perceived REP implementation as necessitating greater support. For these subscales, mean item scores equal to or less than 3 would be optimal, suggesting advanced parental or school system involvement (e.g., from administrators or trainers) is not necessary to implement REP with fidelity.

## Results

### Randomization Test

To determine whether randomization was successful in establishing equivalent groups, an independent samples t-test was conducted using pretest scores. The Levene's Test for Equality of Variances failed to reject the null hypothesis, indicating that the intervention and control groups were not statistically significantly different across groups for all pretest scores. The t-test for equality of means indicated that the intervention and control groups were equivalent across all measures with the exception of the School Support pretest scores ($p < .05$). This indicated that the CASSS School Support subscale pretest mean scores were significantly different between the

intervention ($M$ = 51.48, $SD$ = 16.83) and control groups ($M$ = 39.53, $SD$ = 19.31), such that intervention group participants reported higher levels of school support than participants in the control group. Overall, 19 of the 20 pretest variables indicated group equivalence.

**Missing Data**

A missing values analysis was conducted across all outcome variables and covariates. One participant in the waitlist control group dropped out of the study after moving to homeschooling prior to completion of posttest measures. Upon listwise deletion of this student from analyses, the final sample was equal to 38. Due to missed or skipped items, one participant failed to complete the entirety of the SRCS at posttest, while five participants failed to complete the entirety of the CASSS. Overall, this led to 2.6% missing YSR and TRF post-test data, 5.1% missing CASSS School Importance pre-test subscale scores, 5.1% missing CASSS Teacher Support post-test subscale scores, and 17.9% missing CASSS Teacher Importance post-test subscale scores. Little's Missing Completely at Random (MCAR) test was conducted to examine the pattern of data missingness. Results indicated the missing values reported above were likely to be missing completely at random, $x^2$ (368) = 31.32, $p > .05$. Given this MCAR assumption as well as the low level of overall data missingness (2.68%), single imputation of missing data using expectation maximization (EM) was conducted. All of the demographic, youth self-report, and teacher report variables were included in the EM algorithm to create a complete dataset.

**Assumption Testing**

The appropriateness of the data analysis plan was reviewed by testing the assumptions of the data for MANCOVA. This included a review of homogeneity of variance, independence of observations, and data normality. Homogeneity of variance was met through an independent samples t-test in which the intervention and waitlist control group pre-test scores were

determined to have equal variances. Though students were nested within schools and grade levels, thereby violating the assumption of independence, the sample was not large enough to account for nesting (e.g., via hierarchical linear modeling). This remains a direction of future inquiry. There were nine variables that violated normality, with skewness or kurtosis (+/- 1). These variables were transformed using logarithmic, square root, and reciprocal transformations. Following transformation, all variables were approximately normally distributed.

**Research Question 1**

Table 2 presents descriptive statistics for pre and posttest YSR and TRF scores across the intervention and control groups. The MANCOVA test specific to YSR scores revealed that while the main treatment effect was non-statistically significant, it was associated with a large effect size, Wilks' lambda $F(4) = 1.48$, $p = .24$, $\eta_p^2 = .19$. Similarly, the MANCOVA test specific to TRF scores indicated that while the main treatment effect was non-statistically significant, it was associated with a large effect size, Wilks' lambda $F(4) = 1.16$, $p = .352$, $\eta_p^2 = .15$. ANCOVA analyses specific to the broader Internalizing Problems scale scores revealed non-statistically significant effects for both the YSR and TRF. Effect sizes corresponding to these tests were small for YSR ($\eta_p^2 = .04$) and medium ($\eta_p^2 = .14$) for TRF.

Table 3 presents results of follow-up ANCOVA tests specific to YSR and TRF subscales. Although results indicated all main treatment effects were non-statistically significant, all of the YSR and TRF subscale effects were in the anticipated direction of change, in favor of the intervention group. For YSR outcomes, effect sizes fell in the small range for Anxious/Depressed, Social Problems, and Somatic Complaints, and large range for Withdrawn/Depressed. For TRF outcomes, medium effects were noted for

Withdrawn/Depressed, Anxious/Depressed, and Social Problems. The effect for Somatic Complaints was not found to reach the threshold for "small."

