School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis

Aliza Werner-Seidler PhD, Yael Perry PhD, Alison L. Calear PhD, Jill M. Newby PhD, Helen Christensen PhD

PII: S0272-7358(15)30140-9
DOI: doi: 10.1016/j.cpr.2016.10.005
Reference: CPR 1556

To appear in: Clinical Psychology Review

Received date: 13 October 2015
Revised date: 21 October 2016
Accepted date: 21 October 2016

Please cite this article as: Werner-Seidler, A., Perry, Y., Calear, A.L., Newby, J.M. & Christensen, H., School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis, Clinical Psychology Review (2016), doi: 10.1016/j.cpr.2016.10.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis

Aliza Werner-Seidler, PhD
Yael Perry, PhD
Alison L. Calear, PhD
Jill M. Newby, PhD
Helen Christensen, PhD

1 Black Dog Institute, University of New South Wales, Sydney, NSW, Australia
2 National Institute for Mental Health Research, The Australian National University, Canberra, ACT, Australia
3 Clinical Research Unit for Anxiety and Depression, School of Psychiatry, University of New South Wales, Sydney, NSW, Australia

Corresponding author: Aliza Werner-Seidler.
Address: Black Dog Institute
University of New South Wales
Hospital Rd, Randwick, NSW 2031
Australia.
Email: a.werner-seidler@blackdog.org.au.
Telephone: +61 2 9382 3769
Fax: +61 2 9382 8208
Abstract

Depression and anxiety often emerge for the first time during youth. The school environment provides an ideal context to deliver prevention programs, with potential to offset the trajectory towards disorder. The aim of this review was to provide a comprehensive evaluation of randomised-controlled trials of psychological programs, designed to prevent depression and/or anxiety in children and adolescents delivered in school settings. Medline, PsycINFO and the Cochrane Library were systematically searched for articles published until February, 2015. Eighty-one unique studies comprising 31,794 school students met inclusion criteria. Small effect sizes for both depression (g=.23) and anxiety (g=.20) prevention programs immediately post-intervention were detected. Small effects were evident after 12-month follow-up for both depression (g=.11) and anxiety (g=.13). Overall, the quality of the included studies was poor, and heterogeneity was moderate. Subgroup analyses suggested that universal depression prevention programs had smaller effect sizes at post-test relative to targeted programs. For anxiety, effect sizes were comparable for universal and targeted programs. There was some evidence that externally-delivered interventions were superior to those delivered by school staff for depression, but not anxiety. Meta-regression confirmed that targeted programs predicted larger effect sizes for the prevention of depression. These results suggest that the refinement of school-based prevention programs have the potential to reduce mental health burden and advance public health outcomes.

Systematic Review Registration Number: PROSPERO - CRD42015023328

Keywords: Meta-analysis; systematic-review; school-based; depression; anxiety
Anxiety and depression are common, debilitating mental health problems that often emerge for the first time during adolescence. Up to 20% of young people will experience a depressive episode or an anxiety disorder by the age of 18 years (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Lewinsohn, Rohde, & Seely, 1998; Merry et al., 2011). Both depression and anxiety disorders tend to run a chronic and recurring course, with comorbidity levels of between 10-50% (Garber & Weersing, 2010; Kessler, Ayenewoli, & Merikangas, 2001; Scholten et al., 2013). This poses a significant public health burden, with depression already the leading cause of disease burden in Australia (Murray et al., 2015).

Earlier onset of depression and anxiety is associated with a worse clinical course over the lifespan, and in youth is associated with drug and alcohol abuse, risky sexual behaviour, suicide risk, poor academic outcomes and physical health problems (Birmaher et al., 1996; Donovan & Spence, 2000; Kessler et al., 2001; Rao et al., 1995). One way in which to address this disease burden is via prevention. Prevention programs can be divided into universal or targeted approaches. Universal prevention is delivered to all individuals within an identified population regardless of risk. For example, universal prevention programs for youth are typically delivered on a large scale in the school environment to every child in the grade (e.g., Neil & Christensen, 2009). Conversely, targeted prevention approaches are directed either towards those who have an increased risk profile for the disorder such as familial risk or poverty (selective prevention), or who have sub-clinical symptoms (indicated prevention; Mrazek & Haggerty, 1994; Muñoz, Cuijpers, Smit, Barrera, & Leykin, 2010).

Prevention programs are associated with a number of advantages. First, prevention, in some cases, will prevent the incidence of a disorder occurring altogether, with research suggesting that it is possible to prevent 22% of new depression cases each year (Cuijpers, van Straten, Smit, Mihalopoulos, & Beekman, 2008). In cases where disorder cannot be prevented, prevention programs can delay the onset of clinically significant symptoms.
(Bienvenu & Ginsburg, 2007; Merry et al., 2011), which confers considerable benefit, including reduced disability and service use (Donovan & Spence, 2000; Muñoz et al., 2010). Many young people who have mental health disorders are unable to, or do not access mental health treatment services (Gulliver, Griffiths, & Christensen, 2010; Gulliver, Griffiths, Christensen, & Brewer, 2012). Taking a prevention approach means that this problem may be averted altogether. Second, implementing prevention programs at an early age when behaviour is more amenable to change is likely to produce better outcomes than treatment delivered when rigid patterns of cognition and behaviour have already been established and are engrained (Craske & Zucker, 2001; Gladstone, Beardslee, & O'Connor, 2011). Indeed, the failure to respond to treatment is often a consequence of established patterns of behaviour that are difficult to reverse (Donovan & Spence, 2000).

Utilising the school system as the context for implementation of prevention programs provides a natural and accessible way to reach young people. The school environment is particularly advantageous because it offers unparalleled access to youth (Masia-Warner, Nangle, & Hansen, 2006). Schools are a place of learning and provide tremendous opportunity to provide young people with many of the skills and strategies that may protect against, or delay, the onset of emotional difficulties. School-based programs integrated into the school curriculum can also alleviate many typical barriers to accessing treatment, such as time, location and cost (Barrett & Pahl, 2006).

While there has been an increase in studies examining preventive programs for youth over the past decade, the field has been hampered by a lack of consistent terminology, poor research methods and inadequately powered trials, leading to confusion about whether prevention programs are effective and should be pursued (e.g., Callear & Christensen, 2010; Nehmy & Wade, 2014). On the whole, the evidence suggests that there is a modest but positive effect of prevention programs for depression and anxiety (Merry et al., 2011;
Stockings et al., 2016; van Zoonen et al., 2014), particularly those based on Cognitive
Behaviour Therapy (CBT; Hetrick, Cox, & Merry, 2015). Larger effects sizes have generally
been detected for targeted programs, relative to those delivered universally (Calear &
Christensen, 2010; Merry et al., 2011), although most universal trials are likely to have been
underpowered (Muñoz et al., 2010). There has also been some indication that potentially
larger effect sizes are associated with programs delivered by mental-health professionals,
relative to programs delivered by teachers (e.g., Calear & Christensen, 2010; Stallard et al.,
2014). Moreover, online prevention programs also require consideration in light of advancing
technology and burgeoning interest in the area, as well as the potential for online delivery to
reduce burden on teachers and professionals (Calear, Christensen, Mackinnon, Griffiths, &
O'Kearney, 2009). Furthermore, variability in the types of control conditions included in
previous studies have further complicated the picture, with effect sizes being potentially
smaller when an attentional control group serves as the comparison, relative to an inactive
comparison, such as a waitlist control (e.g., Calear & Christensen, 2010). Rigorous evaluation
of the impact of control group type on effect size is now needed. Finally, there is not yet
consensus regarding the optimal temporal window within which to intervene. Given that
anxiety typically precedes the onset of depression in youth (Kessler et al., 2005; Merikangas,
Nakamura, & Kessler, 2009), earlier prevention might impact anxiety and depressive
symptoms differentially.

To date, there have been several systematic reviews and meta-analyses focused on
mental health prevention programs in young people. However, most of these reviews do not
focus exclusively on the school environment (e.g., Fisak, Richard, & Mann, 2011; Stockings
et al., 2016), and there has been a bias towards evaluating depression-prevention studies over
anxiety (Horowitz & Garber, 2006; van Zoonen et al., 2014). Most notably, there was a
comprehensive Cochrane review of youth-focused depression prevention programs (Merry et
al., 2011), which updated an earlier review conducted more than ten years ago (Merry, McDowell, Hetrick, Bir, & Muller, 2004). The findings of these two reviews reported a modest but genuine prevention effect on symptoms outcomes ($g=.26$ in the 2004 review, $g=.20$ in the 2011 review). However, together with those reviews listed above, these did not separate school-based from community-delivered programs, which is problematic because the circumstances of young people in schools are likely to be different from those targeted in community settings (e.g., youth offenders, special needs groups). Two previous systematic reviews that have specifically examined school-based prevention programs for anxiety and depression (Calear & Christensen, 2010; Neil & Christensen, 2009), found promising effects (depression: median $d=0.64$ for children, $d=0.60$ for adolescents; anxiety: median $d=0.57$ for children, $d=0.32$ for adolescents). A comprehensive review is now needed, particularly in light of the advances and substantially increased interest in school-based prevention over recent years.

**Aims**

The aim of the current review was therefore to provide a comprehensive overview and evaluation of randomised-controlled trials of school-based prevention programs for depression and/or anxiety in children and adolescents, and to conduct a meta-analysis of intervention effects. More specifically, the review aims to: (1) identify the overall effect of school-based depression and/or anxiety prevention programs on symptoms of depression and anxiety; (2) identify the time period over which treatment gains are maintained; (3) establish the relative effect of these programs according to (i) control group type (i.e., no intervention (NI) vs. waitlist control (WL) vs. attention control (AC)), (ii) prevention type (i.e., universal vs. targeted), (iii) delivery format (i.e., classroom teacher vs. health professional), (iv) program content (i.e., CBT vs other), and, (v) age of target sample (i.e., children vs. early
adolescence vs. late adolescence); and (4) determine, where possible, fidelity of program
delivery, program attendance and completion rates, and the cost-effectiveness of these
interventions.

