Empirical Research

School-based Mental Health Interventions Targeting Depression or Anxiety: A Meta-analysis of Rigorous Randomized Controlled Trials for School-aged Children and Adolescents

Qiyang Zhang 1 · Jun Wang 1 · Amanda Neitzel 1

Received: 13 July 2022 / Accepted: 28 September 2022 / Published online: 13 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Past meta-analyses in mental health interventions failed to use stringent inclusion criteria and diverse moderators, therefore, there is a need to employ more rigorous methods to provide evidence-based and updated results on this topic. This study presents an updated meta-analysis of interventions targeting anxiety or depression using more stringent inclusion criteria (e.g., baseline equivalence, no significant differential attrition) and additional moderators (e.g., sample size and program duration) than previous reviews. This meta-analysis includes 29 studies of 32 programs and 22,420 students (52% female, 79% White). Among these studies, 22 include anxiety outcomes and 24 include depression outcomes. Overall, school-based mental health interventions in grades K-12 are effective at reducing depression and anxiety ($ES = 0.24$, $p = 0.002$). Moderator analysis shows that improved outcomes for studies with anxiety outcomes, cognitive behavioral therapy, interventions delivered by clinicians, and secondary school populations. Selection modeling reveals significant publication and outcome selection bias. This meta-analysis suggests school-based mental health programs should strive to adopt cognitive behavioral therapy and deliver through clinicians at the secondary school level where possible.

Keywords School-based mental health · Depression · Anxiety · Meta-analysis

Introduction
Given the escalating mental health crisis (Wong et al., 2021), there is a substantial need for effective ways to improve depression and anxiety among school-age children. There exists a prolific amount of school-based mental health interventions, yet the quality of such interventions varies greatly. Previous meta-analyses provided many insights on effective interventions; however, they failed to use stringent inclusion criteria. There is a need to review high-quality randomized controlled trials to generate new and evidence-based insights. This study aims to synthesize research on existing mental health interventions targeting depression and anxiety of school-aged children and adolescents to provide updated guidance on effective interventions.

Prevalence of Depression and Anxiety in School-aged Children and Adolescents
In recent years, depression and anxiety have increased rapidly among 6-17 years old American children (Centers for Disease Control and Prevention (2022)). Approximately 9.4% 3–17-year-old children were diagnosed with anxiety problems (Centers for Disease Control and Prevention (2022)) and 31.5% 13-18-year-old children have experienced depression (Feiss et al., 2019). These statistics are concerning not only because of what they tell, but also because of what they do not tell. In the mental health area, numbers often underestimate the actual prevalence of mental health problems due to diagnostic challenges (Mathews et al., 2011), stigmatization (Moses, 2010), and subsequent reluctance to seek help (Reavley et al., 2010). On top of that, there is a gap between those who are diagnosed and those who receive treatment: around half of diagnosed American adolescents receive mental health treatments in the form of medication or counseling (Zablotsky, 2020). More needs to be done to provide accessible mental health support for school-aged children.

Qiyang Zhang
qzhang74@jhu.edu

1 School of Education, Johns Hopkins University, Baltimore, MD, USA
and adolescents (i.e., aged 10–19; World Health Organization (2022)).

Apart from the increasing prevalence and inadequate treatment in depression and anxiety, the associated consequences highlight the need to intervene early and effectively. Past literature found that depression and anxiety among children are associated with poor academic outcomes (Owens et al., 2012), deteriorating physical health (Naicker et al., 2013), substance abuse or dependence (Conway et al., 2006), negative coping strategies (Cairns et al., 2014), self-injury (Giletta et al., 2012), and suicidal attempts (Nock et al., 2013). In addition, anxiety and depression among adolescents are likely to be recurrent (Gillham et al., 2006), chronic (Costello et al., 2003), and persistent through adulthood (Lee et al., 2018). Treating depression and anxiety effectively can create great social, educational, and economic benefits. A meta-analysis of randomized controlled trials can produce guidance for policymakers by providing insights into the effectiveness of all related interventions. This meta-analysis intends to provide evidence to help school districts understand what types of mental health interventions work in the school environment.

School-based Mental Health Interventions

The importance of addressing mental health in children and adolescents cannot be understated. The long-term adverse outcomes outlined above exist not only for those who meet diagnostic criteria, but also for those with subclinical levels of depression (Copeland et al., 2021). Schools provide an ideal setting within which to both implement preventative interventions as well as identify and serve those with or at-risk of depression or anxiety. School settings can provide access to all school-age children, while overcoming barriers such as location, time, and stigma (Stephan et al., 2007). Compared to primary-care settings, school-based mental health programs can reach larger populations, provide more convenient access, and enhance social relationships between classmates and teachers (van Loon et al., 2020). An additional advantage of school-based interventions is that they can serve to identify students at high risk for depression and anxiety and provide them with clinical support. This is especially important because one major reason for untreated depression is the failure to identify or diagnose depressive symptoms (Hirschfeld et al., 1997). This challenge can be mitigated by school-based mental health programs. With these advantages, school-based mental health interventions are increasingly gaining popularity (Werner-Seidler et al., 2017).

Past Meta-analyses

Past meta-analyses have conflicting views on the effectiveness of school-based interventions targeting depression and anxiety. One review identified 118 randomized controlled trials with 43,924 participants and found that these interventions had a small effect on depression and anxiety (Werner-Seidler et al., 2021). Similarly, another recent review analyzed 18 included studies and found that school-based programs had a small positive effect on self-reported anxiety symptoms (Hugh-Jones et al., 2021). These conclusions were challenged by another meta-analysis, where authors identified 137 studies with 56,620 participants and found little evidence that school-based interventions which focused solely on the prevention of depression or anxiety are effective (Caldwell et al., 2019). These three meta-analyses focused on children within the age group of 4–19 years old. When meta-analysts narrowed their focus to adolescents (11–18 years old) in the USA, they found significant effects of school-based programs on both depression and anxiety, but not on stress reduction (Feiss et al., 2019). One possible reason behind these conflicting conclusions is the distribution of age groups; perhaps school-based interventions are generally more effective for adolescents compared to younger children. Another possible source of conflict may be the subjective inclusion criteria created by researchers, which can bias the results (Cheung and Slavin, 2013). When applying more stringent inclusion criteria, the magnitude and statistical significance of effect sizes tends to diminish (Neitzel et al., 2022). The association between inclusion criteria and outcomes was demonstrated in another school-based meta-analysis, where the authors found that removing low-quality studies led to changes in average effect sizes (Tejada-Gallardo et al., 2020).

Outcome domain

Depression is a clinical symptom that involves persistent sadness and loss of interest in previously enjoyable activities (National Institute of Mental Health, 2018). Anxiety disorder refers to persistent anxiety that interferes with daily life (National Institute of Mental Health, 2022). In past meta-analyses, there are conflicting views on the effectiveness of depression- or anxiety-focused school-based interventions, ranging from no evidence of effectiveness in either depression or anxiety (Caldwell et al., 2019), to effectiveness dependent on program features (Feiss et al., 2019), to a small positive average effect size in both outcome domains (Werner-Seidler et al., 2021), to sustainable positive effect sizes even after 12 months (Hugh-Jones et al., 2021).

Program type

The universal approach delivers the treatment to the whole population (e.g., class, school, cohort) regardless of their
conditions while the targeted approach delivers the treatment to a selective group of students who show elevated symptoms of depression or anxiety. In targeted interventions, students were first screened using self-reported surveys to detect at-risk students. Universal programs are preventive in nature while targeted programs are curative. Each approach has its advantages and disadvantages. Targeted interventions may be difficult to apply at a large scale due to the tremendous amount of screening efforts required (Merry et al., 2004). Moreover, compared to the universal approach, the targeted approach involves taking students from classes, which may cause unintended effects of labeling and stigmatizing students (Huggins et al., 2008). Past research has shown that students selected for targeted mental health treatments feel embarrassed and may have negative attitudes towards receiving medication (Biddle et al., 2006). Universal programs avoid potential dangers of social stigmatization and can reach more children, but they may be more costly due to the larger group of participating populations (Ahlen et al., 2012). Furthermore, targeted programs were more effective compared to universal programs (Werner-Seidler et al., 2021). Since both approaches have pros and cons, investigating the effectiveness of one compared to another can generate useful scientific and implementation implications.

**Program content**

Cognitive behavioral therapy (CBT) is a traditional type of program targeting depression and anxiety (Feiss et al., 2019). In clinical settings, CBT treatment usually involves changing thinking and behavioral patterns, helping individuals to understand the problem, and developing a treatment strategy together with the psychologist (American Psychological Association, 2017). In K-12 school settings, CBT techniques and components can be adjusted and applied to the behavioral and emotional needs of students even without clinical diagnosis (Joyce-Beaulieu & Sulkowski, 2019). Through CBT interventions, students gain skills to understand and cope with their own feelings, such as using relaxation techniques and interacting with peers more effectively (Joyce-Beaulieu and Sulkowski, 2019). School-based interventions with CBT components are found effective to reduce depressive symptoms (Rooney et al., 2013) and anxiety (Lewis et al., 2013), and to improve coping strategies (Collins et al., 2014). Apart from CBT, evidence suggests promising effects of other innovative program features, such as physical education (Olive et al., 2019), student-family-school triads (Singh et al., 2019), and Hatha yoga sessions (Quach et al., 2016). One past meta-analysis suggested that there are no significant differences between CBT programs and other approaches in treating either depression or anxiety (Werner-Seidler et al., 2021). Yet, another meta-analysis reported weak evidence of CBT’s effectiveness in reducing anxiety in both elementary and secondary school populations (Caldwell et al., 2019). Comparing traditional program types to other new program types can help us understand components that make interventions effective.