**Research Question 2**

Table 4 presents descriptive statistics for pre and posttest scores across the various measures of REP change mechanisms. MANCOVA tests including all subscales from these measures revealed a non-statistically significant main treatment effect, Wilks' lambda $F(10) = .86$, $p = .59$. However, the corresponding effect size fell in the large range ($\eta_p^2 = .38$). Follow-up univariate ANCOVA tests were then conducted for each of the individual subscales (see Table 3). None of the main treatment effects were statistically significant. Small effect sizes were found in favor of the intervention group for Teacher Support ($\eta_p^2 = .02$) on the CASSS, Perceived Control of Internal States ($\eta_p^2 = .02$), and the Distancing subscale on the SRCS ($\eta_p^2 = .05$). Small effect sizes were found for School Support ($\eta_p^2 = .04$) and medium effect sizes were found for Importance of School Support on the CASSS ($\eta_p^2 = .08$), both in favor of the control group. Large effect sizes were found for Importance of Teacher Support on the CASSS ($\eta_p^2 = .19$), in favor of the intervention group. The remaining effects did not reach the threshold for "small."

**Research Question 3**

Descriptive statistics were used to summarize scores for each of the URP-IR subscales. Mean item ratings were equal to 4.82 ($SD = .79$) with a range from 2 to 6 for Acceptability (e.g., "This intervention is a good way to handle the child's behavior problem"); 4.80 ($SD = .98$ ) with a range from 2 to 6 for Understanding (e.g., "I understand how to use this intervention"); 4.30 ($SD = 1.15$) with a range from 1 to 6 for Home-School Collaboration (e.g., "A positive home-school relationship is needed to implement this intervention"); 4.67 ($SD = 1.00$) with a range

from 1 to 6 for Feasibility (e.g., "I would be able to allocate my time to implement this intervention"); 4.97 (*SD* = 0.77) for Systems Climate with a range from 2 to 6 (e.g., "My administrator would be supportive of my use of this intervention"); and 3.95 (*SD* = 1.12) for System Support with a range from 2 to 5 (e.g., "I would need consultative support to implement this intervention"). The mean item ratings for the Acceptability, Understanding, Feasibility, and Systems Climate subscales fell in the optimal range of endorsement, with mean item scores equal to or greater than 4. However, the higher Home-School Collaboration and System Support mean item ratings indicate that home-school collaboration and system support are needed for implementation.

## Discussion

This study aimed to address a gap in the literature regarding feasible Tier 2 interventions for internalizing concerns for implementation in the school setting. An underpowered RCT design was employed with a sample of 38 elementary and middle school students who were randomly assigned to a REP intervention or waitlist control group. The study's primary purpose was to evaluate REP efficacy in reducing internalizing concerns for students, as reported by teacher and youth self-report measures. The secondary purpose of the current study was to evaluate the impact of REP on the variables presumed to serve as the change mechanisms by which primary outcomes would be observed. A final purpose of the study was to examine the degree to which educator implementors found REP to be acceptable, feasible, and usable.

### Research Question 1

Based on prior SCED findings (Allen et al., 2019), it was hypothesized REP would decrease internalizing concerns, as reported by students and teachers. In accordance with expectations given the intentionally underpowered nature of this study, MANCOVA tests

indicated there were not statistically significant differences in posttest internalizing concerns between the intervention and control groups. However, a review of descriptive statistics for YSR and TRF scores revealed a positive trend, wherein REP participants demonstrated reductions in internalizing concerns while control participants remained stable or worsened. Notable changes were documented on the TRF Anxious/Depressed and Internalizing Problems scales, with average *T* scores decreasing by more than half a standard deviation in the REP group.

This trend in findings was further supported by effect size estimates. MANCOVA tests suggested that REP was associated with large and practically significant treatment effects relative to YSR and TRF outcomes. Follow-up ANCOVA tests examined REP efficacy relative to the broad Internalizing Problem scale and related subscales. Findings suggested REP was associated with mostly small effects when considering youth self-report scores – this was with exception of student reported Withdrawn/Depressed scores, for which REP yielded a large treatment effect. Results were slightly more promising when considering teacher reports of student internalizing concerns, as REP was associated with medium and practically significant effects for three of the four subscales (with the exception of Somatic Complaints) and a large effect for the Internalizing Problems scale.