Method

Protocol and registration

In line with Preferred Reporting Items for Systematic Reviews and Meta-Analyses
(PRISMA) recommendations (Moher, Liberati, Tetzlaff, & Altman, 2009), this review was
registered with PROSPERO [CRD42015023328].

Eligibility criteria

(i) Types of participants: Participants were children or adolescents with a mean age
between 5-19 years. Age was used to categorise participants according to whether they were
children (<10 years), early adolescents (10-14 years), or older adolescents (>14 years).
Diagnostic status was not used to include or exclude participants, as a majority of school-
based studies do not conduct a diagnostic assessment prior to program delivery.

(ii) Types of interventions: Included interventions were manualised psychological
or psycho-educational programs, including individual, group or computerised interventions
such as CBT (including relaxation and progressive muscle relaxation approaches),
interpersonal psychotherapy (IPT), mindfulness-based cognitive therapy (MBCT), wellbeing
therapy (WBT) and psycho-educational approaches. For multi-component programs, the
psychological or educational component was required to constitute >75% of the programs’
content. Studies were included if they used a program designed to prevent depressive or
anxiety symptoms, and/or promote wellbeing. Studies evaluating drug and alcohol, physical
activity, nutritional or pharmacological interventions were not eligible for inclusion.

Interventions needed to be school-based, which in this context refers to a program that
is endorsed by the school and delivered in the classroom during school hours, or before or
after school on school premises. The school context could not simply provide the location for private/external programs to be delivered. The program was required to be school-supported with recruitment occurring within and facilitated by the school. For multi-setting studies (e.g., partly at school, partly at a primary health care setting), the school-based component needed to comprise >75% of the overall program. There were no restrictions on whether participants were receiving other forms of therapy or medication.

(iii) Types of comparisons: Studies were included in which the effects of the school-based intervention was compared to either a no intervention control group or a school-as-usual control condition (NI), a waitlist control condition (WL), or an attention control condition/alternate educational/psychological condition (e.g., bibliotherapy; AC).

(iv) Types of outcomes: Studies were included if they reported symptoms of depression and/or anxiety at both baseline and post-intervention at a minimum. The two primary outcomes were depression and anxiety symptoms. Outcome measures needed to be valid and reliable rating scales suitable for children and adolescents. When more than one continuous measure was described, the primary outcome was used. If the primary outcome was not specified, the data from the measure reported first were extracted. For studies meeting inclusion criteria, means, standard deviations, and sample size of completers at post-intervention and at each follow-up time point thereafter, were extracted. In studies in which outcome data were not reported, AWS contacted the authors of the study to obtain this information.

(v) Length of follow-up: Reporting on follow-up outcomes was not required for inclusion in the current review. However, available follow-up data were extracted and categorised. Each data-point was categorised as being post-intervention, short-term (0–6

---

1 One study (Stallard et al, 2013) reported ‘post-intervention’ results at 12-month follow-up. This study was included, but the data were compared to other studies reporting medium term follow-up. Another study (Hunt et al., 2009) also reported ‘post-intervention’ data only at 24 month follow-up. Again this study was included but compared to other studies reporting long term follow-up.

2 Several authors did not provide the data requested and so these studies (n=4) could not be included in the meta-analysis. These are noted in Table 1.
months inclusive), medium-term (6-12 months inclusive), or long-term (greater than 12 month) follow-up. Waitlist control groups were required to remain waitlist for the follow-up assessment point for data to be extracted (e.g., waitlist groups could not be delivered the intervention prior to follow-up for follow-up data to be included). The follow-up data categories were based on the time points most frequently reported by authors. When there was more than one follow-up assessment during a particular time-frame in a particular study (e.g., 18 and 24 months), the follow-up period closest to other studies in that category was included.

(vi) Types of studies: Studies were eligible for inclusion if they used quantitative randomised controlled trial (RCT) methodology, including cluster RCTs. Studies were included if they were published in English language, peer-reviewed journals.

Search strategy

The electronic databases PsycINFO, PubMed and the Cochrane Library were systematically searched. The following terms were used as text words and key words: (depress* OR mood OR affect OR anxiety OR anxious) AND (prevent* OR early intervent*) AND (school* OR school-based OR adolescen* OR child* OR youth) AND control. The search was conducted on 12th of February, 2015. See Appendix I for search strategy used for PubMed. A separate electronic search (using the same terms and keywords) was also conducted in the peer-reviewed journal Internet Interventions, as it is not currently listed in these databases. An additional search in the peer-reviewed journal Early Intervention in Psychiatry was carried, as there is a delay between the online publication of articles in this journal and their inclusion into the databases. Finally, reference lists from studies meeting inclusion criteria, as well as recent reviews in the field, were hand-searched.

Data extraction process and management
Study characteristics and outcome measures were extracted by AWS, and independently checked by YP. The following data were extracted into a piloted spreadsheet: author, year of publication, program target (depression, anxiety, depression and anxiety), prevention type (universal, targeted), age range, sample size, program name, control group conditions (no intervention, waitlist, active), program format (face-to-face group, individual, computerised), program content (e.g., CBT, IPT), mode of delivery (school staff, external mental health professionals/researchers), and number of sessions. Parental involvement, fidelity and completion information were extracted when reported. Outcome data for depression and anxiety symptoms were extracted for the primary outcome analysis. Studies were distinguished according to whether they targeted depression, anxiety or both, as outlined in the aims and objectives of the study. For studies targeting either depression or anxiety, symptom data were only extracted for the target of the program (e.g., depressive symptoms for a depression-prevention program). If a program targeted both depression and anxiety, both outcomes were extracted. In cases where a study targeted either depression or anxiety but included both depression and anxiety symptoms as outcomes, symptom data were extracted from both for the purposes of the meta-analysis. This decision was made because it provides a more comprehensive evaluation of both anxiety and depression outcomes, which is warranted given the similarities between internalising disorder symptoms as well as the commonalities between prevention programs for these disorders.

**Risk of bias in individual studies**

Quality and risk of bias was assessed using the Cochrane Collaboration ‘Risk of Bias’ tool (Higgins & Green, 2011). This tool allows possible sources of bias to be assessed. In the current review, included studies were assessed against those criteria deemed to be most relevant to school-based prevention randomised controlled trials. As such, studies were assessed in relation to: a) generation of their condition allocation sequence, b) concealment of
this sequence, c) reporting of incomplete outcome data, d) selective reporting of data, and e) protection against contamination\(^3\). Quality ratings were made independently by AWS and either YP or JN, and disagreements were resolved through discussion and consultation with ALC. Cochrane recommends against using summed scales to make an overall judgement about level of bias because different forms of bias are likely to be more or less relevant depending on the nature of the research (Higgins & Green, 2011). Therefore, risk of bias for each criterion identified above has been reported individually, rather than in aggregated format (see Table 1).

**Statistical analyses**

**Calculation of effect sizes:** Comprehensive Meta-Analysis (version 3.0, Biostat Inc.) was used to calculate individual study and pooled effect sizes. For each comparison between a prevention intervention and control group, effect size was calculated using Hedges' g. This statistic is the standardised mean difference between the two groups at post-treatment, which includes an adjustment to address small sample sizes (Hedges & Olkin, 1985). The 95% confidence interval around effect size is also reported. In cases where studies had multiple comparison conditions, the number of participants in the prevention program group was divided equally over the comparison conditions so that each participant was only represented once in the meta-analysis. One study (Arnarson & Craighead, 2009, 2011) reported dichotomous outcomes only so this data were transformed into Hedges g using Comprehensive Meta-Analysis software. Effect sizes of 0.2, 0.5 and 0.8 refer to small, moderate and large effect sizes respectively (Cohen, 1988). As considerable heterogeneity among studies was expected (as is commonly reported in the field e.g., Merry et al., 2011), a random effects model was used, which assumes that the true effect size varies from one study

---

\(^3\) Contamination was included as it is relevant to school-based studies and refers to whether the unit of allocation was at the school level or not. When randomisation occurs at the individual or class level, there is potential for risk of contamination across conditions through sharing materials or information, and so a risk of bias is reported. Cluster RCTs (i.e., randomisation at the school level) protect against this source of bias (Craig et al., 2008).
to the next, and that the studies in the analysis represent a random sample of effect sizes that could have been observed.

**Testing homogeneity:** Homogeneity of effect sizes was tested using the $I^2$ statistic, which indicates heterogeneity in percentages. Zero percent indicates no heterogeneity, while 25%, 50% and 75% indicate low, moderate, and high heterogeneity, respectively.

**Subgroup analyses:** A number of subgroup analyses were planned, in which prevention type (targeted vs. universal prevention), personnel delivering the program (classroom teachers/counsellors vs. mental health professional/researcher), control condition type (no intervention vs. waitlist vs. active), age at which the intervention was delivered (childhood vs. early adolescence vs. late adolescence), and content of the programs delivered (CBT vs. other) were examined, using a mixed-effects model.

**Meta-regression:** We conducted a multivariate meta-regression with effect size as the dependant variable, using a mixed effects model. This enabled us to simultaneously examine the moderator variables we examined in the subgroup analyses as predictors of outcome.

**Testing for and managing publication bias:** The funnel plots of the primary outcome measures (depression, anxiety) were examined to test for publication bias (Egger, Smith, Schneider, & Minder, 1997). In cases where publication bias was indicated, Duval and Tweedie’s Trim and Fill procedure (Duval & Tweedie, 2000) was conducted within Comprehensive Meta-Analysis, which yields an adjusted effect size estimate. This procedure corrects for the variance of the effects and provides a best estimate of the unbiased effect size.