**Delivery personnel**

To meet the rapidly growing demand for school-based mental health services, the teacher’s role shifts from being a supportive figure to being a service provider (Park et al., 2020). The appeal of hiring teachers to deliver interventions is increasing, yet there is lack of research that compares teacher-delivered interventions to clinician-delivered interventions. In school-based mental health interventions, interventions are normally delivered by either trained teachers or certified clinicians. Although trained psychologists or clinicians have more professional and practical knowledge than teachers, hiring specialists is more expensive compared to training schoolteachers through a short workshop. Moreover, students spend most of their time during schools with teachers and have developed rapport and trust with each other. In contrast to teachers, clinicians are less familiar with students’ backgrounds and personalities. Very few meta-analyses were able to include the provider as a moderator in their meta-regression because very few interventions were delivered by teachers. Findings on providers are mixed, ranging from no significant moderation effect on delivery personnel (Ahlen et al., 2015), to external providers being more effective than school staff (Werner-Seidler et al., 2021), to teachers being more effective (Neil and Christiansen, 2009), and to teachers being more effective under some treatment conditions (Franklin et al., 2017). This article contributes to the literature by comparing teacher-delivered interventions to clinician-delivered interventions.

**Program duration**

Past research has established that interventions delivered within a short span of time tend to produce much larger effect sizes than long-duration interventions (Cheung & Slavin, 2013). Factors contributing to this effect include novelty factors, a more experimentally-stable environment, and the feasibility of conditions only sustainable for short-duration interventions (Cheung & Slavin, 2013). Those findings suggest using 12 weeks as a benchmark to separate short-duration and long-duration studies.

**Sample size**

Study sample size has also been found to strongly impact effect sizes, with small sample sizes tending to inflate effect sizes.
sizes (Slavin & Smith, 2009). One reason behind this observation is the “superrealization” effect (Cronbach et al., 1980), which means that the high implementation fidelity maintained within a small sample can hardly be scaled to a larger sample. Another reason may be that small-scale studies are more likely to use researcher-developed measures compared to standardized tests (de Boer et al., 2014). Lastly, publication bias may have contributed to this phenomenon since small-scale studies have limited statistical power, which often requires higher effect sizes than large-scale studies to reach statistical significance. To the best of the authors’ knowledge, no prior meta-analysis has included sample size as a moderator.

Student age

As summarized in the previous section (i.e., past meta-analyses), results of the meta-analysis may be different across age groups. For example, one investigation of academic achievement found that students in the elementary grades gain much more academic progress than secondary school students in one academic year (Bloom et al., 2008). A separate meta-analysis of academic interventions found the same disparity between age groups, but the significance of these results disappeared when they restricted inclusion criteria to only randomized and quasi-experimental designs, which further demonstrated the importance of applying rigorous inclusion criteria (Cheung & Slavin, 2013).

Current Study

Past meta-analyses in mental health interventions failed to use stringent inclusion criteria and diverse moderators; therefore, there is a need to employ more rigorous methods to provide evidence-based and updated results on this topic. In order to identify characteristics of effective depression or anxiety interventions and provide evidence-based suggestions for current and future practices, this meta-analysis systematically reviews school-based mental health programs that serve K-12 students. The article aims to answer two research questions: What are the overall impacts of school-based randomized-controlled-trial programs on depression and/or anxiety reduction? (Research question 1). This study hypothesizes that school-based randomized-controlled-trial programs can effectively reduce depression and anxiety in school-age children, as has been demonstrated in prior reviews (Feiss et al., 2019; Hugh-Jones et al., 2021; Werner-Seidler et al., 2021). (Hypothesis 1). To what extent do intervention outcomes differ according to methodological criteria, such as sample size and program duration; and intervention criteria, such as program type (targeted vs. universal); program delivery personnel (trained teachers vs. certified clinicians); program content (cognitive behavior therapy vs. others); and student age? (Research question 2). This study hypothesizes that the identified factors significantly moderate the impact of the programs. (Hypothesis 2).

Methods

Literature Search

To ensure a thorough literature search, the first author conducted database search, handsearching, andbackward citation chasing. Using the list of keywords (Table 1), the first author searched for relevant articles in the Education Resources Information Center, PsychINFO, and Google Scholar. After the database search, the first author also conducted complementary handsearching in a set of reputable field-specific journals (listed in Table 2). The first author used Paperfetcher, a browser-based tool, to conduct both handsearching and backward citation chasing (Pallath & Zhang, 2022). All literature search was completed in June 2021. There are three steps in the screening: first, the first author used ASReview’s machine learning algorithms to rank all retrieved articles by their relevance and exported the most relevant 2308 studies (van de Schoot et al., 2021). This decision was made by the first author by manually browsing through the titles of the most relevant studies and making a cut-off point when studies become obviously irrelevant. To be cautious, the first author also sampled 10% of the 5864 studies excluded through ASReview’s machine learning ranking to confirm that they would be excluded. Then, these 2308 studies were imported into Covidence (Covidence systematic review software, 2013) for double-blinded screening. The authors chose Covidence here because it enables full-text review and the authors’ affiliated institutions provide free software licenses for this tool (Zhang & Neitzel, 2021). The first and second authors manually screened the title and abstract through a single-reviewer decision-making. Then the first and second authors manually screened full-text through a double-blinded decision-making. The proportionate agreement in this full-text screening stage is 0.85. The third author was invited for a group discussion to resolve all conflicts and 100% agreement was reached eventually.

1 In the US educational system, K-12 refers to formal education covering elementary and secondary school grades through kindergarten to 12th grades (Department of Homeland Security, 2022). American children generally start formal education at age five or six. However, educational systems in other countries may be different from the US. For example, kindergarten is not included in elementary school grades in some countries. Thus, this study examines the effects of interventions for children and adolescents in kindergarten, elementary school, middle school, and high school.
Eligibility Criteria

This article used the following inclusion criteria for full-text review to ensure consistency and high standards in study quality:

1) Randomized controlled trials (RCT) must have at least 30 students per experimental condition to reduce bias in small studies (Cheung & Slavin, 2016) and at least 2 teachers/schools per condition to eliminate confound due to sample size (What Works Clearinghouse, 2020a). For example, one study was excluded because it used a one school vs. one school design (Harnett & Dadds, 2004). Another study was excluded because it had 22 and 24 students in control and treatment conditions respectively (Burckhardt et al., 2016).

2) Program duration from program start to posttest must be at least four weeks to remove particularly short interventions. A study was excluded because it was a one-week intervention (Link et al., 2020).

3) Studies must have taken place in the following countries: USA, Canada, Europe (European Union + U.K. + Switzerland + Norway), Israel, Australia, and New Zealand. This geographical restriction intends to narrow down the scope of review to countries that share similar economic and political situations. A study was excluded because it took place in South Africa (Fernald et al., 2008).

4) Studies must use randomization to focus on studies of the highest level of internal validity. The level of random assignment may be schools, classes, or students. For example, a study was excluded because it allocated students to treatment and control based on number of students and gender composition (Kowalenko et al., 2005).

5) Differences between conditions at baseline on depression/anxiety measure must be less than 0.25 standard deviations (SDs) to reduce bias from unreliable statistical analyses (Rubin, 2001). For example, a study was excluded because its depression measurements at baseline were inequivalent (Ardic & Erdogan, 2017).

6) Differential attrition between treatment and control groups must be less than 15% to reduce bias (What Works Clearinghouse, 2020a).

7) A control group must be present.

8) Intervention or instruction should be delivered by non-researchers. Treatments had to be delivered by ordinary teachers, not by researchers, because effect sizes are inflated when researchers deliver the treatment (Scammaca et al., 2007). For example, a study was excluded because the intervention was delivered by lead author and master’s students in clinical psychology (Garcia-Escalera et al., 2020). Similarly, a study was excluded because the intervention was delivered by lead author and a research assistant (Burckhardt et al., 2016).

9) Outcomes of interest measurements must include quantitative measures of either depression/depressive symptoms/depression literacy or anxiety/anxious symptoms. For example, a study was excluded because outcomes did not measure depression or anxiety (Daunic et al., 2012).

10) Text must be available in English.

11) Articles must be published on or after January 1st, 2000. This means that this study reviewed experimental studies in the last two decades for the most updated evidence.

12) Children participating in the programs must come from K-12 school grades. This means that this study includes...
studies that focus on students studying in kindergarten, elementary, secondary (including middle and high schools). For example, a study was excluded because the intervention focused on college students (Xiong et al., 2022).