When taken together, it appears teacher perceptions of REP effects were greater than that of student participants. This might suggest that while behavioral manifestations of student emotional functioning improved, students did not necessarily perceive a substantial change in their own emotions and internal states. Reasons for this disparity are unclear. It could be that while the abbreviated implementation period (i.e., 7 weeks vs. the intended 10) was sufficient to change student behavior, it was insufficient to change the more fundamental emotions that underlie that behavior. There is the potential that more sustained implementation would result in

greater change. It is also possible that REP is not efficacious enough to support change in student emotions and internal states. Additional research supporting more sustained REP implementation is necessary to understand REP efficacy vis-à-vis both behavioral and emotional functioning.

**Research Question 2**

Based upon the REP theory of change, it was hypothesized that REP would increase student emotional control and social support. It was further hypothesized these the waitlist control group would not demonstrate an increase in emotional control. Interestingly, though the omnibus MANCOVA test was once again non-statistically significant, the associated effect size was quite large , indicating 38% of the variance between the groups in change mechanism scores was due to group assignment (after controlling for pretest scores and demographics). Closer examination of descriptive statistics and follow-up ANCOVA findings revealed few statistically significant between group differences between groups. However, a few interesting findings were noted.

First, a large, and practically significant effect was found for the CASSS subscale of Teacher Importance. This finding would suggest REP implementation resulted in increased student recognition of the importance of social support from teachers, including that related to the provision of information and resources, behavioral praise, and emotional support. Interestingly, REP did not impact student perceptions of the extent to which such support was actually available or provided to them or student perceptions of social support from people at their school more broadly. This finding might initially be considering somewhat concerning. However, increased student recognition of the importance of social support from teachers can still serve a key role in treatment, as openness to support increases youth help-seeking for internalizing concerns and eventual treatment outcomes (Ando et al., 2018). These findings are

in accordance with the theory of planned behavior, which suggests that intentions to engage in a behavior (e.g., seek out social support) are often a precursor of that behavior (Ajzen, 1991). Thus, there is the potential that improvements in the perceived importance of social support from teachers could be vital for REP success and should be incorporated into a revised theory of change (pending replication of findings).

Second, REP students self-reported a greater increase in emotional control. Specifically, there was an average 5.34-point increase in PCOISS scores for the intervention group, whereas there was only a 0.53-point increase in PCOISS scores for the waitlist control group. Though this difference was only representative of a small effect size (after controlling for covariates), it is nevertheless notable given that it corresponds to an increase of half a standard deviation (per both pre and posttest descriptive statistics). The finding of improvement in emotional control is certainly promising. REP CBI was designed to enhance students' perception that they are in control of their emotions and they can exert that control by using their cognitive-behavioral skills. Furthermore, research and theory have supported the role of emotional control within the treatment of internalizing concerns. Specifically, scholars have suggested emotional control is a key mediator of cognitive-behavioral interventions (Hofmann, 2000), as individuals exhibiting internalizing concerns frequently report a perceived lack of control over external events (White et al., 2006). That REP has the potential to influence student emotional control is important in documenting REP efficacy and the means by which it influences internalizing concerns.

**Research Question 3**

The acceptability of REP as a Tier 2 intervention to address internalizing concerns was measured using the URP-IR. Descriptive statistics indicated the mean of scores for the Acceptability, Understanding, and Feasibility subscales fell between the Slightly Agree and

Agree anchors. This indicates that educators who participated as CICO mentors or classroom teachers found REP acceptable and easy to understand and use. In addition, Systems Climate results indicated educators felt their administration would support REP and that it would align with existing intervention efforts. Educators also indicated that home-school collaboration would be important in applying REP. This finding is understandable given typical CICO protocols and the expectation that parents would work to also support student behavior in the home. Finally, results suggested educators indicated a need for system support when implementing REP, such as through the provision of training or consultation. This finding is supported by fidelity results, which suggested educators accurately implemented CICO following a brief training. Of course, it should be noted that these social validity findings are specific to the CICO portion of REP, as research assistants were responsible for CBI delivery.