**Results**

**Study selection**
See Figure 1 for the PRISMA flowchart illustrating the inclusion of studies. A total of 5917 articles were identified, from which duplicate articles \((n=486)\) were removed. The remaining titles and abstracts \((n=5431)\) were screened by the first author to determine their relevance to this review. These were independently screened by a masters-level research assistant for relevance. Of the abstracts, 5212 were deemed irrelevant according to both raters, and therefore excluded. Two authors (AWS, YP)\(^4\) then independently screened the full text articles of the remaining 219 records, of which 129 were excluded as they did not meet inclusion criteria. Any disagreements were resolved through discussion and consultation with ALC. This resulted in 90 articles being included in the current review.

**Study characteristics**

Characteristics of included studies are presented in Table 1. There were 81 unique studies identified in the current review, which included a total of 31,794 participants \((n=16,848\) in prevention program conditions, and \(n=14,946\) in control conditions). Sample size of the included studies varied considerably from between 21 participants (Hains & Ellmann, 1994; Hains & Szyjakowski, 1990) to 2,512 participants (Araya et al., 2013), with a median of 208 participants. Of the 81 studies, 40 were studies of depression prevention \((n=15,844)\), 24 were anxiety prevention studies \((n=8,580)\), and 17 were mixed depression and anxiety studies \((n=7,370)\).

**Prevention type.** Overall, more than half of the studies were of a universal program (44 studies), just under one third were of an indicated program (25 studies), nine studies tested a selective prevention program, two studies evaluated a blended indicated/selective program, while the final study involved a blended universal/indicated program. Focusing on the 40 depression studies, 42.5% were universal, 40% indicated, 12.5% selective, 2.5% were blended selective/indicated and 2.5% universal/indicated. Of the 24 anxiety programs, 62.5%

\(^4\) A subset of full-text articles \((n = 89)\) were independently screened and selected by AW-S and a doctoral-level research assistant (screening and selection was completed for the other 130 articles by AW-S and YP).
were universal, 25% indicated, 8.25% selective, and 4.25% blended selective/indicated. Of the 17 mixed depression/anxiety studies, 71% were universal, 24% were indicated, and 5% selective. The selective interventions defined ‘risk’ as: a negative attributional style (Arnarson & Craighead, 2009), living in a low-income area (Cardemil, Reivich, & Seligman, 2002; Kindt, Kleinjan, Janssens, & Scholte, 2014), elevated anxiety sensitivity (Balle & Tortella-Feliu, 2010), conduct or behavioural problems (King & Kirschenbaum, 1990), personality risk factors (Castellanos & Conrod, 2006), exposure to community or political violence (Cooley-Strickland, Griffin, Darney, Otte, & Ko, 2011; Tol et al., 2008), or parental divorce (Pedro-Carroll & Cowen, 1985). For the meta-analysis, given the relatively few selective (n=9) and blended trials (n=3), these were combined with indicated trials and collectively labelled ‘targeted’ trials (n=37; 46%), resulting in a relatively even proportion of universal and targeted programs included within the review.

**Participant group.** Just over one-fifth (21%; n=17) of the studies identified in the current review delivered the intervention to children with a mean age of less than ten years. A majority (11 of 17) of these programs targeted anxiety only, four were mixed anxiety/depression, and only two studies focused exclusively on depression. Thirty-nine trials (48%) involved early adolescents, with the participant mean age between ten and 14 years. There was more variability in the mental health target amongst this age group, with 11 studies focusing on anxiety, seven mixed anxiety/depression, and 21 depression-only trials. The remaining 25 studies (31%) were delivered to individuals in the later adolescent years, with a mean age of between 14 and 19 years. Most programs directed toward older adolescents included a depression outcome measure, 16 exclusively, six with anxiety outcomes as well, and only three were exclusively anxiety focused. Relatively few studies (16%) screened participants prior to study entry with a diagnostic interview, and excluded participants based on the presence of a clinical disorder.
Randomisation. Studies varied in terms of whether randomisation occurred at the school (33% studies overall), grade (1%), class (16%) or individual level (49%). One study randomised condition by county (1%).

Program content. CBT comprised the basis of 84% of the programs identified in the current review. Other therapeutic approaches included five combined CBT/IPT programs (6%), two pure IPT programs (2.5%), two social skills programs (2.5%), one blended CBT and creative-expressive experiential therapy, one mindfulness-based cognitive therapy program, one wellbeing therapy program, and one psychoeducational program (each making up 1.2%).

Program format and mode of delivery. More than half \((n=51, 63\%)\) of the included programs were delivered by personnel external to the school environment, while 35\% \((n=28)\) were delivered by school staff (including two that were delivered by computer, but delivery was supported by school staff). Two studies involved both school and external staff, and two studies compared a school-led version of the program to a health-professional delivered version (Barrett & Turner, 2001; Stallard et al., 2014). Of the programs delivered by personnel external to the school, 35 (43\% overall) were delivered by mental health professionals/researchers, six (7\%) by graduate students, and ten (12\%) by a combination of the two. For programs delivered by school staff, five (6\% overall) of these involved both teachers and school health staff (counsellors or nurses), 18 (22\%) were delivered exclusively by classroom teachers, and five (6\%) were delivered exclusively by school health staff.

Interventions in all but four studies (95\%) were delivered in a group format, most frequently in small groups of 6-10 individuals. Two studies used a combination of group and individual sessions, while the two remaining programs were delivered individually online, although one of these involved a group discussion of the material following each module (Wong, Kady, Mewton, Sunderland, & Andrews, 2014).
Program sessions. The length of the programs identified ranged from two to 40 sessions, with most programs (70%) being delivered in between 8-12 sessions (median=10 sessions), usually on a weekly or fortnightly basis. Some programs delivered the intervention in two to seven sessions (16%), while others were administered in 13-24 sessions (13%). One study delivered the intervention weekly over the academic year (approximately 40 weeks). There was substantial variability in the duration over which these programs were delivered, ranging from between 3 to 40 weeks (median=10 weeks). A majority of sessions ran for between 45-60 minutes (53%; n=43; median=60 minutes), although some went for longer (between 75-120 minutes; 30%; n=23). The remaining studies (n=15) did not report session length. Only a small proportion (11%) of the programs involved the delivery of booster sessions, with seven studies offering two booster sessions, one program providing up to five booster sessions, and one study offering a single booster session.

Control groups. Forty of the included studies compared the prevention program to a NI control group (49%), 21 used a WL (26%), eight studies involved an AC only (10%), and 12 studies (15%) involved more than one comparison condition. Of these twelve studies involving more than one comparison, ten included an NI as well as one AC arm, and two involved a WL arm as well as one AC condition.

Outcome measures. Of the studies that evaluated depression symptoms as a primary outcome, the most frequent symptom measures included the Children’s Depression Inventory (CDI; 38%; Kovacs, 1992), followed by the Centre for Epidemiological Studies – Depression Scale (CES-D; 11%; Radloff, 1977), and the Reynolds Adolescent Depression Scale (RADS; 11%; Reynolds, 2010). Anxiety symptoms were measured most frequently with the Spence Children’s Anxiety Scale (SCAS; 24%; Spence, 1998), followed by the Multidimensional Anxiety Scale for Children (MASC; 21%; March, Parker, Sullivan, Stallings, & Conners,
1997), and the Revised Children’s Manifest Anxiety Scale (RCMAS; 12%; Reynolds & Richmond, 1978).

**Parental involvement.** Most programs (58%) did not involve parents in any way. In studies that did request parental involvement, parents were asked to complete questionnaires about themselves or their children (18%), or were offered information sessions so that they could learn about the program that their child was participating in (4%). However, the attendance level of these information sessions was either low (<32%), or not reported. Some studies (21%) involved parents as partial recipients of the intervention; with between two to seven sessions being provided to parents, in which they were provided psychoeducation about cognitive behavioural skills. Levels of uptake of parental sessions varied considerably, with most studies reporting at least some involvement from one parent in at least one session.

**Fidelity to program.** Two of the studies were delivered online, protecting the fidelity of the program (Calear et al., 2009; Wong et al., 2014). Of the remaining 79 studies, 33 (42%) did not report on fidelity. The remaining 58% of studies reported some level of program fidelity monitoring, ranging from program leaders completing self-report checklists (n=14), to independent ratings of session audio and visual recordings (n=19).

**Program completion rates.** A minority of the studies (36%) reported on program completion rates. These varied substantially across studies and the metric in which they were reported (e.g., number attending all sessions, mean number of sessions attended, number of sessions missed), although 22 studies (27%), reported high attendance and completion rates – with at least more than half of their participants completed more than half of the program.

**Cost-effectiveness.** Only one study evaluated cost-effectiveness (Stallard et al., 2013). It was reported that the targeted depression intervention was not cost effective, relative to regular Personal, Social, Health and Economic (PSHE) classes.
Risk of bias. The methodological quality of the studies reported varied significantly (see Table 1 for individual study quality ratings). There was evidence of selection bias, with only twenty-two studies (27%) reporting that the allocation sequence had been adequately generated. This means that 73% of studies either did not provide sufficient detail to evaluate how groups were randomised (54 studies), or did not use a randomisation procedure that ensured comparability between groups (e.g., did not use a random number table; five studies). Similarly, 29 studies (36%) indicated adequate concealment of allocation, suggesting that a majority of studies did not report enough information to ascertain whether intervention allocations could have been foreseen prior to, or during enrolment. Risk of attrition bias was lower, with thirty two studies indicating low risk of bias for addressing incomplete outcome data (40%), which indicates that the proportion of missing data was comparable across the study conditions. Bias from selective reporting was difficult to judge, with five studies clearly indicated low risk of bias (6%). Most studies (85%) did not report enough information in order to rule selective reporting out, which is largely contingent on publication of a study protocol to ensure authors report on outcomes they have previously indicated they will. Approximately a third of studies (36%) had adequate protection against contamination by randomising at the school level. This means that there was potential for contamination between conditions in 64% of studies, where participants in the control condition may have had access to material covered in the prevention condition, though contact with participants. Overall, only one study was classified as having a low risk of bias for all five indices.