Analytical Plan

This study conducted the meta-analysis in R (R Core Team, 2021) using the metafor package (Viechtbauer, 2010). The three authors performed double coding in Google spreadsheet and held discussion to eventually reach 100% inter-judge reliability in codification. This study used weighted mean effect sizes and meta-analytic tests such as Q statistics. Weights were then assigned to each study based on inverse variance (Lipsey & Wilson, 2001) and adjusted weights (Hedges, 2007). A random-effects model was used in meta-regression since there was no single true effect size but a range of effect sizes that may have depended on other factors (Borenstein et al., 2010). For cluster randomization, this study added adjustments for clustering adapted from What Works Clearinghouse (2020b). This study analyzed six pairs of moderators (more details in the next section) and examined differential effects by including interaction terms. All moderators and covariates were grand mean centered to facilitate interpretation of the intercept. All reported mean effect sizes come from this meta-regression model, which adjusts for potential moderators and covariates. To assess publication bias, this study adopted selection modeling instead of other traditional techniques (e.g., funnel plot, Egger’s regression, fail-safe N) because of the limitations in these traditional techniques. Selection modeling involves a model of the selection process that uses a weight function to estimate the probability of selection in random-effect meta-analysis (Hedges, 1992). Selection modeling is the most recommended method to investigate meta-analyses’ publication bias (Terrin et al., 2005). This study used weightr package (Coburn & Vevea, 2019) to apply the weight-function model (Vevea & Woods, 2005). To conduct risk of bias analysis, the authors decided to use JBI Checklist for Randomized Controlled Trials (Joanna Briggs Institute, 2017) because it provides a clear guideline for systematic review and meta-analysis on randomized controlled trials, which fits well with this meta-analysis.

Following the open science movement, the complete dataset and code are publicly available at https://github.com/qiyanngzh/School-based-Mental-Health-Interventions-Targeting-Depression-or-Anxiety-A-Meta-analysis.

Moderators

This study added moderators to explain the difference in impacts based on outcome domains, methodological criteria, intervention criteria, program delivery personnel, program content, and student age.

Outcome domain: depression vs. anxiety

Outcome domain was coded as depression if the outcome used a depression scale, such as Children’s Depression Inventory and Beck Depression Inventory II. It was coded as anxiety if the outcome included used an anxiety scale, such as Spence Children’s Anxiety Scale and Multi-dimensional Anxiety Scale for Children.

Program type: universal vs. targeted

Programs were coded as either universal or targeted. Universal means the program was delivered to entire classes, schools, or cohorts, regardless of risk level. Targeted means the program was delivered to specific groups of students showing elevated levels of depression of anxiety.

Program content: cognitive behavioral therapy vs. others

Programs were coded as either cognitive behavioral therapy or others. CBT programs have design features that involve CBT techniques and components. Non-CBT programs refer to a wide range of other designs, including mental health education, yoga, physical education, mindfulness, coping skills.

Delivery personnel: teachers vs. clinicians

Delivery personnel was coded as either teachers or clinicians. Among non-teacher delivered programs, one intervention was delivered through a self-directed online program; others were delivered by personnel with psychological backgrounds, such as psychologists, mental health workers, clinicians, social workers, and counselors. In the following analysis, this study treats all non-teacher delivery personnel as clinicians.

Program duration: short duration (<12 weeks) vs. long duration (≥12 weeks)

This article codes interventions shorter than 12 weeks as short duration and interventions equal to or longer than

---

2 The evaluation of funnel plots is subjected to meta-analysts’ interpretations, which are often misled by the plot shapes (Terrin et al., 2005). In addition, visual assessment of funnel plots or Egger’s regression only presents part of the publication bias related to small-study bias. The fail-safe N technique is also largely abandoned in the field due to its arbitrary choice of zero, ignorance of heterogeneity in primary studies (Iyengar & Greenhouse, 1988), and other major flaws (Becker, 2005).
12 weeks as long durations to compare the effect sizes of the two types.

**Sample size: small sample (<250) vs. large sample (≥250)**

Sample size was either small or large. When the total participating student sample is smaller than 250, the program was coded as having a small sample size. When the sample is equal or larger than 250, the program was coded as having a large sample size.

**Student age: grade level comparison between elementary and secondary schools**

Student age was coded as 1 for kindergarten and elementary schools, 2 for middle schools, and 3 for high schools. In analysis, 2 and 3 are combined to refer to secondary school students.

**Results**

**Descriptive Results**

This study retrieved both published studies and unpublished studies to minimize publication bias in this review. Figure 1 presents the PRISMA screening process in Covidence. Among the 218 excluded studies in the full-text review stage, the top five reasons for exclusion are inadequate outcome measures (n = 52), irrelevant (n = 46), wrong design (n = 33), meta-analysis/review (n = 23) and outside the geographical scope (n = 17). After applying the inclusion criteria, this review found a total of 29 qualified studies evaluating 32 programs (Fig. 1, Table 3). In total, these programs have 22,420 K-12 students (n = 12,174 in treatment group, and n = 10,246 in control group). Among these students, 52% are female and 79% are White. Table 4 presents the descriptive statistics of the 29 included studies. Among the included studies, most were conducted in Australia and USA (13 and nine respectively, Table 4). For grade levels, 17 programs (53.1%) focused on the elementary school population and 15 (46.9%) focused on the secondary school population. Overall, 22 programs (68.8%) employed Cluster Randomized Controlled Trials (CRCT) and the other 10 programs (31.2%) used RCT design. In terms of intervention design, 22 programs (68.8%) examined interventions with CBT components and 10 programs (31.2%) reported interventions with non-CBT strategies. In terms of duration, 10 programs (31.2%) evaluated interventions that last at least 12 weeks (i.e., long duration), while 22 programs (68.8%) reported interventions that last

---

**Fig. 1 PRISMA Diagram**

| Additional records identified through handsearching and citation tracking (n = 4935) | Records identified through various database searching (n = 6,111) |
| --- | --- |
| Total records identified (n = 11,046) | Duplicates removed (n = 2627) |
| Records screened (n = 8419) | Records excluded (n = 8172): ASReview machine learning algorithm (n = 5864) Manual screening (n = 2308) |
| Full-text articles assessed for eligibility (n = 247) | Full-text articles excluded (n = 218) |
| Citations included in review (n = 29) |
less than 12 weeks (i.e., short duration). Sample sizes of included programs vary greatly, ranging from 68 students (Chaplin et al., 2006) to 5,634 students (Sawyer et al., 2010). In total, 17 programs (53.1%) have small sample sizes (less than 250) and 15 (46.9%) have large sample sizes. The majority (78.1%) of the interventions were universal, and seven programs (21.9%) are targeted. This demonstrates universal programs’ increasing popularity. This study analyzed 79 effect sizes related to depression or anxiety. Among these outcomes, 40 effect sizes (50.6%) were related to anxiety, and 39 effect sizes (49.4%) were related to depression. In addition, 40 (50.6%) studies evaluated teacher-delivered interventions and 39 (49.4%) studies investigated non-teacher-delivered interventions.

Meta-analysis Results

As shown in Table 5, the overall mean effect size for these 32 programs is 0.24 ($p = 0.002$) while holding all moderators fixed at their mean. The 95% predictive interval ranges from −0.91 to 1.39.

Outcome domain: depression vs. anxiety

Overall, outcomes using anxiety-related measurements have a significant weighted mean effect size ($ES = 0.44$, $p = 0.001$, Table 6). But no significant mean effect size was found in outcomes measuring depression ($ES = 0.04$, $p = 0.723$). The difference between anxiety outcomes and depression outcomes was 0.4 $SD$s on average ($p = 0.025$).

Program type: universal vs. targeted

Intervention type was not a significant moderator of effect sizes. The mean effect size for interventions focused on targeted populations was 0.42 ($p = 0.021$), while the mean effect size for interventions focused on universal populations was 0.18 ($p = 0.028$).

Intervention design: cognitive behavioral therapy vs. others

Intervention design was a significant moderator of impact. On average, CBT programs have significantly higher effect sizes than those without CBT components ($p = 0.016$). CBT programs have a mean effect size of 0.33 ($p = 0.002$), while programs without CBT elements have a non-significant mean effect size of −0.15 ($p = 0.260$).

Intervention delivery: teachers vs. clinicians

Intervention delivery personnel were a significant moderator of impact. On average, the effect size of teacher-delivered programs is 0.39 $SD$s lower than programs delivered by non-teacher personnel ($p = 0.013$). Programs delivered by non-teacher personnel have, on average, a significant mean effect size of 0.44 ($p = 0.007$) while programs delivered by teachers have an average non-significant mean effect size of 0.05 ($p = 0.371$).

Intervention duration: short duration (<12 weeks) vs. long duration (≥12 weeks)

No significant difference was found between the mean effect size of long-duration and short-duration programs. On average, short duration programs had a statistically significant mean effect size ($ES = 0.28$, $p = 0.003$), compared to long duration programs with a non-significant mean effect size of 0.14 ($p = 0.222$).

Sample size: small sample (<250) vs. large sample (≥250)

The weighted mean effect size of interventions with small sample sizes ($n < 250$) is 0.35 ($p = 0.019$), while the mean effect size of large-sample interventions ($n \geq 250$) is 0.13 ($p = 0.221$). The difference between the mean effect sizes of small-sample and large-sample interventions is not significant ($p = 0.271$).

Grade level: elementary vs. secondary schools

On average, there was no significant effect found in the interventions implemented in elementary schools ($ES = 0.06$, $p = 0.547$). Interventions implemented in secondary schools had a significant average effect size of 0.42 ($p = 0.006$). The difference in the effect sizes between secondary school population and elementary school population is 0.36 $SD$s, which is marginally statistically significant ($p = 0.076$).

