2014-03-14

A Meta-Analysis of School-Based Depression Prevention Programs for Children and Adolescents

Sherry Lynn Cowen
Brigham Young University - Provo

A Meta-Analysis of School-Based Depression Prevention

Programs for Children and Adolescents

Sherry L. Cowen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Randal D. Day, Chair
Alan J. Hawkins
Richard B. Miller
Susanne Olsen Roper
Terrance D. Olson

School of Family Life

Brigham Young University

March 2014

ABSTRACT

A Meta-Analysis of School-Based Depression Prevention
Programs for Children and Adolescents

Sherry L. Cowen
School of Family Life, BYU
Doctor of Philosophy

School-based depression prevention programs are being implemented in schools across the world in efforts to inoculate children and adolescents from depressive symptoms. This meta-analysis examined 56 manuscripts with a total of 82 studies which focused on school-based programs to determine how they affect depression, anxiety, cognitive skills, self-esteem, coping, and internalizing behaviors. For these specific outcomes, effect sizes ranged from .08 to .25. All combined outcomes yielded a significant effect size of .15. Moderator analyses revealed key differences that identified characteristics of the most effective programs. Targeted programs servicing at-risk students yielded an effect size of .31, while universal programs produced a significant but small effect size of .07. However, the program facilitator seemed to impact the effectiveness of all types of programs. Non-school personnel produced a .39 effect size with targeted samples, and .17 with universal samples, while school personnel produced about one half to a third of the effect.

ACKNOWLEDGEMENTS

I express my special appreciation and thanks to my advisor Dr. Randal Day, because he has been a tremendous mentor to me from day one. I thank him for encouraging me as a novice researcher and for coaching me all along the way, while still supporting me in keeping a healthy balance of being a wife, mother, and PhD student. His advice, strategies, and patience in helping me to see the big picture of writing more effectively have been priceless. I thank my committee members, Dr. Alan Hawkins, Dr. Susanne Olsen Roper, Dr. Richard Miller, and Dr. Terrance Olson for serving as my committee members and for seeing me through this challenging, but rewarding process. Dr. Hawkins was always there to reply to frequent emails, and to discuss and explore the many questions and decisions that arose in meta-analysis.

I give a special thanks to my family. My husband provided constant encouragement and listening ears for all my exciting and disappointing findings and moments along this incredible journey. He should be earning a PhD in tech support for hours of help at the computer as I learned statistical programs and created professional graphs, charts, and tables. Thank you for backing up my flash drive regularly to several different locations, all for my peace of mind…and probably yours too. And I sincerely acknowledge Andrew, Becky, and Rachel, who provided much of my motivation to learn all I could about marriage, families, parenting, adolescents, well-being, and flourishing. I have watched them grow and flourish over these past four years and their lives truly inspire me.

Table of Contents

List of Tables

List of Figures

Introduction

Depression incidence and prevention programs designed to respond to depression in young people are becoming an urgent topic in child and adolescent research. Researchers estimate that one-third of young people experience challenges in psychological functioning to the extent that it impedes academic and psychosocial development (Cunningham, Brandon, & Frydenberg, 2002). In fact, depression is the leading cause of disability in the U.S. among Americans ages 15 to 44 (National Institute of Mental Health, 2013). In response to increasing reports of depression in young people, researchers and mental health professionals have been designing and implementing prevention programs for children and adolescents aimed at diminishing the severity of depressive symptoms. These programs typically focus on strategies that center on challenging and correcting habitual cognitive tendencies that promote inaccurate thoughts and negative beliefs (Cicchetti & Curtis, 2006).

There are three key goals of the current meta-analysis. First, an assessment of overall program effectiveness will be presented which includes an identification of successful programmatic and methodological strategies. Second, this analysis seeks to understand which programs have the greatest impact on primary and secondary depressive symptoms. And, finally, this analysis helps pinpoint which population subgroups can be most helped by the different kinds of depression prevention programs.