Synthesis of results

Primary analysis

Initial meta-analyses were conducted to compare the intervention and control groups on the primary outcomes at post-intervention and follow-up (see Figures 2 and 3 for forest plots of depression and anxiety symptoms at post-intervention, respectively). The overall
effect size at post-intervention for depression was small ($n^2=74$, $g=.23$, 95%CI: .19-.28), with moderate heterogeneity ($I^2=57$, CI: 0-.97). The effect for depression was small at short-term ($n=41$, $g=.20$, 95%CI: .14-.26); medium-term ($n=34$, $g=.12$, 95%CI: .07-.17); and long-term ($n=14$, $g=.11$, 95%CI: .04-.18) follow-up. For anxiety, the overall effect size at post-intervention was small ($n=49$, $g=.20$, 95%CI: .14-.25), with moderate heterogeneity ($I^2=55$, CI: .38-.67). The effect size for anxiety was comparable at the first two follow-up periods, with conventionally small effects at short- ($n=11$, $g=.23$, 95%CI: .09-.37) and medium-term ($n=20$, $g=.23$, 95%CI: .13-.33) follow-up. At long-term follow-up, the effect size for anxiety was also small ($n=5$, $g=.13$, 95%CI: .04-.22).

**Subgroup analyses**

**Program type**

A subgroup meta-analysis (see Table 2 for results) was conducted to investigate if post-intervention effect sizes differed according to prevention type (i.e., universal vs. targeted). For depression, there was a statistically significant difference ($Q=6.05$, $df=1$, $p=.01$) in the effect size obtained for universal ($n=39$, $g=.19$, 95%CI: .14-.24) compared to targeted ($n=35$, $g=.32$, 95%CI: .23-.41) prevention programs. For anxiety, there was no statistically significant difference between the effect size as a function of prevention type ($Q=0.12$, $df=1$, $p=.73$; universal: $n=32$, $g=.19$, 95%CI: .13-.26; targeted: $n=17$, $g=.22$, 95%CI: .09-.34). No significant differences were evident at the short, medium or long-term follow up time points between universal and targeted programs. 

**Program personnel**

---

5 $n$ represents number of comparisons included in each analysis

6 Short-term follow-up (depression: $Q=8.2$, $df=1$, $p=.37$; universal: $n=17$, $g=.18$, 95%CI: .10-.26; targeted: $n=24$, $g=.23$, 95%CI: .14-.31); anxiety: $Q=1.50$, $df=1$, $p=.22$; universal: $n=5$, $g=.17$, 95%CI: .01-.32; targeted: $n=6$, $g=.36$, 95%CI: .11-.61); medium-term (depression: $Q=9.05$, $df=1$, $p=.33$; universal: $n=18$, $g=.09$, 95%CI: .04-.23; anxiety: $Q=1.62$, $df=1$, $p=.20$; universal: $n=14$, $g=.26$, 95%CI: .13-.40; targeted: $n=6$, $g=.14$, 95%CI: .00-.27); or long-term (depression: $Q=1.01$, $df=1$, $p=.32$; universal: $n=5$, $g=.09$, 95%CI: .02-.20; targeted: $n=9$, $g=.16$, 95%CI: .07-.27); anxiety: $Q=0.8$, $df=1$, $p=.78$; universal: $n=3$, $g=.12$, 95%CI: .00-.01; targeted: $n=2$, $g=.15$, 95%CI: .06-.36).
A second sub-group meta-analysis (see Table 2 for results) was conducted to explore if the personnel involved in delivering the program (classroom teachers/school health staff vs. external providers) influenced the size of the effects obtained. For depression, there was a significant effect, $Q=7.41$, $df=1$, $p=.006$ with externally-delivered programs having a larger effect ($n=51$, $g=.30$, 95% CI: .22-.37), than programs delivered or supported by school staff ($n=23$, $g=.17$, 95% CI: .11-.22). This difference remained significant at the short-term follow-up period, $Q=11.56$, $df=1$, $p=.001$, with externally-delivered programs showing stronger effects ($n=26$, $g=.29$, 95% CI: .20-.38), compared to those delivered by school staff ($n=15$, $g=.11$, 95% CI: .05-.16). At medium- and long-term follow-up this difference was no longer significant (medium term: $Q=.16$, $df=1$, $p=.69$, external: $n=24$, $g=.11$, 95% CI: .06-.16; school staff: $n=10$, $g=.14$, 95% CI: .02-.25; long-term: $Q=.17$, $df=1$, $p=.68$, external: $n=9$, $g=.15$, 95% CI: .04-.26; school staff: $n=5$, $g=.12$, 95% CI: .00-.23). For anxiety symptoms, there was no difference between these subgroups, at post-intervention, $Q=0.37$, $df=1$, $p=.55$, with similar effect sizes found for externally-delivered programs ($n=30$, $g=.21$, 95% CI: .14-.29), and school staff delivered/supported programs ($n=19$, $g=.18$, 95% CI: .10-.26). No differences were detected at any of the follow-up periods.

**Control condition**

Further subgroup meta-analyses were conducted to compare if the magnitude of effect sizes differed as a function of the control condition the program was compared to. For depression programs, there was no overall statistically significant difference as a function of control group type ($Q=1.96$, $df=2$, $p=.38$). The effects for all three control group types were in the small to medium range (NI: $n=47$, $g=.22$, 95% CI: .16-.27; WL: $n=8$, $g=.36$, 95% CI: .16-.56; AC: $n=19$, $g=.24$, 95% CI: .13-.35). No significant differences were apparent at any

\footnote{Anxiety outcomes at short-term ($Q=2.93$, $df=1$, $p=.09$, externally delivered: $n=8$, $g=.33$, 95% CI: .13-.52; school staff delivered: $n=3$, $g=.11$, 95% CI: .06-.27), medium term ($Q=0.11$, $df=1$, $p=.75$, externally delivered: $n=12$, $g=.25$, 95% CI: .12-.38; school staff delivered: $n=8$, $g=.21$, 95% CI: .04-.38), long-term ($Q=0.09$, $df=1$, $p=.76$, externally delivered: $n=3$, $g=.15$, 95% CI: .01-.31; school staff delivered: $n=2$, $g=.12$, 95% CI: .01-.23).}
of the three follow-up time points (all \( p > .05 \)). For anxiety programs, there was a trend-level difference in effect size as a function of control group, \( Q = 5.78, df = 2, p = .06 \), with the WL control comparisons yielding larger effect sizes, \( n = 16, g = .29, 95\% CI: .16-.43 \), relative to NI, \( n = 23, g = .19, 95\% CI: .11-.26 \), and AC, \( n = 10, g = .10, 95\% CI: .01-.19 \) comparison types. However, it should be noted that there was a high level of heterogeneity for the WL control group (\( I^2 = 73 \)). At the short-term follow-up, this trend level difference was no longer evident, \( Q = 4.17, df = 2, p = .13 \), (WL: \( n = 4, g = .34, 95\% CI: .20-.49 \); NI: \( n = 6, g = .19, 95\% CI: -.02-.40 \); AC: \( n = 1, g = .05, 95\% CI: -.22-.31 \)). There were no differences in the medium term follow-up, \( Q = 4.26, df = 2, p = .12 \). Only studies including no-intervention control groups (\( n = 5 \)) conducted long-term follow-ups and so no further analyses could be performed.

**Program delivery age**

Subgroup analyses were conducted to determine if the age at which programs were delivered had an impact on the size of intervention effects (see Table 2). For depression, no significant between-group differences were found, \( Q = 3.07, df = 2, p = .22 \), (children: \( n = 5, g = .50, 95\% CI: .19-.80 \), early adolescents: \( n = 32, g = .23, 95\% CI: .16-.30 \), older adolescents: \( n = 37, g = .22, 95\% CI: .15-.28 \)). For anxiety outcomes, there was no overall effect for age of program delivery, \( Q = 3.31, df = 2, p = .19 \), (children: \( n = 15, g = .23, 95\% CI: .09-.38 \), early adolescents: \( n = 22, g = .21, 95\% CI: .15-.28 \), older adolescents: \( n = 12, g = .12, 95\% CI: .02-.21 \)). No differences emerged over the follow-up periods for depression or anxiety outcomes (all \( p > .05 \)).

**Program content**

The comparison of outcomes as a function of program content indicated that there was no statistically significant difference for CBT vs other programs for depression, \( Q = 1.98, df = 1, p = .16 \), (CBT: \( n = 66, g = .22, 95\% CI: .18-.27 \), other: \( n = 8, g = .38, 95\% CI: .17-.59 \), or
anxiety, $Q=60$, $df=1$, $p=.44$, (CBT: $n=46$, $g=.19$, 95% CI: .13-.25, other: $n=3$, $g=.36$, 95% CI: -.06-.78). There were no differences at any of the follow-up intervals (all $ps>.05$).

**Meta-regression**

We conducted a multivariate meta-regression with effect size at post-intervention as the dependant variable, and study characteristics used in the subgroup analysis entered into the regression model as predictors of outcome. To avoid collinearity among the predictors in the model, we first examined correlations between the variables. All correlations were less than $r=.23$, suggesting that none of the variables were highly confounded and were therefore retained in the model. Within each category (prevention type, personnel delivering program, control group comparison, age of delivery, program content), we first defined a reference group. Reference group information and results are presented in Table 3. For depression, only one variable emerged as significant, with targeted studies having larger effect sizes than universal programs. At the trend level ($p<.10$), programs delivered by facilitators external to the school had larger effect sizes, and programs delivered to older adolescents has smaller effect sizes than those delivered to children (although it is noted that there were only five depression studies delivered to children). There were no significant predictors of effect size for anxiety prevention studies.