Interactions

The interaction analysis identified a marginally significant differential effect for personnel by presence of CBT ($\beta = -0.75$, $p = 0.059$). On average, clinicians have higher effect sizes with CBT ($ES = 0.59$, $p = 0.007$) than they do without CBT ($ES = -0.27$, $p = 0.325$). Teachers have smaller average effect sizes, with average effect sizes for CBT of 0.07 ($p = 0.295$) and for without CBT of −0.04 ($p = 0.732$).

In addition, a significant differential effect for personnel also was identified by intervention type ($\beta = 0.83$, $p = 0.024$). On average, interventions delivered by clinicians for targeted populations have a larger effect size ($ES = 0.92$, $p = 0.023$) compared to interventions for the universal population ($ES = 0.26$, $p = 0.044$). On the contrary, although no significant effect sizes were found by teacher-delivered interventions for both universal...
| Study                  | Program name                  | Age       | Country   | Type | N (T, C) | Duration                                      | Design   | Control          | Program content                                                                 | Targeted outcomes                                                                 | Program delivery                                                                 |
|-----------------------|-------------------------------|-----------|-----------|------|----------|------------------------------------------------|----------|------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| Barnes et al. (2012)  | The William LifeSkills        | 14–17     | USA       | U    | 159 (86,73) | Twelve 50-min weekly sessions, 3 months       | CRCT     | Health education | Workshop trains in anger and stress coping skills                              | Anger and stress management (anxiety and blood pressure)                        | Teachers at school                                                            |
| Barrett & Turner (2001) | Friends for Children          | 10–12     | Australia | U    | 325 (188, 137) | 75-min per session, 10 sessions              | CRCT     | Usual care       | CBT                                                                            | Depression and anxiety                                                          | Psychiatrists and Teachers                                                     |
| Britton et al. (2014) | Mindfulness meditation        | 11–12     | USA       | U    | 100 (52, 48)  | Daily meditation for 6 weeks                 | CRCT     | Experiential activity | Based on integrative contemplative pedagogy                                   | Student acceptability, engagement, clinical symptoms                           | Two history teachers who completed mindfulness training |
| Chaplin et al. (2006) | Penn Resiliency Program       | 11–14     | USA       | U    | 103 (65, 38)  | 90 min/week for 12 weeks                     | RCT      | Co-ed            | CBT and social problem-solving intervention                                  | Depression                                                                      | Teachers, guidance counselors with a week-long training                      |
| Collins et al. (2014) | Lessons for living: Think well, Do well | 9–10      | UK        | U    | 238 (103, 135) | Ten-lesson program                             | CRCT     | Regular personal and social education | The intervention was theoretically grounded in CBT for the development of coping skills. | Anxiety, coping                                                              | Psychologists and Teachers                                                |
| DeRosier (2004)       | S.S.GRIN (Social Skills GRoup INtervention) | 8–11      | USA       | T    | 381 (187, 194) | 8 weeks, 50 to 60 mins, weekly               | RCT      | No-treatment control group | Social responsibility training, S.GRIN combines social learning and cognitive–behavioral techniques | Social anxiety and depression                                                 | School’s counselor and a trained undergraduate intern                |
| Haugland et al. (2020)| Cool Kids Program             | 12–16     | Norway    | T    | 201 (106, 95)  | 10 weeks: total 15 h plus two 90-minute parents-only sessions | RCT      | Waitlist         | CBT                                                                            | Anxiety                                                                        | School personnel or mental health workers                                    |
| Vaag                  |                               |           |           |      | 177 (82, 95)  | 5 sessions: with weekly sessions of 45 to 90 min (total 5.5 hours) over the first 4 weeks, followed by a final session 5 weeks later | RCT      | Waitlist         | CBT                                                                            | Anxiety                                                                        | School personnel or mental health workers                                    |
| Study                  | Program name                | Age   | Country       | Type | N (T, C)         | Duration                      | Design | Control | Program content                                                                 | Targeted outcomes                 | Program delivery                  |
|-----------------------|-----------------------------|-------|---------------|------|-----------------|-------------------------------|--------|---------|---------------------------------------------------------------------------------|-----------------------------------|----------------------------------|
| Kraag et al. (2009)   | Learn Young, Learn Fair     | 10    | Netherlands   | U    | 1384 (652, 732) | 8 weekly one-hour lessons, 5 weekly one-hour booster sessions | CRCT   | Waitlist | Addressed stress, stress awareness, coping skills and possible barriers and was developed using the Intervention Mapping protocol | Anxiety, depression, stress symptoms, coping, stress awareness | Teachers                          |
| Lewis et al. (2013)   | Positive Action             | 12–13 | USA           | U    | 1170 (585, 585) | 6 years                       | CRCT   | Business as usual | Based on self-esteem enhancement theory | Positive affect, life satisfaction, depression, anxiety | Teacher, counselor, and family training |
| Lowry-Webster et al. (2001) | FRIENDS                 | 10–13 | Australia     | U    | 531 (392, 139)  | 10 sessions                   | CRCT   | Waitlist | CBT | Depression and anxiety | Teachers                          |
| Merry et al. (2004)   | RAP-Kiwi Program           | 13–15 | New Zealand   | U    | 332 (178, 154)  | 11 sessions for 11 weeks or 6 weeks | RCT    | Placebo | CBT and interpersonal therapy | Depression | Teachers with 2.5-days of training |
| Mifsud & Rapee (2005) | Cool Kids Program          | 9–10  | Australia     | T    | 91 (51, 40)     | 1-hour sessions, 8 sessions, one per week | CRCT   | Waitlist | Education about anxiety | Anxiety, internalizing problems | School counselors and social workers |
| Miller et al. (2011)  | FRIENDS Program            | 10.1  | Canada        | T    | 180 (61, 119)   | 9 weeks                       | CRCT   | Attention-control | CBT | - | Implemented by the trained school person (i.e., teacher) paired with a trained school counselor (or trained psychology graduate student) |
| O’Kearney et al. (2009)| MoodGYM                    | 15–16 | Australia     | U    | 157 (67, 90)    | 6 weeks                       | RCT    | Usual health education | Coping skills to tackle depression | Depression and anxiety | Online intervention |
| Olive et al. (2019)   | LOOK                       | 8     | Australia     | U    | 642 (335,307)   | Two 50-minute sessions/week over four years | CRCT   | Similar to intervention but moderated | Specialist-taught physical education: emphasize enjoyment of movement, de-emphasize individual competition, reflective discussion, engaging every child | Depression, body image, and stress | University trained and qualified PE teachers |
| Quach et al. (2016)   | Mindfulness Meditation     | 12–17 | USA           | U    | 107 (54, 53)    | 45 min/session, 2 sessions/week for 4 weeks | RCT    | Waitlist | Based on the mindfulness-based stress reduction (MBSR) program developed by Jon | Working memory, capacity, stress, anxiety | Two instructors with 5-10 years of experience in meditation |
| Study | Program name | Age | Country | Type | N (T, C) | Duration | Design | Control | Program content | Program delivery | Targeted outcomes |
|-------|--------------|-----|---------|------|---------|----------|--------|---------|----------------|----------------|------------------|
| Roberts et al. (2003) | Penn Prevention Program | 11–13 | Australia | T | 189 (90, 99) | 12 sessions | CRCT | Usual health education classes | Kabat-Zinn, breathing techniques, discussion, formal meditation, CBT | Depression, anxiety | School psychologists and nurses |
| Roberts et al. (2010) | Ottawa Optimism Program | 8–9 | Australia | U | 428 (257, 191) | 20 weeks | CRCT | Usual health classes | Depressive symptoms, CBT intervention strategies | Depressive and anxiety | Teachers |
| Rooney et al. (2006) | The Positive Thinking Program | 9–10 | Australia | U | 910 (467, 443) | Ten weekly 60-min sessions | CRCT | Health education | CBT intervention strategies | Internalizing problems and protective factors | Classroom teachers and nurses |
| Rooney et al. (2013) | Aussie Optimism: Positive Thinking Skills Program | 9–10 | Australia | U | 5634 (3037, 2597) | 60-minute session, 20 weeks | CRCT | Regular health curriculum | CBT intervention strategies | Depression | Teachers |
| Sawyer et al. (2010) | Beyondblue Program | 9 | Australia | U | 5634 (3037, 2597) | Ten weekly 60-min sessions | CRCT | Business as usual | Integrate two major CBT components: cognitive restructuring and problem-solving skills training | Depression, anxiety, and externalizing problems | Teachers |
| Shochet et al. (2001) | Resourceful Adolescent Program (RAP) | 13–15 | Australia | U | 1248 (634, 614) | One 45-min session per week for 8 weeks | CRCT | No intervention | CBT | Observation, previous cohort simply complete the measures | Teachers |
| Sinclair (2016) | Think, Be, Do Program | 14–21 | USA | U | 115 (74, 41) | 10 sessions over five weeks | CRCT | Business as usual | CBT | Based on CBT and ecological theory | Knowledge, mental health, academic and social skills training |
| Skryabina et al. (2016) | FRIENDS Program | 9–10 | UK | U | 844 (457, 387) | 12 months | CRCT | Usual personal, social health, and education | Health care staff | Depression, anxiety, and externalizing symptoms | Teachers |