**Limitations and Future Directions**

Certain limitations to this study should be noted. The first limitation pertained to the small sample size. As noted above, the study was underpowered given the early stage of REP research. Though this approach can potentially generate unbiased effect size estimates, it is nevertheless associated with certain limitations. For instance, an underpowered study does not permit more complex statistical analyses, including those that would have allowed testing of the hypothesized mediation effects (e.g., structural equation modeling). Future REP efficacy studies should therefore build upon this investigation by employing larger sample sizes in which intervention effects at the elementary and middle school levels can be separately evaluated. The importance of such disaggregated analyses should not be understated, as the manner in which REP is implemented will differ between elementary and middle school contexts; this is particularly true of the CICO portion of REP, as middle school students will need to interact with

multiple classroom teachers throughout the day, while elementary students will only interact with one teacher. Samples in future studies should also be representative of diverse populations within the United States to assess REP impact with youth from various racial/ethnic and cultural backgrounds. They should also be sufficiently large to limit the potential for sampling bias that could result in effects that are not representative of population-level parameters (Levine, Asada, & Carpenter, 2009).

Second, although internalizing concerns and REP change mechanisms were measured at pre and posttest, the study lacked a true longitudinal design. A longitudinal approach inclusive of additional measurement waves would afford multiple benefits. For instance, collecting change mechanism data prior to internalizing concern data would permit a more stringent evaluation of the extent to which change mechanism variables in fact mediate change in internalizing concerns. Collection of follow-up measurements subsequent to posttest would permit examination of (a) the extent to which changes in internalizing concerns were maintained over time, and (b) whether REP also influences the distal outcomes specified in the REP theory of change (see Figure 1). Given these benefits, future REP research should look to employ more intensive longitudinal designs.

Third, and related to the previous point, intervention duration was abbreviated within this study given recruitment delays and scheduling concerns; accordingly, the response rate for the URP-IR measure was lower due to limited time at the end of the school year to collect the information. The low response rate indicates that the URP-IR findings may not represent the perceptions of all implementers. Results should therefore be interpreted with caution; furthermore, future studies should seek to attain more representative and complete usability data from implementers to increase the accuracy and generalizability of usability findings. In

addition, the effect sizes resulting from this study are not characteristic of full REP implementation across a minimum of 10 weeks. Given the current effect sizes might underestimate REP efficacy, future research should ensure REP implementation occurs across the intended timeline. Fourth, given the observed lower rates of fidelity for certain elements of the CICO implementation (e.g., students returning DPR sheets, mentor prompts to use CBI skills), future studies should include additional supports during the intervention, such as targeted feedback sessions for CICO mentors based on fidelity checks, to encourage high rates of intervention fidelity.

Fifth, additional measures of outcomes should be analyzed in future studies. As previously noted, the YSR measure of self-reported internalizing concerns is validated for youth between the ages of 11-18; this is a limitation of the current study and future studies should employ self-reported measures that are validated for younger ages. Examination of a multi-informant approach across self-reported, teacher, and parent informants is an important consideration in evaluating the efficacy of REP. In addition to a quantitative evaluation of outcomes, quantitative and qualitative feedback from students regarding their perceived acceptability of the intervention components should be included in future studies.

Sixth, this study did not examine the role of parents in supporting REP implementation. As previously mentioned, social-ecological theory proposes that interactions with individuals such as family members may play a role in addressing internalizing concerns (Kilgus et al., 2015). Though REP was designed to involve parents as part of CICO procedures (i.e., through youth feedback and home-based reinforcement), the participating schools elected to not employ parent-related components due to concerns associated with fidelity of parent implementation. Future research should then look to include these components in examining to what extent they

might influence REP effects. Such research also might include novel REP components intended to enhance home-school collaboration and fidelity. Such components might include parent skills training, which research suggests can support parents in developing skills known to promote positive child behavior and reduce problem behaviors, including those related to internalizing concerns (Cartwright-Hatton et al., 2005).

Finally, though it is best practice in the evaluation of effect sizes, we were unable to examine the observed effect sizes in the context of previously reported effects (Wilkinson & Task Force on Statistical Inference, 1999). This was given the paucity of group design research specific to school-based targeted internalizing interventions (Sanchez et al., 2018). As this field of research continues to grow, we anticipate it will support the derivation of generalizable interpretive criteria against which the efficacy and effectiveness such interventions might be better judged.