**Publication bias**

There was some evidence of publication bias for the depression studies, as demonstrated by inspection of the funnel plot (Appendix II). After adjusting for publication bias using Duval and Tweedie’s trim and fill procedure, the estimate of the mean effect size at post-intervention reduced from $g=.23$ to $g=.15$ (23 studies removed). There was no evidence of publication bias for anxiety studies.

**Discussion**
The aim of this study was to provide a review and evaluation of studies investigating school-based programs designed to prevent depression and/or anxiety in children and adolescents. Across the eighty-one included RCTs involving 31,794 participants, our findings show that school-based prevention programs have a small beneficial effect on depressive and anxiety symptoms when compared to a control condition.

For the studies focused on depression, the effect sizes at post-intervention and short-term follow-up indicated a small effect (\(g=0.23\) and \(g=0.20\) respectively). There was evidence of a very small effect of the depression-prevention programs at the medium-term follow-up (6-12 months; \(g=0.12\)), and long term follow-up (more than 12-months, \(g=0.11\)). It is encouraging that 14 studies included follow-up intervals of more than 12-months, which is important in prevention where the end-goal is to detect prevention effects which necessarily become evident over time. That said, the effect sizes for study follow-up assessment points were modest. Several studies included follow-ups for longer-term periods and suggest that gains can be maintained at 24 month follow-up period (e.g., Stice, Rohde, Gau, & Wade, 2010), but tend to dissipate at longer intervals (e.g., Hunt, Andrews, Crino, Erskine, & Sakashita, 2009; Johnstone, Rooney, Hassan, & Kane, 2014).

The pattern of findings for the anxiety prevention programs was similar to that of depression programs, with a small effect at post-intervention (\(g=0.10\)), which was maintained at the short-term and medium-term follow-ups (\(g=0.23\) for both). By the long-term follow-up, the effect of the anxiety programs was marginal (\(g=0.13\)). Long term follow-ups for anxiety programs were less common, with only five studies including long-term follow-up intervals, again with little evidence of gains being maintained beyond the 24-month follow-up period (e.g., Johnstone et al., 2014). Taken together, these data suggests that the effect of anxiety prevention programs is maintained at 6-12 months after the program is delivered and highlights the need for long-term follow-up assessments to establish whether gains remain at
and beyond this point. Indeed, more research is needed to accurately assess if prevention programs in general have the potential for prolonged effects. It is possible that the deterioration in effects is due to natural decay over time, or could reflect reduced power with smaller samples being retained over protracted periods. It will be important for future studies to identify the parameters under which prevention effects are most likely to be maintained. For example, it is possible that prevention effects might be augmented by the delivery of booster sessions. Alternatively, low intensity prevention strategies such as automated, computerised modules could be incorporated into standard classroom activities as intermittent reminders following a more intensive program, if this is likely to maintain preventive effects.

A common criticism of prevention programs is that they are associated with small effect sizes. While this is true, it is important to keep in mind that in prevention (as compared to treatment), even small effect sizes are likely to be associated with meaningful improvements particularly at a population level. A small effect size difference in this context is likely to have implications for preventing the onset of these disorders in youth. Analyses that have been conducted on diagnostic outcomes following prevention programs consistently support this assertion, with a recent review showing that preventive programs are associated with a 53% decrease in risk of internalising disorder onset in the 6-9 months following program delivery (Stockings et al., 2016). These studies are difficult to do because many school-based trials often do not include diagnostic outcomes, so symptom levels provide a meaningful proxy for disorder. Emerging online diagnostic screening tools may enable researchers to more often collect diagnostic information in the school environment (Gibbons et al., 2012).

The effect sizes reported in the current review are comparable to those reported in previous reviews of depression (Calear & Christensen, 2010; Merry et al., 2011), and anxiety prevention (Fisak et al., 2011; Stockings et al., 2016). Slight variability in effect size
estimates is likely to be due to a number of factors, including the absence of pooled effect sizes in previous reviews (e.g., Calear & Christensen, 2010), and the more conservative method in which effect sizes were calculated in the present review (Hedges $g$ rather than Cohen’s $d$ in order to adjust for small sample sizes).

The quality assessments undertaken in the current review indicated that 80 of the 81 RCTs included some degree of bias. This suggests that there is substantial room for improvement in the rigour and quality of the research conducted in this field, and in the reporting of study methodologies and outcomes. Ways in which the quality of school-based prevention trials may be improved include using random sequence generation methods and adequate allocation concealment. It is possible that these methods actually were followed in many of the trials included, but a majority of authors failed to provide sufficient detail around these possible selection biases. Moreover, the increasingly common requirement to publish study protocols will mitigate the risk of selective reporting moving forward, and authors are encouraged to be transparent in their report of outcome data and attrition. Finally, randomisation at the school-level for school-based trials protects against contamination. The potential for contamination in 64% of the included trials would be expected to actually reduce outcomes because participants in control conditions may have accessed material covered in the prevention condition. This means that current effect size estimates may be conservative. Cluster randomisation at the school level in prevention research will protect against this risk. These changes, as well as many journals now requiring submissions follow the CONSORT statement (Schulz, Altman, & Moher, 2010) will help to improve the overall quality of RCTs in the area.

Moderate levels of heterogeneity were also found across the studies included in this review and meta-analysis. This finding was not surprising, given that prevention trials can vary enormously in terms of their sample, the type of program evaluated and how it is
delivered. The results from sub-group analyses reveal some possible reasons for this heterogeneity, as discussed below.

**Sub-group and meta-regression analyses outcomes**

The analysis comparing universal and targeted prevention programs found that for depression-prevention programs, targeted prevention yielded a significantly greater effect size relative to universal programs at post-test, even when other factors (e.g., facilitator type, age) were controlled for. This was not the case for anxiety, and was not maintained at any of the follow-up periods for either depression or anxiety. The existing literature on this question has produced mixed results, with some previous meta-analyses reporting results consistent with those reported here (e.g., Merry et al., 2011), although null and opposite effects have been reported (e.g., Fisak et al., 2011; Stockings et al., 2016). However, a key difference between the current a previous reviews is that previous reviews did not distinguish between school-based and community settings. The data from the current review suggests for depression programs delivered in the school environment, targeted intervention may be more efficacious. However, several other factors need to be taken into account in the targeted vs universal debate. First, research evaluations of universal prevention programs require much larger sample sizes to detect effects than targeted programs, which are often impractical or prohibitively expensive to conduct (Muñoz et al., 2010). Therefore, smaller effect sizes reported in the literature may be due to limits in statistical power to detect effects, not because they do not work. Additionally, universal prevention in the school context confers several advantages over targeted prevention that include: (i) removal of the need for screening; (ii) minimisation of stigma because no student is singled out, and; (iii) capturing youth who may not be at risk yet, but will go on to develop symptoms in the future. It is notable that school administrators often prefer universal programs due to ease of scheduling programs (Horowitz, Garber, Ciesla, Young, & Mufson, 2007). Both universal and targeted
prevention programs yielded small effect sizes in the current review, and it is likely that there is value in pursuing both kinds of prevention. Indeed, no differences for anxiety outcomes as a function of prevention type were evident. A suggestion for how future studies may be improved is to take a stepped care approach where a universal program is delivered in the first instance, and followed up with a targeted program for at risk or symptomatic individuals who do not respond to the universal program.

Analyses comparing facilitator type as a predictor of effect size revealed that depression programs were more efficacious when delivered by individuals external to the school environment (e.g., researchers, mental health professionals, graduate students), relative to school-staff at post-intervention and short-term follow-up. This finding is in accordance with a previous review reporting a series of larger effect sizes for programs delivered by individuals external to the school environment (Calear & Christensen, 2010) and another suggesting that there is not yet enough evidence to show that programs delivered by internal providers are effective (Brunwasser & Garber, 2015). It is noteworthy to consider that targeted programs were more likely to be delivered by external facilitators than universal programs were (74% externally delivered for targeted vs 64% externally delivered for universal), although the meta-regression confirmed this difference was not statistically significant. The differential effect sizes as a consequence of personnel delivering the intervention was not replicated for anxiety programs, with no difference in effect size between programs delivered by school vs external staff at post-intervention, or any of the follow-up time points. We speculate that this could be due to the fact that anxiety programs tended to be delivered to younger participants, whom may feel more comfortable with familiar school staff than unfamiliar external providers.

Provider type is a particularly important factor to consider because programs delivered by school staff provide information about whether the program is likely to be
effective under conditions that are suitable for large-scale implementation. One avenue likely to reduce the resources required to deliver prevention programs in schools are computerised therapies (Richardson, Stallard, & Velleman, 2010), which show promise by reducing the training required for the delivery of both school and mental health staff. Although it was not possible to do a formal analysis on the two computer-delivered programs, symptom differences between the experimental group and controls were detected at post-intervention in both studies ($d=0.15$ for anxiety, Cear et al., 2009; $d=0.14$ for depression, $d=0.18$ for anxiety, Wong et al., 2014). An additional benefit of computer-delivered programs is that they protect the fidelity of the content being delivered. Therefore, outcomes might actually be augmented by having teachers support their delivery, as was the case in one study whereby teachers led a classroom discussion following each treatment module to reinforce the ideas presented by the program (Wong et al., 2014).

A major limitation of the school-based prevention field has been the lack of active control groups (e.g., Merry et al., 2011). The results of the current review suggest that studies are increasingly including active attention control groups, which typically involve a structured program (not including active elements of the prevention program). This is a much stronger comparison than NI and WL groups, which tend to involve regular classes as usual. The present review found that 25% of studies evaluating prevention programs included an attention control condition. This compares more favourably to previous reviews of anxiety and depression programs that reported attention control conditions in between 9% and 15% of trials (Cear & Christensen, 2010; Neil & Christensen, 2009). However, the sub-group analyses suggested that the control group used did not influence the size of the effect. This finding may suggest one of two things: firstly, that non-specific program factors, such as being part of a group or receiving a program outside of regular school classes does not impact program outcomes, or secondly, that there are yet to be a sufficient number of trials with
attention control conditions to accurately detect a difference. Further research involving attention control conditions will help to ascertain if the control group utilised in a trial can inflate intervention effects.