Table 3 (continued)
Table 3 (continued)

| Study | Program name | Age | Country | Type | N (T, C) | Duration | Design | Control | Program content | Targeted outcomes | Program delivery |
|-------|--------------|-----|---------|------|----------|----------|--------|---------|-----------------|------------------|------------------|
| Spence et al. (2005) | The problem solving for life | 12–14 | Australia | U | 1500 (751, 749) | 45-50 min weekly, 8 sessions | CRCT | Monitoring control condition | CBT | Depression | Teachers |
| Stein et al. (2003) | Cognitive-Behavioral Intervention for Trauma in Schools (CBITS) | 9–10 | USA | T | 117 (54, 63) | 10 sessions | RCT | Delayed intervention | CBT | PTSD, depression, psychosocial dysfunction, classroom problems | Trained school mental health clinicians with 2-day training and weekly group supervision |
| Tak et al. (2016) | Op Volle Kracht (OVK, based on Penn Resilience Program) | 14 | Netherlands | U | 1341 (634, 707) | 16 50-min lessons, a 2 h booster session delivered at 12 months follow up | CRCT | School as usual | Based on several psychological theories and principles, namely CBT, the ABC model (Ellis, 1991), and the hopelessness theory of depression | Depression | Psychologists with varying degrees of experience in CBT and teaching |
| Volanen et al. (2020) | Mindfulness-Based interventions | 12–15 | Finland | U | 2258 (1157, 1101) | Nine weekly 45-min sessions | CRCT | Both | Mindfulness: attending, observing, and accepting | Resilience, socio-emotional functioning, and depressive symptoms | Trained, certified teachers, or experienced leaders of well-being groups |

Note. *RCT* refers to randomized controlled trial, *CRCT* refers to cluster randomized controlled trial, *N* is the total number of participants at the post test, *T, C* means number of treatment participants and number of control participants respectively, *U* means universal and *T* means targeted, CBT Cognitive Behavior Therapy, PTSD Posttraumatic stress disorder.
The effect sizes for different grade levels also varied significantly by delivery personnel ($\beta = 0.78$, $p = 0.049$). At secondary school level, the effects of interventions delivered by clinicians have a larger mean effect size of 0.81 ($p = 0.012$) compared to interventions delivered by teachers with a mean effect size of 0.03 ($p = 0.711$). At elementary school level, no difference in effect size was identified by delivery personnel. Effect sizes for teacher-delivered interventions were similar across grade levels, with an average effect size of 0.06 ($p = 0.316$) at the elementary school level and an average effect size of 0.03 ($p = 0.711$) at the secondary school levels.

Moreover, significant differential effects were identified for outcome types by delivery personnel ($\beta = 0.89$, $p = 0.017$). Interventions with outcomes on anxiety delivered by clinicians have a significantly larger effect size ($ES = 0.86$, $p = 0.004$) compared to those delivered by teachers ($ES = 0.03$, $p = 0.657$). Those with outcomes on depression have a different trend. Teacher-delivered interventions on depression have a somewhat higher effect size ($ES = 0.07$, $p = 0.368$) than those of clinician-delivered interventions on depression ($ES = 0.01$, $p = 0.980$).

No significant differential effects were found for intervention outcomes (depression vs. anxiety) by intervention design ($\beta = -0.39$, $p = 0.219$).

**Exploratory and Sensitivity Analyses**

In exploratory analysis, an additional moderator called waitlist is included in the meta-regression model. The waitlist moderator is coded as 1 when the control group adopts business as usual or waitlist, which are considered as inactive group. It is coded as 0 when the control group has an active intervention serving as a placebo as compared to the real treatment. After adding this moderator, estimates and significance values were broadly similar to the original results.

Two types of sensitivity analyses were performed on the two moderators with arbitrary cut-off points: program duration and sample size. In the first round of sensitivity analysis, duration and sample sizes were coded as continuous variables. In the second round, they were coded as categorical variables with three levels using 33% and 66% percentile values as

---

### Table 4 Descriptive Statistics of Included Studies

| Category          | Level | Overall |
|-------------------|-------|---------|
| Study Level       |       |         |
| Total Studies     | 32    |         |
| Country           |       |         |
| Australia         | 13 (40.6) |     |
| Canada            | 2 (6.2) |     |
| Finland           | 1 (3.1) |     |
| Netherlands       | 2 (6.2) |     |
| New Zealand       | 1 (3.1) |     |
| Norway            | 2 (6.2) |     |
| UK                | 2 (6.2) |     |
| USA               | 9 (28.1) |     |
| Universal         |       |         |
| No                | 7 (21.9) |     |
| Yes               | 25 (78.1) |    |
| Elementary        |       |         |
| No                | 17 (53.1) |    |
| Yes               | 15 (46.9) |    |
| Long Duration     |       |         |
| No                | 22 (68.8) |     |
| Yes               | 10 (31.2) |     |
| Design            |       |         |
| CRCT              | 22 (68.8) |    |
| RCT               | 10 (31.2) |    |
| CBT               |       |         |
| No                | 10 (31.2) |    |
| Yes               | 22 (68.8) |    |
| Small Sample      |       |         |
| No                | 15 (46.9) |     |
| Yes               | 17 (53.1) |     |
| Outcome Level     |       |         |
| Total Effect Sizes| 79    |         |
| Teacher           |       |         |
| No                | 39 (49.4) |    |
| Yes               | 40 (50.6) |    |
| Construct         |       |         |
| Anxiety           | 40 (50.6) |    |
| Depression        | 39 (49.4) |    |

**Note.** Long duration refers to interventions lasting at least 12 weeks. Design: **RCT** refers to randomized controlled trial, **CRCT** refers to cluster randomized controlled trial. Small sample refers to sample sizes less than 250. **CBT** Cognitive Behavior Therapy

---

### Table 5 Overall Results for Meta-regression

| Coefficient | beta  | SE   | t    | df  | $P$ value |
|-------------|-------|------|------|-----|-----------|
| Null Model  |       |      |      |     |           |
| Intercept   | 0.28  | 0.13 | 2.20 | 29.41| 0.036     |
| Meta-Regression |     |      |      |     |           |
| Intercept   | 0.24  | 0.06 | 3.95 | 11.87| 0.002     |
| Long duration | -0.14| 0.13| -1.06| 10.90| 0.310     |
| Small sample | 0.21  | 0.18 | 1.16 | 11.22| 0.271     |
| Universal   | -0.24 | 0.18 | -1.32| 10.57| 0.215     |
| Depression  | -0.40 | 0.16 | -2.50| 14.71| 0.025     |
| Teacher     | -0.39 | 0.13 | -3.00| 10.32| 0.013     |
| CBT         | 0.48  | 0.16 | 2.91 | 9.54 | 0.016     |
| Elementary  | -0.36 | 0.19 | -1.89| 16.75| 0.076     |
| Teacher × CBT | -0.75| 0.34| -2.20| 7.89 | 0.059     |
| Universal × Teacher | 0.83 | 0.32| 2.60 | 11.30| 0.024     |
| Teacher × Elementary | 0.78| 0.37| 2.13 | 15.88| 0.049     |
| Depression × Teacher | 0.89 | 0.33| 2.69 | 14.44| 0.017     |
| Depression × CBT | -0.39| 0.28| -1.36| 6.53 | 0.219     |

**Note.** SE standard error, df degrees of freedom, **CBT** Cognitive Behavior Therapy
cut-off points. In both sensitivity analyses, estimates and significance values were broadly similar to the original results. Results of the exploratory and sensitivity analyses are available from the authors upon request.

**Publication Bias**

Applying the weight-function model, this study found significant publication bias. In the adjusted model, the test for heterogeneity is significant ($Q[df = 92] = 1485, p < 0.001$). Likelihood ratio test for the model result is significant ($x^2 = 81.88, p < 0.001$). This means that the estimated pooled effect of school-based programs is upwardly adjusted, suggesting that statistically significant positive effects were 38.90 times more likely to be reported than non-significant results. Readers should exercise caution when interpreting the results since there is a good chance that the overall effect is overestimated.

**Risk of Bias Analysis**

Since this meta-analysis applied stringent inclusion criteria, all the studies fulfill the JBI Checklist’s criteria except the sixth criteria, therefore only sixth criteria were coded: Were...
outcomes assessors blind to treatment assignment? The risk implied in the sixth criteria is that when assessors of the targeted outcomes are aware of participants’ allocation to the treatment or control, there may exist measurement errors. Among the 29 included studies, seven are blinded, 16 are not blinded, and six are unclear. Table 7 presents the risk of bias analysis. Overall, there is a low risk of bias among the included studies.

Discussion

There is an urgent need for schools to provide more mental health services and support; however, existing meta-analyses are insufficient in providing evidence-based insights on high-quality interventions. School-based mental health interventions are promising tools to protect school-aged children, so there is a pressing need to identify and disseminate evidence-based models to address the increasing number of children with depression and anxiety in schools. This meta-analysis aims to identify elements of effective school-based mental health interventions targeting depression and anxiety for K-12 students. This meta-analysis only included RCTs and used more stringent inclusion criteria (e.g., baseline equivalence, no significant differential attrition) and additional moderators (e.g., sample size and program duration) than previous reviews. The results indicate that, overall, compared to control groups, there was a significant positive mean effect of school-based interventions on symptoms of depression and anxiety. However, this result may be somewhat inflated due to publication bias, which was found to be likely in this study.