**Practical Implications**

The current study provides additional research on a school-based internalizing intervention to overcome the barriers of existing interventions within the school setting, including lack of time, resources, and training for staff to implement interventions. REP is a time-limited intervention with five weekly 30-45-minute CBI lessons that can be implemented by a school mental health provider such as a school psychologist, school counselor, or school social worker. Daily CICO can be implemented by school staff. Based on fidelity outcomes monitored in the current study and acceptability and feasibility findings, the current study's brief CICO training appeared to adequately address training needs for CICO implementation.

With the inclusion of weekly school-based CBI lessons in a small group format, REP was perceived as a socially valid Tier 2 intervention according to teacher ratings of acceptability,

feasibility, and understanding on the URP-IR, which demonstrates promise as a potential intervention that can be implemented in the school setting. Future research should evaluate the impact of home-school collaboration on internalizing outcomes, particularly that of social support for students. In addition, school-based mental health providers generally have experience and training with small-group instruction, and it is expected that the brief nature of training provided in the current study would be acceptable for school mental health providers to implement CBI lessons. Future studies employing educators serving as implementers of the CBI curriculum should look to expand this social validity evidence to that portion of the REP intervention. Through further evaluation of the effectiveness of REP on decreasing internalizing concerns and increasing positive student coping skills and sense of social support, larger scale studies may provide further support for REP as a feasible school-based Tier 2 internalizing intervention.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. https://doi.org/10.1080/08870446.2011.613995

Allen, A. N., Kilgus, S. P., & Eklund, K. (2019). An initial investigation of the effectiveness of the Resilience Education Program (REP). *School Mental Health, 11*(1), 163-178. https://doi.org/10.1007/s12310-018-9276-1

Ando, S., Nishida, A., Usami, S., Koike, S., Yamasaki, S., Kanata, S., ... & Hiraiwa-Hasegawa, M. (2018). Help-seeking intention for depression in early adolescents: Associated factors and sex differences. *Journal of Affective Disorders*, *238*, 359-365.

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & Ehrenreich-May, J. (2011). *The unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. New York: Oxford University Press.

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology, 51*(1), 81-96. https://doi.org/10.1016/j.jsp.2012.08.006

Cartwright-Hatton, S., McNally, D., White, C., & Verduyn, C. (2005). Parenting skills training: An effective intervention for internalizing symptoms in younger children? *Journal of Child and Adolescent Psychiatric Nursing*, *18*(2), 45-52. https://doi.org/10.1111/j.1744-6171.2005.00014.x

Causey, D. L., & Dubow, E. F. (1992). Development of a self-report coping measure for elementary school children. *Journal of Clinical Child and Adolescent Psychology*, *21*(1), 47-59. https://doi.org/10.1207/s15374424jccp2101_8

Dart, E. H., Furlow, C. M., Collins, T. A., Brewer, E., Gresham, F. M., & Chenier, K. H. (2015). Peer-mediated check-in/check-out for students at-risk for internalizing disorders. *School Psychology Quarterly*, *30*(2), 229-243. https://doi.org/10.1037/spq0000092

Fergusson, D. M., Horwood, L. J., Ridder, E. M., & Beautrais, A. L. (2005). Subthreshold depression in adolescence and mental health outcomes in adulthood. *Archives of General Psychiatry*, *62*(1), 66-72. https://doi.org/10.1001/archpsyc.62.1.66

Hofmann, S. G. (2000). Treatment of social phobia: Potential mediators and moderators. *Clinical Psychology: Science and Practice*, *7*(1), 3-16.

Huang, F. L., Reinke, W. M., Thompson, A., Herman, K. C., & County Schools Mental Health Coalition. (2019). An investigation of the psychometric properties of the early identification system–student report. *Journal of Psychoeducational Assessment*, *37*(4), 473-485. https://doi.org/10.1177%2F0734282918758791

Kilgus, S. P., Bonifay, W., von der Embse, N. P., Allen, A. N., & Eklund, K. (2018). Evidence for the interpretation of Social, Academic, and Emotional, Behavior Risk Screener (SAEBRS) scores: An argument-based approach to screener validation. *Journal of School Psychology, 68,* 129-141. https://doi.org/10.1016/j.jsp.2018.03.002

Kilgus, S. P., & Eklund, K. (2017). *Resilience Education Program*. Unpublished curriculum and workbook.