Sub-group analyses suggested that the age at which a prevention program was delivered influenced the size of the intervention effects obtained, with the exception of a trend towards depression programs delivered to children being more effective than those delivered to older adolescents. Previous research has suggested the need to deliver prevention programs prior to the onset of a clinical disorder (Cuijpers et al., 2008). Drawing from epidemiological research, it would therefore be expected that depression programs would be best delivered in childhood or early adolescence, as the mean age of onset for depression is between 11 and 14 years (Merikangas et al., 2009), while for anxiety disorders, delivery during childhood would be best given the early age of onset of many anxiety disorders (Kessler et al., 2005; Merikangas et al., 2010). The results of the current review may lend some support for this suggestion, although there were only five studies delivering depression prevention programs to children aged younger than ten years old. It is important to note that studies were included even if they did not screen for the presence of current or past diagnoses, and so the effects may have been obscured by individuals essentially receiving treatment or relapse prevention rather than prevention.

Other factors

Almost half of the included studies did not measure program fidelity, and among those that did, the use of program leader self-report checklists was pervasive. Less than one third of studies recorded sessions for independent fidelity ratings. Measuring and protecting program fidelity must be a priority for future studies. A promising solution to this problem is the advent of internet-based psychological interventions, which by their very nature protect
program fidelity. An additional benefit associated with internet-based programs is that they can automatically collect adherence information. Mapping a dose-response outcome as a function of program adherence will provide important data regarding the parameters needed to achieve preventive effects.

The current review also established that parents vary substantially in rates of participation in school-based prevention programs, with some studies reporting very poor parental attendance levels (e.g., Lock & Barrett, 2003). The impact of involving parents on outcome is not yet clear, although there is some preliminary evidence that their involvement may support the longer term maintenance of intervention effects (Manassis et al., 2014). Therefore, future studies will need to quantify the impact of parental involvement, and explore strategies designed to incentivise and improve parental uptake rates.

**Future research directions**

Overall, the findings from the current meta-analysis suggest that there is merit in continuing to develop, deliver and evaluate school-based prevention programs for depression and anxiety. This review establishes that even when an ecologically valid approach is taken, and stringent entry criteria that exclude those who may already be experiencing significant symptoms is *not* imposed, meaningful effects are still detected. A related challenge for the future is improving reach, coverage, and the availability of prevention programs. Advances in technology provide one avenue by which to address reach and availability, although how this might be delivered at a population level has not yet been established. Finally, enhancing youth engagement in these programs is necessary if program adherence is to be improved. Engaging asymptomatic young people is challenging because the material may be viewed as irrelevant. Consulting with young people and gaining insight into their perspective on the issue may help to elucidate the barriers to engagement.
A priority for the field moving forward is a closer consideration of ways to accelerate efficacy findings into the effectiveness domain. A recent review of depression prevention concluded that there is strong evidence for program efficacy, but evidence for effectiveness under real-world conditions is still accumulating (Brunwasser & Garber, 2015). To ensure efficacious programs get translated into programs ready for wide-scale rollout, efforts now need to be directed towards pragmatic trial designs that enable testing the infrastructure and personnel that will ultimately required for sustainable, long-term programs to be adopted by schools.

**Limitations**

The results of this meta-analysis need to be interpreted in the context of several limitations. First, the meta-analysis was based exclusively on self-report symptom measures and not clinician-rated measures. Self-report symptoms measures were used in the current review as the majority of the studies that met the inclusion criteria did not report clinician-rated diagnostic outcomes and, from a practical perspective, it is unlikely that when delivered under real-world settings, programs will include clinician assessments because they are expensive and time consuming to administer. Self-report measures are acceptable in the field to measure intervention effects, with many researchers arguing that they are as important as clinical diagnoses, as they are associated with comparable degrees of impairment (Gillham, Shatté, & Reivich, 2001; Horowitz et al., 2007).

A second limitation is that studies were included in the review even if they did not exclude participants with significant symptoms, of which more than 80% of included studies did not. The rationale for this was to emulate procedures that are likely to be employed under real world conditions, as is typically characteristic of universal prevention programs. While this may have impacted effect size estimates as a consequence of including symptomatic
individuals, it did allow an overall indication as to the effects of programs as they would be delivered in the real world.

Finally, there was evidence of publication bias in the depression studies and so the effect size estimates may have been inflated. The increasing trend towards publication of study protocols should help to protect against selective publication of studies reporting significant group differences.

**Conclusion**

Overall, the findings from the current meta-analysis suggest that there is merit in continuing to evaluate and deliver school-based prevention programs for depression and anxiety. While prevention type and personnel delivering the intervention account for aspects of the heterogeneity observed, more research is needed to identify how program completion and fidelity impacts outcome, at what age program delivery is optimal, and whether there is a need to involve parents. The overall quality of included RCTs was low, and heterogeneity was moderate. Large-scale effectiveness trials that evaluate implementation efforts that are embedded within the school system are now needed in order to identify the most successful ways to roll-out prevention programs on scale.
References


Clinical Child & Adolescent Psychology, 1, 1-21.
doi:10.1080/15374416.2015.1020541

doi:10.1016/j.adolescence.2009.07.004


doi:http://dx.doi.org/10.1016/S0962-1849(01)80012-3


* = studies included in the review
Table 1: School based prevention programs for depression, anxiety and both depression/anxiety