The overall findings disagree with a previous systematic review and meta-analysis that found a lack of evidence of the effectiveness of school-based interventions focusing on depression or anxiety (Caldwell et al., 2019). One reason to explain this disparity may lie in the method: the present study used a random-effects model looking across studies whereas the previous study used a network meta-analysis approach that is better suited to comparing the relative effectiveness of different interventions. In addition, the difference in inclusion criteria result in very different samples of studies analyzed in each study.

Apart from main findings, the moderator analyses help us gain a better understanding of the characteristics of effective depression- or anxiety-focused interventions. One interesting finding is that interventions focused on anxiety are more effective than those on depression for the K-12 population. Furthermore, the results find that CBT programs were more common and had significantly higher effect sizes than programs of other types. This supports the existing wide-usage of CBT programs (Werner-Seidler et al., 2017) and confirms previous research establishing CBT as an essential component in depression and anxiety reduction for school-aged children (Rooney et al., 2013). For intervention delivery, this study find that teacher-delivered programs had a lower mean effect than clinician-delivered programs, such that while clinician-delivered programs had significant, positive impacts, teacher-delivered programs had null effects. This finding is consistent with one previous meta-analysis on this topic, which found that programs delivered by non-school personnel are more effective (Werner-Seidler et al., 2021). Grade level was a marginally significant moderator, with interventions being more effective on the secondary-school population compared to the elementary-school population. No significant impacts were found in other moderators: program type (universal vs. targeted), program duration, and sample size. The result on program type disagrees with a previous meta-analysis that suggests targeted programs being more effective compared to universal programs (Werner-Seidler et al., 2021). The null finding from program type may be a result of offsetting effects between targeted programs’ pros and cons. Targeted programs’ screening and selection process may embarrass students, which further exacerbates their depression and anxiety. The positive effects brought by targeted programs may be nullified by the negative consequences of labeling and stigmatizing students (Huggins et al., 2008). This supports the notion that both targeted and universal program types are useful, and more benefits can be reaped if suitable programs are matched with suitable delivery personnel and treatment populations.

Moreover, the moderator analyses indicate that study-related factors are not significant moderators of impacts. While short-duration interventions had a significantly positive mean effect while long-duration interventions had no significant effect, this difference was not significant. Similarly, small sample sizes were associated with significantly positive effect sizes, while large sample sizes had no significant effects, yet the difference between these was not significant. While intervention duration and sample size may not be significant moderators, another possible explanation is that intervention duration and sample size do not have an impact on the intervention results when using the arbitrary cutoff in the analyses (intervention duration: 12 weeks; sample size: 250).

Interesting findings come from interaction analyses related to delivery personnel, an additional moderator many previous meta-analyses failed to include. The results demonstrated differential effects for delivery personnel by four pairs of interacting moderators. First, the effect sizes vary for personnel by intervention design. Clinicians had higher effect sizes with CBT programs than with non-CBT programs. This suggests that CBT programs have better effects when they are delivered by personnel with psychological backgrounds and professional training, while
### Table 7 Risk of Bias Analysis

| Study                  | Criteria 1 | Criteria 2 | Criteria 3 | Criteria 4 | Criteria 5 | Criteria 6 | Criteria 7 | Criteria 8 | Criteria 9 | Criteria 10 | Criteria 11 | Criteria 12 | Criteria 13 |
|------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|-------------|-------------|-------------|
| Barnes et al. (2012)   | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Barrett and Turner (2001) | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Britton et al. (2014)  | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Chaplin et al. (2006)  | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Collins et al. (2014)  | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| DeRosier (2004)        | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Haugland et al. (2020) | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Kraag et al. (2009)    | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Lewis et al. (2013)    | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Lowry-Webster et al. (2001) | Y          | N          | Y          | U          | N          | N          | U          | Y          | Y          | Y           | Y           | Y           | Y           |
| Merry et al. (2004)    | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Mifsud and Rapee (2005) | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Miller et al. (2011)   | Y          | N          | Y          | U          | N          | N          | U          | Y          | Y          | Y           | Y           | Y           | Y           |
| O’Kearney et al. (2009)| Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Olive et al. (2019)    | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Quach et al. (2016)    | Y          | N          | Y          | U          | N          | U          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Roberts et al. (2003)  | Y          | N          | Y          | U          | N          | N          | Y          | Y          | Y          | Y           | Y           | Y           | Y           |
| Roberts et al. (2010)  | Y          | N          | Y          | U          | N          | N          | Y          | Y          | Y          | Y           | Y           | Y           | Y           |
| Rooney et al. (2006)   | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Rooney et al. (2013)   | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Sawyer et al. (2010)   | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Sheffield et al. (2006)| Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Shochet et al. (2001)  | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Sinclair (2016)        | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Skryabina et al. (2016)| Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Spence et al. (2005)   | Y          | N          | Y          | U          | N          | U          | Y          | Y          | Y          | Y           | Y           | Y           | Y           |
| Stein et al. (2003)    | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Tak et al. (2016)      | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |
| Volanen et al. (2020)  | Y          | N          | Y          | U          | N          | N          | N          | Y          | Y          | Y           | Y           | Y           | Y           |

**Note.** Criteria 1. Was true randomization used for assignment of participants to treatment groups? Criteria 2. Was allocation to treatment groups concealed? Criteria 3. Were treatment groups similar at the baseline? Criteria 4. Were participants blind to treatment assignment? Criteria 5. Were those delivering treatment blind to treatment assignment? Criteria 6. Were outcomes assessors blind to treatment assignment? Criteria 7. Were treatment groups treated identically other than the intervention of interest? Criteria 8. Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed? Criteria 9. Were participants analyzed in the groups to which they were randomized? Criteria 10. Were outcomes measured in the same way for treatment groups? Criteria 11. Were outcomes measured in a reliable way? Criteria 12. Was appropriate statistical analysis used? Criteria 13. Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

_Y Yes, N No, U Unclear_
non-CBT programs have null effects no matter who they are delivered by. This makes sense because delivering CBT programs needs rigorous training and experience working in mental health fields.

Second, the effects of interventions vary for personnel by intervention type. Clinician-delivered programs had higher effect sizes for the targeted population than for the universal population (though both were significant and beneficial); while teacher-delivered programs showed the opposite pattern, although both were non-significant. Clinicians tend to have better effects when delivering programs to the targeted population. In targeted approaches, children demonstrate symptoms of depression or anxiety. In such cases, clinicians have more expertise in treating these symptoms compared to classroom teachers, so clinicians are likely to use more professional strategies when delivering programs to the targeted population.

Third, the effect sizes of interventions also vary for personnel by grade level. Compared to teachers, clinicians had a noticeably larger mean effect size at the secondary school level. Teachers had similar (non-significant) effect sizes at the elementary school level and at the secondary school level. One explanation for this may be that because adolescents experience a more severe level of mental health impairment than children (Olfson et al., 2015), adolescents may be more receptive to CBT interventions delivered by clinicians. It may also be that older students are more able to engage with and understand the content of the intervention, due to their more mature cognitive processes (Stice et al., 2009). This also fits with recommendations to target these symptoms early (ages 11–15), before those behaviors and beliefs become ingrained (Gladstone et al., 2011).

Fourth, the effects of interventions vary for personnel by intervention outcomes. For clinicians, they had larger impacts on average on anxiety outcomes, with null effects on depression outcomes, while for teachers, they had null effects on both depression and anxiety outcomes.

Limitations and Strengths

Readers may want to take note of several limitations when interpreting the results. First, this study only focuses on depression and anxiety outcomes. In the process of coding, the authors found some other interesting outcomes worth investigating, such as internalizing problems, stress, externalizing problems (Fung et al., 2019), worry (Skryabina et al., 2016), depression literacy (O’Kearney et al., 2009), stigma, and help-seeking tendencies (Link et al., 2020). The limited number of studies investigating these outcomes restricts this study’s ability to perform meaningful meta-regression analysis. Future studies can focus on these commonly under-researched outcomes of interest. Second, the authors were not able to extract information on socio-economic status (SES) for all studies, which could be a valuable moderator to the outcomes. SES may have impacts on the prevalence of mental health issues, access to treatment and support, the delivery personnel of mental health interventions. SES can be reflected in free and reduced lunch eligibility, household median income, single or double parenting etc. Future studies can examine the moderator effect of SES on depression and anxiety. Third, two of the twelve inclusion criteria used arbitrary cut-off points, which may have led to the exclusion of quality articles. However, during the screening process, only 10 studies were excluded because they have a sample size smaller than 60 and six studies were excluded because the program duration is less than five weeks. In total, this means that 7.33% of the studies were excluded in the full-text review stage due to these arbitrary cut-offs, so the results are likely not changed substantially by their exclusion. Fourth, publication bias was identified in the sample. That likely means there are many studies that have not been made public, and those studies likely have null effects. The average effect sizes reported here are likely larger than what would be seen if all studies were reported.