Kilgus, S. P., Reinke, W. M., & Jimerson, S. R. (2015). Understanding mental health intervention and assessment within a multi-tiered problem-solving framework:

Contemporary science, practice, and policy. *School Psychology Quarterly, 30*(2), 159-165. https://doi.org/10.1037/spq0000118

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, *76*(3), 286-302. https://doi.org/10.1080/03637750903074685

Levitt, V. H., & Merrell, K. W. (2009). Linking assessment to intervention for internalizing problems of children and adolescents. *School Psychology Forum: Research to Practice, 3*(1), 13-26.

Malecki, C. K., Demaray, M. K., Elliott, S. N., & Nolten, P. W. (2000). *The Child and Adolescent Social Support Scale.* DeKalb, IL: Northern Illinois University.

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L, … Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Study-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(10), 980-989. https://doi.org/10.1016/j.jaac.2010.05.017

Mychailyszyn, M. P., Méndez, J. L., & Kendall, P. C. (2010). School functioning in youth with and without anxiety disorders: Comparisons by diagnosis and comorbidity. *School Psychology Review*, *39*(1), 106-121.

Pallant, J. F. (2000). Development and validation of a scale to measure perceived control of internal states. *Journal of Personality Assessment*, *75*(2), 308-337. https://doi.org/10.1207/S15327752JPA7502_10

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in

educational research. *Educational Research Review, 6*(2), 135-147. https://doi.org/10.1016/j.edurev.2010.12.001

Sanchez, A. L., Cornacchio, D., Poznanski, B., Golik, A. M., Chou, T., & Comer, J. S. (2018). he effectiveness of school-based mental health services for elementary-aged children: A meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(3), 153-165. https://doi.org/10.1016/j.jaac.2017.11.022

Webb, C. A., Stanton, C. H., Bondy, E., Singleton, P., Pizzagalli, D. A., & Auerbach, R. P. (2019). Cognitive versus behavioral skills in CBT for depressed adolescents: Disaggregating within-patient versus between-patient effects on symptom change. *Journal of Consulting and Clinical Psychology*, *87*(5), 484-490. http://dx.doi.org/10.1037/ccp0000393

Weist, M. D., Eber, L., Horner, R., Splett, J., Putnam, R., Barrett, S., ... & Hoover, S. (2018). Improving multitiered systems of support for students with "internalizing" emotional/behavioral problems. *Journal of Positive Behavior Interventions*, *20*(3), 172-184. https://doi.org/10.1177%2F1098300717753832

Weisz, J. R. (2015). Bridging the research-practice divide in youth psychotherapy: The deployment-focused model and transdiagnostic treatment. *Verhaltenstherapie*, *25*(2), 129-132. https://doi.org/10.1159/000430432

White, K. S., Brown, T. A., Somers, T. J., & Barlow, D. H. (2006). Avoidance behavior in panic disorder: The moderating influence of perceived control. *Behaviour Research and Therapy*, *44*(1), 147-157. https://doi.org/10.1016/j.brat.2005.07.009

Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association (1999). Statistical methods in psychology journals: Guidelines and

explanations. *American Psychologist*, *54*(8), 594–604. https://doi/10.1037/0003-066X.54.8.594

World Health Organization (2017). *Child and adolescent mental health.* http://www.who.int/mental_health/maternal-child/child_adolescent/en/

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, *67*(4), 272. https://doi.org/10.1037/a0028015