In the past ten years, eight published meta-analyses explored the overall effectiveness of depression prevention programs. In the following section I will discuss what the previous meta-analyses were missing and how this analysis fills those gaps. For example, in most of the meta-analyses, secondary outcomes were not previously analyzed for effect sizes, even though most depression prevention programs analyze some combination of depression, anxiety, self-esteem,

cognitive attributions, coping, and internalizing behavior outcomes; however one analyzed depression and anxiety, but only examined 28 studies. Furthermore, researchers used a variety of depression measures in the interventions, but previous analyses did not analyze or compare effect sizes, or did not have statistical power to do so, for those various measures. Lastly, previous meta-analyses were not solely focused on school-based programs, except for Corrieri and colleagues (2013) who analyzed 28 studies, even though K-12 schools provide the most feasible way to reach the largest population of children and adolescents

As mentioned, none of the previous meta-analyses examined the direct influence of interventions on the full scope of key secondary outcomes which are also contributing factors of depression (Brunwasser, Gillham, & Kim, 2009; Corrieri, Heider, Conrad, Blume, Konig, & Riedel-Heller, 2013; Cuijpers, van Straten, Smit, Mihalopoulos, & Beekman, 2008; Horowitz & Garber, 2006; Jane-Llopis, Hosman, Jenkins, & Anderson, 2003; Merry, Hetrick, Cox, Brudevold-Iversen, Bir, & McDowell, 2011; Merry, McDowell, Hetrick, Bir, & Muller, 2004; Stice, Shaw, Bohon, Marti, & Rohde, 2009). While these research efforts have furthered our understanding of depression prevention programming, it was found that most programs not only focus on responding to depression symptoms, but also commonly address and measure secondary, depression-related outcomes such as coping, self-esteem, anxiety, internalizing behavior, and cognitive attributions. While many of these constructs can be associated with the course of depression, vital information is lost when an analysis collapses measures, or does not utilize all the measures, and researchers do not attend to finer grained differences that are provided within the target studies (Brunwasser et al., 2009; Gotlib & Hammen, 2009; Horowitz & Garber, 2006). In other words, meta-analytic researchers usually did not use all the measured outcomes, or they combined measured outcomes, and analyzed the effects of programs using a

narrower and limited construct of depressive symptoms. The current meta-analysis examines depression program effectiveness on overall, combined depressive symptoms, but also includes separate analyses of secondary, depression-related outcomes such as coping, self-esteem, anxiety, internalizing behaviors, and cognitive attributions. Furthermore, the actual measurement of depressive symptoms is often confusing. Researchers have used at least fifteen different measures to assess the actual depressive symptomology, not including coping, self-esteem, anxiety, and cognitive attributions measures (see Table 1). Yet it is unclear, or left unstated, why a specific measure is selected over another in each depression prevention program. This meta-analysis will examine individual effect sizes for the most commonly used depression measures that could apparently yield important differences. Depression rating scales may be problematic in that the data generated can be extremely variable, with standard deviations as large as the reported means in some studies (Merry et al., 2004). This information could guide future depression measurement selection.

For example, researchers have indicated that there is considerable heterogeneity among item content in some of the most widely used measures (Santor, Gregus, & Welch, 2006; Snaith, 1993). The commonly used Beck Depression Inventory (BDI, Beck, Steer, & Brown, 1996; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) contains a higher proportion of cognitive items, whereas the Hamilton Rating Scale for Depression (HRSD, Hamilton, 1960) contains a higher number of somatic items, and the Center for Epidemiologic Studies Depression Scales (CES-D, Radloff, 1977) contain more items that are not specific to depressive disorders, such as perceptions of others, talkativeness, or comparisons with others (Snaith, 1993). The majority of the depression prevention programs use content based on cognitive-behavioral, or social-cognitive intervention strategies, so careful depression measure selection is a potentially

significant element in program effectiveness. Since previous meta-analyses have not analyzed the differences in program effectiveness according to type of depression assessment, this dissertation fills that gap.