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Prevention type</th>
<th>Age range</th>
<th>N</th>
<th>Program</th>
<th>Control</th>
<th>Program content</th>
<th>Mode of delivery</th>
<th>Number sessions</th>
<th>Outcome measure</th>
<th>Quality Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Araya et al. 2013</td>
<td>Chile</td>
<td>Universal</td>
<td>NR</td>
<td>2512</td>
<td>ITFA</td>
<td>WL</td>
<td>CBT</td>
<td>MHP</td>
<td>11 + 2 booster</td>
<td>BDI-II</td>
<td>?</td>
</tr>
<tr>
<td>Cardemil et al. 2002 (Study 1)</td>
<td>America</td>
<td>Targeted (selective)</td>
<td>NR</td>
<td>49</td>
<td>PRP</td>
<td>NI</td>
<td>CBT</td>
<td>MHP</td>
<td>12</td>
<td>CDI</td>
<td>?</td>
</tr>
<tr>
<td>Cardemil et al 2002 (Study 2)</td>
<td>America</td>
<td>Targeted (selective)</td>
<td>NR</td>
<td>103</td>
<td>PRP</td>
<td>NI</td>
<td>CBT</td>
<td>MHP</td>
<td>12</td>
<td>CDI</td>
<td>?</td>
</tr>
<tr>
<td>Castellanos &amp; Conrod, 2006</td>
<td>UK</td>
<td>Targeted (selective)</td>
<td>13-16 years</td>
<td>423</td>
<td>PM-CBI</td>
<td>NI</td>
<td>CBT</td>
<td>MHP</td>
<td>2</td>
<td>BSI-DEP</td>
<td>?</td>
</tr>
<tr>
<td>Chaplin et al. 2006</td>
<td>America</td>
<td>Universal</td>
<td>11-14 years</td>
<td>208</td>
<td>PRP</td>
<td>NI</td>
<td>CBT</td>
<td>Grad + Teacher</td>
<td>12</td>
<td>CDI</td>
<td>+</td>
</tr>
<tr>
<td>Clarke et al. 1993 (Study 1)</td>
<td>America</td>
<td>Universal</td>
<td>NR</td>
<td>622</td>
<td>Psycho-education</td>
<td>NI</td>
<td>EDU Teacher</td>
<td>3</td>
<td>CES-D</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Clarke 1993 et al. (Study 2)</td>
<td>America</td>
<td>Universal</td>
<td>NR</td>
<td>380</td>
<td>BI</td>
<td>NI</td>
<td>CBT</td>
<td>Teacher</td>
<td>5</td>
<td>CES-D</td>
<td>?</td>
</tr>
<tr>
<td>Clarke et al. 1995</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>14-16 years</td>
<td>150</td>
<td>CWSC</td>
<td>NI</td>
<td>CBT</td>
<td>MHP</td>
<td>5</td>
<td>CES-D</td>
<td>?</td>
</tr>
<tr>
<td>Gillham et al. 2007</td>
<td>America</td>
<td>Universal</td>
<td>11-14 years</td>
<td>697</td>
<td>PRP</td>
<td>AC + NI</td>
<td>CBT</td>
<td>Teacher + MHP</td>
<td>12</td>
<td>CDI (Anx: RCMAS)</td>
<td>+</td>
</tr>
<tr>
<td>Gillham et al. 2012</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>10-15 years</td>
<td>408</td>
<td>PRP-A</td>
<td>NI</td>
<td>CBT + IPT</td>
<td>Teacher + School MHP</td>
<td>10</td>
<td>CDI</td>
<td>+</td>
</tr>
<tr>
<td>Horowitz et al., 2007</td>
<td>America</td>
<td>Universal</td>
<td>14-15 years</td>
<td>380</td>
<td>CWSC</td>
<td>AC + NI</td>
<td>CBT + IPT</td>
<td>MHP</td>
<td>8</td>
<td>CDI</td>
<td>+</td>
</tr>
<tr>
<td>Kindt et al. 2014</td>
<td>Holland</td>
<td>Targeted (selective)</td>
<td>11-16 years</td>
<td>1440</td>
<td>OVK</td>
<td>NI</td>
<td>CBT</td>
<td>Teacher</td>
<td>16</td>
<td>CDI</td>
<td>+</td>
</tr>
<tr>
<td>King &amp; Kirschenbaum, 1990</td>
<td>America</td>
<td>Targeted (selective)</td>
<td>8 years</td>
<td>127</td>
<td>WEI</td>
<td>NI</td>
<td>SS</td>
<td>MHP</td>
<td>24</td>
<td>CDRS-R</td>
<td></td>
</tr>
<tr>
<td>*Lamb et al. 1998</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>14-19 years</td>
<td>46</td>
<td>-</td>
<td>NI</td>
<td>CBT</td>
<td>School MHP</td>
<td>8</td>
<td>RADS</td>
<td>?</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Targeted (indicated)</td>
<td>Age Range (years)</td>
<td>Participants</td>
<td>Measures</td>
<td>Interventions</td>
<td>Therapists</td>
<td>Referral</td>
<td>Implementation</td>
<td>Side Effects</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>----------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>----------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>McCarty et al. 2013</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>11-15</td>
<td>120</td>
<td>MFQ</td>
<td>PTA, AC, CBT, MHP</td>
<td>12</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>McCarty et al. 2011</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>NR</td>
<td>67</td>
<td>MFQ</td>
<td>PTA, AC, CBT, MHP</td>
<td>12</td>
<td></td>
<td>?</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Merry et al. 2004</td>
<td>New Zealand</td>
<td>Universal</td>
<td>13-15</td>
<td>392</td>
<td>BDI-II</td>
<td>RAP-KIWI, AC, CBT + IPT, Teacher</td>
<td>11</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Pattison &amp; Lynd-</td>
<td>Australia</td>
<td>Universal</td>
<td>9-12</td>
<td>66</td>
<td>CDI</td>
<td>PPP, AC + NI, CBT, MHP</td>
<td>11</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Stevenson, 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possel et al. 2004</td>
<td>Germany</td>
<td>Universal</td>
<td>13-14</td>
<td>347</td>
<td>CES-D</td>
<td>L-T, NI, CBT, MHP</td>
<td>10</td>
<td></td>
<td>-</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Possel et al. 2004</td>
<td>Germany</td>
<td>Universal</td>
<td>12-13</td>
<td>301</td>
<td>SBB-DES</td>
<td>L&amp;L, NI, CBT, Grad + MHP</td>
<td>10</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Possel et al. 2013</td>
<td>America</td>
<td>Universal</td>
<td>14-16</td>
<td>518</td>
<td>CDI</td>
<td>L&amp;L, AC + NI, CBT, Grad + MHP</td>
<td>10</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Puskar et al. 2003</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>14-18</td>
<td>89</td>
<td>RADS</td>
<td>TKC, NI, CBT, MHP</td>
<td>10 + 1 booster</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Quayle et al. 2001</td>
<td>Australia</td>
<td>Universal</td>
<td>11-12</td>
<td>47</td>
<td>CDI</td>
<td>OLP, NI, CBT, MHP</td>
<td>8</td>
<td></td>
<td>?</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Raes et al. 2014</td>
<td>Belgium</td>
<td>Universal</td>
<td>13-20</td>
<td>408</td>
<td>DASS-DEP</td>
<td>MI, NI, MI, MHP</td>
<td>8</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Roberts et al. 2003;</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>11-13</td>
<td>189</td>
<td>CDI</td>
<td>PPP, NI, CBT, MHP</td>
<td>12</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Rhode et al. 2014</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>13-19</td>
<td>328</td>
<td>K-SADS</td>
<td>- , AC + NI, CBT, School MHP</td>
<td>6</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Rooney et al. 2006</td>
<td>Australia</td>
<td>Universal</td>
<td>8-9</td>
<td>120</td>
<td>CDI</td>
<td>PTP, NI, CBT, MHP</td>
<td>8</td>
<td></td>
<td>?</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Rose et al. 2014</td>
<td>Australia</td>
<td>Universal</td>
<td>9-14</td>
<td>210</td>
<td>RADS-2</td>
<td>RAP + PIR, AC + NI, CBT + IPT, Grad</td>
<td>20</td>
<td></td>
<td>?</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Sheffield et al. 2006</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>13-15</td>
<td>1360</td>
<td>CDI</td>
<td>- , NI, CBT, Teacher</td>
<td>8</td>
<td></td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Spence et al. 2003;</td>
<td>Australia</td>
<td>Universal</td>
<td>12-14</td>
<td>1250</td>
<td>BDI</td>
<td>PSFL, NI, CBT, Teacher</td>
<td>8</td>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Stallard et al. 2012</td>
<td>England</td>
<td>Targeted (indicated)</td>
<td>12-16</td>
<td>1064</td>
<td>SMFQ</td>
<td>RAP-UK, AC + NI, CBT, Grad</td>
<td>9</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Intervention Type</td>
<td>Age</td>
<td>Sample Size</td>
<td>Outcome Measure</td>
<td>Measure Details</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------------------------</td>
<td>-----</td>
<td>-------------</td>
<td>-----------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stice et al. 2007</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>15-22 years</td>
<td>225</td>
<td>BG AC + WL</td>
<td>CBT Grad 4 BDI -</td>
<td>? ? ? -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wijhoven et al. 2014</td>
<td>Holland</td>
<td>Targeted (indicated)</td>
<td>11-15 years</td>
<td>102</td>
<td>OVK NI</td>
<td>CBT MHP 8 CDI +</td>
<td>+ + + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woods &amp; Jose, 2011</td>
<td>New Zealand</td>
<td>Targeted (indicated)</td>
<td>NR</td>
<td>56</td>
<td>ACE-Kiwi NI</td>
<td>CBT School MHP 8 CDI ?</td>
<td>? - ? -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young et al. 2006</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>11-16 years</td>
<td>41</td>
<td>IPT-AST AC</td>
<td>IPT Grad + Researcher 10 CES-D +</td>
<td>? + + ? -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young et al. 2010</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>13-17 years</td>
<td>57</td>
<td>IPT-AST NI</td>
<td>IPT MHP 10 CES-D +</td>
<td>- + + -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Anxiety Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Intervention Type</th>
<th>Age</th>
<th>Sample Size</th>
<th>Outcome Measure</th>
<th>Measure Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticich et al. 2013</td>
<td>Australia</td>
<td>Universal</td>
<td>4-7 years</td>
<td>488</td>
<td>FRIENDS AC + WL</td>
<td>CBT Teacher 10 PAS ?</td>
</tr>
<tr>
<td>Balle &amp; Tortella-Feliu, 2010</td>
<td>Spain</td>
<td>Targeted (selective)</td>
<td>11-17 years</td>
<td>145</td>
<td>FRIENDS WL</td>
<td>CBT Grad + MHP 6 SCAS ?</td>
</tr>
<tr>
<td>Barrett &amp; Turner, 2001</td>
<td>Australia</td>
<td>Universal</td>
<td>10-12 years</td>
<td>489</td>
<td>FRIENDS AC + NI</td>
<td>CBT Teacher vs MHP 10 + 2 SCAS booster</td>
</tr>
<tr>
<td>*Barrett et al. 2005</td>
<td>Australia</td>
<td>Universal</td>
<td>9-16 years</td>
<td>692</td>
<td>FRIENDS NI</td>
<td>CBT Grad + MHP 10 + 2 SCAS booster</td>
</tr>
<tr>
<td>Berger et al. 2007</td>
<td>Israel</td>
<td>Universal</td>
<td>7-11 years</td>
<td>142</td>
<td>OTT WL</td>
<td>CBT Teacher 8 SCARED-GAD -</td>
</tr>
<tr>
<td>Bouchard et al. 2013</td>
<td>Canada</td>
<td>Universal</td>
<td>9-12 years</td>
<td>59</td>
<td>DHS WL</td>
<td>CBT Grad + MHP 10 MASC ?</td>
</tr>
<tr>
<td>Cooley-Strickland et al. 2011</td>
<td>America</td>
<td>Targeted (selective/ indicated)</td>
<td>8-12 years</td>
<td>93</td>
<td>FRIENDS WL</td>
<td>CBT Grad 13 RCMAS ?</td>
</tr>
<tr>
<td>Dadds et al. 1997; 1999</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>7-14 years</td>
<td>128</td>
<td>CK NI</td>
<td>CBT MHP 10 + 2 booster RCMAS ?</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Targeted Level</td>
<td>Age Range</td>
<td>Sample Size</td>
<td>Intervention</td>
<td>Diagnostic Measure</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Essau et al. 2012</td>
<td>Germany</td>
<td>Universal</td>
<td>9-12 years</td>
<td>638</td>
<td>FRIENDS</td>
<td>CBT Grad</td>
</tr>
<tr>
<td>Garaigordobil 2004</td>
<td>Spain</td>
<td>Universal</td>
<td>12-14 years</td>
<td>174</td>
<td>-</td>
<td>NI SS Teacher</td>
</tr>
<tr>
<td>Hiebert et al. 1989 (Study 2)</td>
<td>Canada</td>
<td>Universal</td>
<td>13-14 years</td>
<td>113</td>
<td>-</td>
<td>AC CBT Teacher</td>
</tr>
<tr>
<td>Hunt et al. 2009</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>11-13 years</td>
<td>259</td>
<td>FRIENDS</td>
<td>NI CBT School MHP + Teacher</td>
</tr>
<tr>
<td>Keogh et al. 2006</td>
<td>UK</td>
<td>Universal</td>
<td>15-16 years</td>
<td>80</td>
<td>SMI</td>
<td>NI CBT MHP</td>
</tr>
<tr>
<td>Kiselica et al. 1994</td>
<td>America</td>
<td>Targeted (indicated)</td>
<td>NR</td>
<td>48</td>
<td>SIT</td>
<td>NI CBT MHP</td>
</tr>
<tr>
<td>Lock &amp; Barrett, 2003; Barrett et al. 2006</td>
<td>Australia</td>
<td>Universal</td>
<td>9-16 years</td>
<td>737</td>
<td>FRIENDS</td>
<td>WL CBT Grad + MHP 10+2 booster</td>
</tr>
<tr>
<td>McLoone &amp; Rapee, 2012</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>7-10 years</td>
<td>152</td>
<td>CK</td>
<td>WL CBT School MHP 10</td>
</tr>
<tr>
<td>Miller et al. 2010</td>
<td>Canada</td>
<td>Universal</td>
<td>7-12 years</td>
<td>116</td>
<td>TWD</td>
<td>WL CBT Teacher</td>
</tr>
<tr>
<td>Miller et al. 2011a</td>
<td>Canada</td>
<td>Universal</td>
<td>7-13 years</td>
<td>533</td>
<td>FRIENDS</td>
<td>WL CBT Teacher + School MHP</td>
</tr>
<tr>
<td>Miller et al. 2011b (Study 1)</td>
<td>Canada</td>
<td>Targeted (indicated)</td>
<td>9-12 years</td>
<td>191</td>
<td>FRIENDS</td>
<td>WL CBT Teacher + School MHP</td>
</tr>
<tr>
<td>Miller et al. 2011b (Study 2)</td>
<td>Canada</td>
<td>Universal</td>
<td>9-12 years</td>
<td>253</td>
<td>FRIENDS</td>
<td>WL CBT Teacher + School MHP</td>
</tr>
<tr>
<td>Misfud &amp; Rapee, 2005</td>
<td>Australia</td>
<td>Targeted (indicated)</td>
<td>8-11 years</td>
<td>96</td>
<td>Cool Kids</td>
<td>WL CBT MHP + School MHP</td>
</tr>
<tr>
<td>Pedro-Carroll &amp; Cowen, 1985</td>
<td>America</td>
<td>Targeted (selective)</td>
<td>9-12 years</td>
<td>72</td>
<td>CODIP</td>
<td>WL CBT School MHP, MHP, Grad</td>
</tr>
<tr>
<td>Stallard et al. 2014</td>
<td>England</td>
<td>Universal</td>
<td>9-10 years</td>
<td>1448</td>
<td>FRIENDS</td>
<td>AC + NI CBT Teacher vs MHP</td>
</tr>
</tbody>
</table>