**Table 1**

*Student Participant Demographic Characteristics*

| Characteristic | Total *n* | % | REP Group *n* | Control Group *n* |
|---|---|---|---|---|
| Gender | | | | |
| Female | 25 | 65.8 | 15 | 10 |
| Male | 13 | 34.2 | 6 | 7 |
| Race/Ethnicity | | | | |
| Hispanic | 2 | 5.2 | 2 | 0 |
| Black/African American | 8 | 21.0 | 4 | 4 |
| White | 22 | 57.9 | 12 | 10 |
| Asian/ Pacific Islander | 1 | 2.6 | 0 | 1 |
| More than one ethnicity | 5 | 13.2 | 3 | 2 |
| Grade | | | | |
| Fourth | 13 | 34.2 | 8 | 5 |
| Fifth | 13 | 34.2 | 7 | 6 |
| Sixth | 9 | 23.7 | 5 | 4 |
| Seventh | 3 | 7.9 | 1 | 2 |
| Free-Reduced Lunch Status | | | | |
| Free | 12 | 31.6 | 7 | 5 |
| Reduced | 3 | 7.9 | 3 | 0 |
| Full Price | 9 | 23.7 | 4 | 5 |
| Not provided | 14 | 36.8 | 7 | 7 |

**Table 2**

*Descriptive Statistics for Pre and Posttest ASEBA Scales and Subscales*

| Measure | Variable | *Pre M* | *Pre SD* | *Post M* | *Post SD* | Δ[1] |
|---|---|---|---|---|---|---|
| YSR | Anxious/Depressed | | | | | |
| | Intervention | 60.90 | 10.33 | 60.57 | 9.21 | 0.33 |
| | Control | 65.82 | 11.71 | 65.65 | 11.79 | 0.17 |
| | Withdrawn/Depressed | | | | | |
| | Intervention | 63.71 | 12.22 | 62.52 | 11.89 | 1.19 |
| | Control | 67.24 | 10.60 | 71.06 | 11.07 | -3.82 |
| | Somatic Complaints | | | | | |
| | Intervention | 68.10 | 14.05 | 66.24 | 12.63 | 1.86 |
| | Control | 66.47 | 11.77 | 66.82 | 11.14 | -0.35 |
| | Social Problems | | | | | |
| | Intervention | 66.05 | 10.40 | 63.00 | 11.75 | 3.05 |
| | Control | 66.24 | 11.64 | 66.35 | 9.73 | -0.11 |
| | Internalizing Problems | | | | | |
| | Intervention | 63.19 | 13.73 | 63.05 | 11.81 | 0.14 |
| | Control | 67.65 | 11.09 | 68.71 | 10.49 | -1.06 |
| TRF | Anxious/Depressed | | | | | |
| | Intervention | 62.81 | 8.89 | 56.90 | 5.92 | 5.91 |
| | Control | 63.29 | 9.04 | 61.35 | 9.96 | 1.94 |
| | Withdrawn/Depressed | | | | | |
| | Intervention | 59.52 | 9.10 | 56.95 | 8.21 | 2.57 |
| | Control | 62.00 | 7.96 | 60.94 | 7.77 | 1.06 |
| | Somatic Complaints | | | | | |
| | Intervention | 57.90 | 8.96 | 55.05 | 7.39 | 2.85 |
| | Control | 57.29 | 8.21 | 55.88 | 8.04 | 1.41 |
| | Social Problems | | | | | |
| | Intervention | 60.76 | 8.98 | 58.14 | 6.15 | 2.62 |
| | Control | 60.18 | 8.36 | 60.12 | 8.75 | 0.06 |
| | Internalizing Problems | | | | | |
| | Intervention | 62.14 | 9.75 | 56.48 | 8.13 | 5.66 |
| | Control | 63.18 | 9.85 | 61.41 | 11.02 | 1.77 |

*Note.* YSR = Youth Self-Report; TRF = Teacher Report Form; Pre = pre-test score*;* Post = post-test score*; M* = mean; *SD* = standard deviation; Δ = difference between pre-test and post-test mean scores.

[1]A positive mean change indicates a decrease in concern from pre to posttest and therefore overall improvement, whereas a negative mean change indicates an increase in concern.