Despite these limitations, this study has a number of strengths and makes significant contributions to existing literature. Unlike previous meta-analyses on this topic, this article adds program duration, sample size, delivery personnel, and grade level as moderating factors. Understanding the effects of these additional moderators is important for making evidence-based decisions as well as designing future interventions. The findings related to the heterogeneity of delivery personnel and its interactions with other program features have implications for policymakers and practitioners.

Recommendations for Future Research

Future meta-analyses may benefit from doing further moderator and interaction analysis in this vein. Moreover, this study applied more stringent inclusion criteria to identify high-quality primary studies. The quality of systematic reviews depends on the quality of primary studies. Many previous meta-analyses acknowledge inconsistencies in primary studies’ quality as one limitation (e.g., Gerlinger et al., 2021). This study only included RCTs that meet certain standards, such as having equivalence in baseline conditions, low attrition rates, etc. However, the field would be stronger with more rigorous studies, that report more details about the population, such as SES data, as well as better descriptions of the counterfactual. Additionally, there is a dearth of effective programs at the elementary level as well as those delivered by teachers. Future work should explore if there are ways for teachers to deliver effective mental health programming or whether it can be implemented with younger students as a preventative approach.
Implications for Policy

This meta-analysis provides updated evidence and has practical implications for policymakers. While child and adolescent mental health has always been a concern, the need for services for school-age children is even greater, as mental health needs have increased as a result of the COVID-19 pandemic (Pfefferbaum, 2021). Given limited resources, investment in students’ mental health and overall well-being often faces competition from investment in academic achievement given the intense pressure to enhance performance (Zhang & Storey, 2022). It is essential that any efforts to provide school-based mental health interventions using this limited funding prioritize evidence-based interventions.

The current study highlights that depression- and anxiety-focused school-based interventions are more effective when delivered by professionals, such as certified clinicians or psychologists, compared to classroom teachers or school-health staff. In reality, this need is unmet by many schools due to a severe shortage of clinicians countrywide, not to mention pediatric clinicians (Elias, 2021). Policymakers should consider implementing more creative incentives, such as reducing costs to obtain certifications or enabling smoother transitions from associate degrees to bachelor’s degrees, to build a pipeline for school-based mental health clinicians (Kirchner & Cuneo, 2022).

Although this meta-analysis concludes that teacher interventions are not effect, some school districts may be unable to employ enough clinicians to provide mental health supports to their students. In those cases, where schools must rely on teachers to deliver interventions, based on data from other meta-analyses and applying a caution approach, we recommend focusing the available mental health professionals on the targeted populations at the secondary level while having teachers provide universal interventions at the elementary level. Another policy takeaway is that innovative programs are not necessarily better than traditional programs. There is a lack of conclusions on new interventions’ (e.g., mindfulness, yoga, positive psychology) effectiveness, but ample evidence on CBT programs’ efficacy.

Conclusion

Past meta-analyses provide many insights, yet they fail to use stringent criteria and diverse moderators. In light of the increasing incidence rates of depression and anxiety among children, there is an imperative to provide proven services to school-aged children. To synthesize rigorous past studies, this meta-analysis only included RCTs and used more stringent inclusion criteria (e.g., baseline equivalence, no significant differential attrition) and additional moderators (e.g., sample size and program duration) than previous reviews. This meta-analysis found that, overall, school-based mental health interventions school settings are effective at reducing depression and anxiety. However, these impacts may be somewhat inflated due to likely publication bias. Moderator analysis shows that anxiety-focused interventions are more effective than depression-focused programs, cognitive behavioral therapy is more effective than other types of programs, clinicians are more effective than teachers, and programs focusing on secondary school adolescents are more effective than those focusing on elementary school children. For practitioners, useful implications from this meta-analysis are that CBT is an effective program type and certified clinicians are more effective as compared to trained teachers in delivering mental health interventions, while efforts should focus on the secondary school level. More research is needed to explain the mechanisms behind effective elements in interventions and differences between children and adolescents. Although trained teachers are less costly and more widely available, this meta-analysis found that hiring professional psychologists or clinicians to deliver interventions produces more visible enhancements. At the same time, this field-specialization can also reduce teachers’ burdens and help them focus on teaching. In the post-pandemic era, identifying evidence-based mental health interventions can help us better prevent the onset of depressive and anxiety symptoms from an early stage.

Acknowledgements The authors want to thank Akash Pallath for providing constructive suggestions, Yiting Xiong for helping on locating journals for handsearching, and Aishwarya Jayabharathi for proofreading the final draft.

Authors’ Contributions Q.Z. conceived of the study, participated in its design and coordination, and drafted the manuscript; J.W. participated in the interpretation of the data and helped to draft the manuscript; A.N. participated in the design of the study and performed part of the statistical analysis. All authors read and approved the final manuscript.

Data Sharing and Declaration The datasets generated and/or analyzed during the current study are available in the github repository: https://github.com/qiyangzh/School-based-Mental-Health-Interventions-Targeting-Depression-or-Anxiety-A-Meta-analysis

Compliance with Ethical Standards

Conflict of Interest The authors declare no competing interests.

References

(References marked with an asterisk indicate studies included in the meta-analysis)

Ahlen, J., Breitholtz, E., Barrett, P. M., & Gallegos, J. (2012). School-based prevention of anxiety and depression: A pilot study in
Sweden. *Advances in School Mental Health Promotion*, 5(4), 246–257. https://doi.org/10.1080/1754730X.2012.730352.

Ahlen, J., Lenhard, F., & Ghaderi, A. (2015). Universal prevention for anxiety and depressive symptoms in children: a meta-analysis of randomized and cluster-randomized trials. *J Prim Prev*, 36, 387–403. https://doi.org/10.1007/s10935-015-0405-4.

American Psychological Association. (2017). What is cognitive behavioral therapy? Clinical Practice Guideline. https://www.apa.org/psd-guideline/patients-and-families/cognitive-behavioral

Arctic, A., & Erdogan, S. (2017). The effectiveness of the COPE healthy lifestyles TEEN program: A school-based intervention in middle school adolescents with 12-month follow-up. *Journal of Advanced Nursing*, 73(6), 1377–1389. https://doi.org/10.1111/jan.13217.

*Barnes, V. A., Johnson, M. H., Williams, R. B., & Williams, V. P. (2012). Impact of Williams LifeSkills* training on anger, anxiety, and ambulatory blood pressure in adolescents. *Translational Behavioral Medicine*, 2(4), 401–410. https://doi.org/10.1007/s13142-012-0162-3.

*Barrett, P., & Turner, C. (2001). Prevention of anxiety symptoms in primary school children: Preliminary results from anuniversal school-based trial. *The British Journal of Clinical Psychology*, 40(Pt 4), 399–410. https://doi.org/10.1348/014466501163887.

Becker, B. J. (2005). Failsafe N or file-drawer number. In Rothstein, H. R. et al. (Eds.), *Publication Bias in Meta-Analysis* (pp. 111–125). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470870168.ch7

Biddle, L., Gunnall, D., Donovan, J., & Sharp, D. (2006). Young adults’ reluctance to seek help and use medications for mental distress. *Journal of Epidemiology and Community Health*, 60(5), 426 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2563980/.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 172–177.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. https://doi.org/10.1002/ress.12.

*Britton, W. B., Lepp, N. E., Niles, H. F., Rocha, T., Fisher, N. E., & Gold, J. S. (2014). A randomizedcontrolled pilot trial of classroom-based mindfulness meditation compared to anactive control condition in sixth-grade children. *Journal of School Psychology*, 52(3), 263–278. https://doi.org/10.1016/j.jsp.2014.03.002.

Burckhardt, R., Manicavasagar, V., Batterham, P. J., & Hadzi-Pavlovic, D. (2016). A randomized controlled trial of strong minds: A school-based mental health program combining acceptance and commitment therapy and positive psychology. *Journal of School Psychology*, 57, 41–52. https://doi.org/10.1016/j.jsp.2016.05.008.

Cairns, K. E., Yap, M. B. H., Pilkington, P. D., & Jorm, A. F. (2014). Risk and protective factors for depression that adolescents can modify: A systematic review and meta-analysis of longitudinal studies. *Journal of Affective Disorders*, 169, 61–75. https://doi.org/10.1016/j.jad.2014.08.006.

Caldwell, D. M., Davies, S. R., Hetrick, S. E., Palmer, J. C., Caro, P., López-López, J. A., Gunnell, D., Kidger, J., Thomas, J., French, C., Stockings, E., Campbell, R., & Welton, N. J. (2019). School-based interventions to prevent anxiety and depression in children and young people: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 6(12), 1011–1020. https://doi.org/10.1016/S2215-0366(19)30403-1.

Centers for Disease Control and Prevention. (2022). *Data and Statistics on Children’s Mental Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/childrensmentalhealth/data.html

*Chaplin, T. M., Gillham, J. E., Reivich, K., Elkon, A. G. L., Samuels, B., Freeres, D. R., Winder, B., & Seligman, M. E. P. (2006). Depressionprevention for early adolescent girls: A pilot study of all girls versus co-educgros. *The Journal of Early Adolescence*, 26(1), 110–126. https://doi.org/10.1177/0727315405282655.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. https://doi.org/10.3102/0013189X16656615.