**Depression & Anxiety Studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Targeted Level</th>
<th>Age Range</th>
<th>Sample Size</th>
<th>Intervention</th>
<th>Diagnostic Measure</th>
<th>Duration</th>
<th>Other Measures</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calear et al. 2009</td>
<td>Australia</td>
<td>Universal</td>
<td>12-17 years</td>
<td>1477</td>
<td>MG</td>
<td>WL icCBT Computer (teacher supported)</td>
<td>5</td>
<td>Dep: CES-D Anx: RCMAS</td>
<td>+</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Theme</td>
<td>Age</td>
<td>Duration</td>
<td>Group/Setting</td>
<td>Main Interventions</td>
<td>Target</td>
<td>Time</td>
<td>Control</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>-------</td>
<td>-----</td>
<td>----------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>Manassis et al. 2010</td>
<td>Canada</td>
<td>Targeted (indicated)</td>
<td>8-12 years</td>
<td>148</td>
<td>FC</td>
<td>AC</td>
<td>CBT</td>
<td>Grad + MHP</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 2. Subgroup analyses at post-intervention

<table>
<thead>
<tr>
<th>Subgroup Analyses</th>
<th>Measure</th>
<th>N</th>
<th>g</th>
<th>95% CI</th>
<th>$I^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevention Type</td>
<td>Depression</td>
<td>39</td>
<td>.19</td>
<td>.14-.24</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35</td>
<td>.32</td>
<td>.23-.41</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>32</td>
<td>.19</td>
<td>.13-.26</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17</td>
<td>.22</td>
<td>.09-.34</td>
<td>59</td>
</tr>
<tr>
<td>Personnel Delivering Program</td>
<td>Depression</td>
<td>51</td>
<td>.30</td>
<td>.22-.37</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td>.17</td>
<td>.11-.22</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>30</td>
<td>.21</td>
<td>.14-.29</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>.18</td>
<td>.10-.26</td>
<td>67</td>
</tr>
<tr>
<td>Control Group Comparison</td>
<td>Depression</td>
<td>47</td>
<td>.22</td>
<td>.16-.27</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>.36</td>
<td>.16-.56</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>23</td>
<td>.19</td>
<td>.11-.26</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>.29</td>
<td>.16-.43</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>.10</td>
<td>.01-.192</td>
<td>0</td>
</tr>
<tr>
<td>Age</td>
<td>Depression</td>
<td>5</td>
<td>.50</td>
<td>.19-.80</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
<td>.23</td>
<td>.16-.30</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37</td>
<td>.22</td>
<td>.15-.28</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>15</td>
<td>.23</td>
<td>.09-.38</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22</td>
<td>.21</td>
<td>.15-.28</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>.12</td>
<td>.23-.21</td>
<td>41</td>
</tr>
<tr>
<td>Program Content CBT</td>
<td>Depression</td>
<td>66</td>
<td>.22</td>
<td>.18-.27</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>8</td>
<td>.38</td>
<td>.17-.59</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>CBT</td>
<td>46</td>
<td>.19</td>
<td>.13-.25</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3</td>
<td>.36</td>
<td>-.06-.78</td>
<td>62</td>
</tr>
</tbody>
</table>

Note: N = number of comparison conditions; $I^2$ = heterogeneity; ‘external’ includes mental health professional, researchers, graduates; ‘school-staff’ includes teachers, counsellors, nurses; and teacher-supported computerised programs.
Table 3. Standardised regression coefficients of study characteristics in relation to the effect size of outcomes at post-test for both depression and anxiety

<table>
<thead>
<tr>
<th>Variable</th>
<th>Outcome</th>
<th>B</th>
<th>SE</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevention Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td>Depression</td>
<td>-0.12</td>
<td>0.05</td>
<td>0.03**</td>
</tr>
<tr>
<td>Targeted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal</td>
<td>Anxiety</td>
<td>-0.03</td>
<td>0.08</td>
<td>0.71</td>
</tr>
<tr>
<td>Targeted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel Delivering Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>Depression</td>
<td>-0.09</td>
<td>0.05</td>
<td>0.08*</td>
</tr>
<tr>
<td>School-staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External</td>
<td>Anxiety</td>
<td>-0.03</td>
<td>0.08</td>
<td>0.66</td>
</tr>
<tr>
<td>School-staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No intervention</td>
<td>Depression</td>
<td>0.10</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>Waitlist</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td>-0.04</td>
<td>0.07</td>
<td>0.58</td>
</tr>
<tr>
<td>No intervention</td>
<td>Anxiety</td>
<td>0.05</td>
<td>0.07</td>
<td>0.46</td>
</tr>
<tr>
<td>Waitlist</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td>-0.01</td>
<td>0.08</td>
<td>0.90</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>Depression</td>
<td>-0.17</td>
<td>0.11</td>
<td>.13</td>
</tr>
<tr>
<td>Early adolescent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older adolescent</td>
<td></td>
<td>-0.22</td>
<td>0.11</td>
<td>0.05*</td>
</tr>
<tr>
<td>Child</td>
<td>Anxiety</td>
<td>-0.01</td>
<td>0.08</td>
<td>0.91</td>
</tr>
<tr>
<td>Early adolescent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older adolescent</td>
<td></td>
<td>-0.09</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Program Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td>Depression</td>
<td>0.15</td>
<td>0.09</td>
<td>.12</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT</td>
<td>Anxiety</td>
<td>0.21</td>
<td>0.02</td>
<td>.18</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Ref = reference group. ** = significant at α<.05, *= trend level at α<.10.
Figure Captions

Figure 1. Study flow chart

Figure 2. Forest plot effect sizes for comparisons between prevention programs and control conditions on depressive symptoms at post-intervention

Figure 3. Forest plot of effect sizes for comparisons between prevention programs and control conditions on anxiety symptoms at post-intervention

Appendix A. Search string example

Appendix B. Funnel plot of depression effect size data at post-intervention
Records identified through database searching (n = 5776)

Additional records identified through other sources (n = 141)

Records after duplicates removed (n = 5431)

Records excluded (n = 5212)

Records screened (n = 5431)

Full-text articles assessed for eligibility (n = 219)

Full-text articles excluded, with reasons (n = 129)
- Not in age range: 6
- Not school-based: 31
- Not prevention: 15
- Not psychological: 9
- Not depression or anxiety focused (with primary outcome): 26
- Not an RCT: 40
- Not in English: 2

Articles included in review (n = 90)

Number original trials included in review (n = 81)*

*77 included in meta-analysis

Fig. 1
Fig. 2
Fig. 3
Appendix A – Search String Example

Search String from PubMed:

((depress*[All Fields]) OR mood [All Fields] OR affect [All Fields/MeSH Terms]) OR anxiety [All Fields] OR anxious [All Fields] AND ((school* [All Fields]) OR school-based [All Fields]) OR adolescen* [All Fields/MeSH Terms]) OR child* [All Fields]) OR youth [All Fields]) AND ((prevent* [All Fields]) OR early intervent* [All Fields])) AND control [All Fields] OR control groups [MeSH Terms] OR control groups [All Fields] AND (Clinical Trial[ptyp] AND English[lang]).
Appendix B

Funnel Plot of Standard Error by Hedges’s g
Highlights

- School-based prevention programs have small effects on depression and anxiety
- Significant prevention effects were detected at 6 and 12 month follow-up
- Prevention type and personnel delivering the prevention program influenced outcomes
- For depression, targeted prevention was more effective than universal prevention
- School-based prevention programs have potential to reduce mental health burden