**Table 3**

*Analysis of Covariance (ANCOVA) Tests of Between-Group Treatment Effects for Internalizing Concerns and Change Mechanism Outcomes*

| Measure | Variable | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| YSR | Anxious/Depressed | 1.96 | .17 | .06 |
| | Withdrawn/Depressed | 6.01 | .02 | .17 |
| | Somatic Complaints | 0.99 | .33 | .03 |
| | Social Problems | 1.49 | .23 | .05 |
| TRF | Anxious/Depressed | 3.21 | .08 | .10 |
| | Withdrawn/Depressed | 2.10 | .16 | .07 |
| | Somatic Complaints | 0.20 | .66 | .01 |
| | Social Problems | 3.94 | .06 | .12 |
| CASSS | Teacher Support | 0.35 | .56 | .02 |
| | Importance of Teacher Support | 5.48 | .03 | .19 |
| | School Support | 0.86 | .36 | .04 |
| | Importance of School Support | 2.02 | .17 | .08 |
| PCOISS | Total Scale | .40 | .53 | .02 |
| SRCS | Seeking Social Support | 0.001 | .99 | .001 |
| | Self-Reliance/Problem-Solving | 0.04 | .84 | .002 |
| | Distancing | 1.12 | .30 | .05 |
| | Internalizing | 0.005 | .94 | .001 |
| | Externalizing | 0.002 | .96 | .001 |

*Note.* YSR = Youth Self-Report; TRF = Teacher Report Form; SRCS = Self-Report Coping Scale; PCOISS = Perceived Control of Internal States Scale; CASSS = Child and Adolescent Social Support Scale; $\eta_p^2$ = partial eta-squared value

**Table 4**

*Descriptive Statistics for Pre and Posttest Change Mechanism Measures*

| Variable | *Pre M* | *Pre SD* | *Post M* | *Post SD* | Δ[1] |
|---|---|---|---|---|---|
| CASSS Teacher Support | | | | | |
| Intervention | 52.91 | 14.38 | 50.25 | 17.37 | -2.66 |
| Control | 57.29 | 15.57 | 52.82 | 14.31 | -4.47 |
| CASSS Teacher Importance | | | | | |
| Intervention | 27.48 | 7.86 | 30.46 | 4.41 | 2.98 |
| Control | 29.41 | 5.93 | 27.20 | 5.48 | -2.21 |
| CASSS School Support | | | | | |
| Intervention | 51.48 | 16.83 | 52.67 | 19.17 | 1.19 |
| Control | 39.53 | 19.31 | 42.76 | 15.45 | 3.23 |
| CASSS School Importance | | | | | |
| Intervention | 27.67 | 6.76 | 31.23 | 5.72 | 3.56 |
| Control | 23.34 | 7.52 | 27.12 | 6.74 | 3.78 |
| PCOISS | | | | | |
| Intervention | 49.95 | 9.57 | 55.29 | 8.72 | 5.34 |
| Control | 49.53 | 12.10 | 50.06 | 14.45 | 0.53 |
| SRCS Seeking Social Support | | | | | |
| Intervention | 23.19 | 8.39 | 22.62 | 9.66 | -0.57 |
| Control | 20.65 | 7.47 | 19.68 | 8.21 | -0.97 |
| SRCS Self-Reliance/Problem-Solving | | | | | |
| Intervention | 21.95 | 5.09 | 22.81 | 8.58 | 0.86 |
| Control | 21.41 | 5.16 | 22.61 | 6.98 | 1.20 |
| SRCS Distancing | | | | | |
| Intervention | 15.57 | 4.18 | 15.76 | 5.24 | -0.19 |
| Control | 13.59 | 5.40 | 16.68 | 4.70 | -3.09 |
| SRCS Internalizing | | | | | |
| Intervention | 19.90 | 5.43 | 17.90 | 5.97 | 2.00 |
| Control | 21.53 | 6.27 | 19.48 | 7.43 | 2.05 |
| SRCS Externalizing | | | | | |
| Intervention | 8.52 | 3.56 | 9.05 | 3.26 | -0.53 |
| Control | 8.59 | 3.32 | 9.34 | 3.95 | -0.75 |

*Note.* Pre = pretest score; Post = posttest score*; M* = mean; *SD* = standard deviation; Δ = difference between pre-test and post-test mean scores. In addition, SRCS = Self-Report Coping Scale; PCOISS = Perceived Control of Internal States Scale; CASSS = Child and Adolescent Social Support Scale.

[1]A positive mean change indicates a decrease in concern from pre to posttest and therefore overall improvement, whereas a negative mean change indicates an increase in concern.

**Figure 1**

*Theory of Change for the Resilience Education Program*