Coburn, K. M., & Vevea, J. L. (2019). Weightr: Estimating weightfunction models for publication bias (2.0.2) [Computer software]. https://CRAN.R-project.org/package=weightr

*Collins, S., Woolfson, L. M., & Durkin, K. (2014). Effects on coping skills and anxiety of a universal school-based mental health interventiondelivered in Scottish primary schools. *School Psychology International*, 35(1), 85–100. https://doi.org/10.1177/0143034312469157.

Conway, K. P., Compton, W., Stinson, F. S., & Grant, B. F. (2006). Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: Results from the national epidemiologic survey on alcohol and related conditions. *The Journal of Clinical Psychiatry*, 67(2), 247–257. https://doi.org/10.4088/jcp.v67n0211.

Copeland, W. E., Ailaie, I., Jonsson, U., & Shanahan, L. (2021). Associations of childhood and adolescent depression with adult psychiatric and functional outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, 60(5), 604–611. https://doi.org/10.1016/j.jaac.2020.07.097.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, 60(8), 837–844. https://doi.org/10.1001/archpsyc.60.8.837.

Covidence systematic review software. (2013). Veritas Health Innovation. www.covidence.org

Cronbach, L. J., Ambrón, S. R., Dornbusch, S. M., Hess, R. O., Hornik, R. C., & Phillips, D. C., et al. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: JosseyBass.

Daunic, A. P., Smith, S. W., Garvan, C. W., Barber, B. R., Becker, M. K., Peters, C. D., Taylor, G. G., Van Loan, C. L., Li, W., & Naranjo, A. H. (2012). Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the tools for getting along curriculum. *Journal of School Psychology*, 50(2), 149–166. https://doi.org/10.1016/j.jsp.2011.09.003.

De Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students’ academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509–545. https://doi.org/10.3102/0034654314540006.

Department of Homeland Security. (2022). *Kindergarten to Grade 12 Students*. Department of Homeland Security. https://studyinthestatus.dhs.gov/students/get-started/kindergarten-to-grade-12-students

*DeRosier, M. E. (2004). Building relationshipsand combating bullying: Effectiveness of a school-based social skills group-intervention. *Journal of Clinical Child & Adolescent Psychology*, 33(1), 196–201. https://doi.org/10.1027/0735742421003301_18.

Elias, P. (2021). Schools can help with the youth mental health crisis. Education Next. https://www.educationnext.org/schools-can-help-with-youth-mental-health-crisis-shorthages-counselors/
Behaviour Change, 18(1), 36–50. https://doi.org/10.1375/bech.18.1.36.

Mathews, R. R. S., Hall, W. D., Vos, T., Patton, G. C., & Degenhardt, L. (2011). What are the major drivers of prevalent disability burden in young Australians. The Medical Journal of Australia, 194(5), 232–235. https://doi.org/10.5694/j.1326-5377.2011.tb02951.x.

Merry, S., McDOWELL, H., Wild, C. J., Bir, J., & Cunliffe, R. (2005). Early Intervention for childhood anxiety and depression prevention program. Journal of the American Academy of Child & Adolescent Psychiatry, 44(10), 996–1004. https://doi.org/10.1097/00004583-200500000-00007.

Mifsud, C., & Rapee, R. M. (2005). Early Intervention for childhood anxiety in a school setting: Outcomes for anecdotally disadvantaged population. Journal of the American Academy of Child & Adolescent Psychiatry, 43(5), 538–547. https://doi.org/10.1097/00004583-200405000-00007.

Moses, T. (2010). Being treated differently: Stigma experiences with mental health disorders. Social & Psychological Studies, 5(7), 985–993. https://doi.org/10.1016/j.saps.2009.12.022.

Naikker, K., Galambos, N., Zeng, Y., Senthilselvan, A., & Colman, I. (2013). Social, demographic, and health outcomes in the 10 years following adolescent depression. Journal of Adolescent Health, 52(5), 533–538. https://doi.org/10.1016/j.jadohealth.2012.12.016.

Neitzel, A. J., Zhang, Q., & Slavin, R. (2022). Effects of Varying Inclusion Criteria: Two Case Studies. https://doi.org/10.35542/osf.io/h258x.

National Institute of Mental Health. (2018). Depression. National Institute of Mental Health (NIMH). https://www.nimh.nih.gov/health/topics/depression.

National Institute of Mental Health. (2022). Anxiety disorders. National Institute of Mental Health (NIMH). Retrieved June 15, 2022, from https://www.nimh.nih.gov/health/topics/anxiety-disorders.

Neil, A. L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. Clin Psychol Rev, 29, 208–215. https://doi.org/10.1016/j.cpr.2009.01.002.

Nock, M. K., Green, J. G., Hwang, I., McLaughlin, K. A., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2013). Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the national comorbidity survey replication adolescent supplement. JAMA Psychiatry, 70(3), 300 https://doi.org/10.1001/2013.jamapsychiatry.55.

O’Kearney, R., Kang, K., Christensen, H., & Griffiths, K. (2009). A controlled trial of a school-based Internet program for reducing depressive symptoms in adolescent girls. Depression and Anxiety, 26, 65–75. https://doi.org/10.1002/da.20507.

Olsson, M., Druss, B. G., & Marcus, S. C. (2015). Trends in mental health care among children and adolescents. New England Journal of Medicine, 372(21), 2029–2038. https://doi.org/10.1056/NEJMsai413512.

Olive, L. S., Byrne, D., Cunningham, R. B., Telford, R. M., & Telford, R. D. (2019). Can physical education improve the mental health of children? The look study cluster-randomized controlled trial. Journal of Educational Psychology, 111(7), 1331–1340. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=eric&AN=EJ1230839&site=ehost-live&scope=site&Auhtype=ip,shib&custid=s355202

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

Pallath, A. & Zhang, Q. (2022). Paperfetcher: A tool to automate handsearching and citation searching in systematic reviews. Research Synthesis Methods. https://arxiv.org/abs/2110.12490.

Park, S., Guz, S., Zhang, A., Beretvas, S. N., Franklin, C., & Kim, J. S. (2020). Characteristics of effective school-based, teacher-delivered mental health services for children. Research on Social Work Practice, 30(4), 422–432. https://doi.org/10.1177/1049731519879982.

Pfeiferbaum, B. (2021). Challenges for child mental health raised by school closure and home confinement during the COVID-19 pandemic. Current Psychiatry Reports, 23(10), 65 https://doi.org/10.1007/s11920-021-01279-z.

Quach, D., Jastrowski Mano, K. E., & Alexander, K. (2016). A randomized controlled trial examining the effect of mindfulness meditation on working memory capacity in adolescents. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine, 58(5), 489–496. https://doi.org/10.1016/j.jadohealth.2015.09.024.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.Rproject.org/.

Reavley, N. J., Cvetkovski, S., Jorm, A. F., & Lubman, D. I. (2010). Help-seeking for substance use, anxiety and affective disorders among young people: Results from the 2007 Australian national survey of mental health and wellbeing. Australian & New Zealand Journal of Psychiatry, 44(8), 729–735. https://doi.org/10.3109/00048671003705458.

Roberts, C. M., Kane, R., Bishop, B., Cross, D., Fenton, J., & Hart, B. (2010). The prevention of anxiety and depression in children from disadvantaged schools. Behaviour Research and Therapy, 48(1), 68–73. https://doi.org/10.1016/j.brat.2009.09.002.

Roberts, C. M., Kane, R., Thomson, H., Bishop, B., & Hart, B. (2003). The prevention of depressive symptoms in rural school children: A randomized controlled trial. Journal of Consulting and Clinical Psychology, 71(3), 622–628. https://doi.org/10.1037/0022-006X.71.3.622.

Rooney, R., Hassan, S., Kane, R., Roberts, C. M., & Nesa, M. (2013). Reducing depression in 9–10 year old children in low SES schools: A longitudinal universal randomized controlled trial. Behaviour Research and Therapy, 51(12), 845–854. https://doi.org/10.1016/j.brat.2013.09.005.

Rooney, R., Roberts, C., Kane, R., Pike, L., Winsor, A., White, J., & Brown, A. (2006). The prevention of depression in 8- to 9-year-old children: A pilot study. Australian Journal of Guidance and Counselling, 16(1), 76–90. https://doi.org/10.1375/ajgc.16.1.76.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. Health Services & Outcomes Research Methodology, 2, 169–188. https://doi.org/10.1016/S1534-180X(00)00007-9.

Saaywer, M. G., Pfeifer, S., Spence, S. H., Bond, L., Graetzel, B., Kay, D., Patton, G., & Sheffield, J. (2010). School-based prevention of depression: A randomised controlled study of the beyondblue schools research initiative. Journal of Child Psychology and Psychiatry, 51(2), 199–209. https://doi.org/10.1111/j.1469-7610.2009.02136.x.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). Interventions for adolescent struggling readers: A meta-analysis with implications for practice. Portsmouth, NH: Center on Instruction, RMC Research.
Qiyang Zhang is a doctoral student in School of Education at Johns Hopkins University. Her major research interests include adolescents’ mental health and well-being, perceived social support, and systematic reviews.

Jun Wang is a doctoral student in School of Education at Johns Hopkins University. Her major research interests include early childhood children’s well-being, social-emotional learning, early bilingualism, and systematic reviews.

Amanda Neitzel is the assistant research professor and deputy director of evidence research in the School of Education at Johns Hopkins University. Her major research interests include systematic reviews, evidence synthesis, and school-based health.