Another deficit in previous meta-analyses is that they did not solely focus on school-based depression prevention programs; sometimes they mixed in college samples, studies based in medical centers, mental health facilities, and community centers, thus camouflaging the actual impact of school-based programs. Yet, K-12 school-based depression prevention programs are essential because they provide the opportunity to reach the greatest amount of children and adolescents (Quayle, Dzuirawiec, Roberts, Kane, & Ebsworthy, 2001). The school setting provides the potential for coordinating with families, teachers, and community providers in order to maximize effectiveness, and thus should be a central focus of intervention (Greenberg, Domitrovich, & Bumbarger, 2001). K-12 schools also provide the environment to implement universal and targeted programs, depending on the need. It is critical to examine how well the school-based programs are doing so that their effectiveness can be verified and so that improvements can be made in order to deliver successful programs to large populations. In this study, programs implemented within the K-12 framework will be examined, to see if the elementary through high school settings are a feasible and effective means of program delivery.

Background

Over several years, researchers have been interested in assessing programs that respond to adolescent depression in school-based settings. In the following section, a number of key terms, constructs, and measures will be introduced. Then, depression will be defined from a clinical perspective, and an explanation of why it is so important to address depressive symptoms in childhood and adolescence will be explored. Next, the theory and prevention science behind

depression prevention programs will be examined and why school-based programs could be ideal for implementing such programs. Last, program content will be described as well as the overall goals of depression prevention programs.

Depression

  Defining terms. First, it is important to understand how depression terminology is used in this paper. In depression prevention programs, researchers use depression measures, such as the BDI, CDI, and CES-D, and these assess actual depressive symptoms. Researchers also use coping, anxiety, cognitive attributions, self-esteem, and internalizing measures to assess these secondary depression-related symptoms. When all of these depression, coping, anxiety, cognitive attributions, self-esteem, and internalizing measures are combined, then overall depressive symptoms are measured. This can be confused with the depressive symptoms measured only by the specific depression measures (BDI, CDI, etc.). Thus, depressive symptoms can be measured by the depression measures alone, or overall depressive symptoms, which refer to combining depression measures with other secondary depression-related measures. So, the term overall depression, or overall depressive symptoms, refers to all the depression measures and secondary depression-related measures combined (See Table 1).

  Defining depression. According to the Diagnostic and Statistical Manual of Mental Disorders: DSM-V (2013), depression is a psychiatric disorder based on the number, severity, persistence, and impairment of symptoms (Kazdin, 1995). Those symptoms include depressed mood (feeling sad, lonely, or empty) or irritability, decreased interest or pleasure in most activities that were once enjoyable, significant weight change or change in appetite, change in sleep, change in activity, fatigue or loss of energy, feelings of guilt or worthlessness, diminished ability to concentrate, anxiety (irrational worry, fears, feeling tense, trouble relaxing, etc.), and

suicidality. The DSM-V (2013) states that an individual who has five to nine of those symptoms

for more than two weeks, nearly every day, combined with impaired function in a social,

occupational, or educational domain could indicate major depressive disorder. Someone who

feels depressed for more than two weeks and has only 1-2 other symptoms, plus mild functional

impairment, could be classified as mildly depressed. So, the number of depressive symptoms

and degrees of severity, persistence, and impairment determine the levels of mild, moderate, or

severe depressive episodes. When children and adolescents have multiple depressive symptoms

and if they persist for more than two weeks, they generally require a formal evaluation and

diagnosis to see if clinical depression is indicated (NIMH, 2013).

     Adolescent depressive symptoms. Developmentally, young people are particularly

susceptible to depressive symptoms because of the increasing prevalence of stressful events

during the adolescent years (Steinberg, 2008).   The majority of individuals who develop

depressive symptoms experience their first significant episode during the transition from early to

late adolescence (15-18), during which there is a considerable six fold increase in the prevalence

of depression (Hankin, Abramson, Moffitt, Silva, McGee, & Angell, 1998; Kessler, Avenei, &

Merikangas, 2001; Nolen-Hoeksema & Hilt, 2009) that steadily increases during adolescence

and then starts to decline after late adolescence (Wight, Sepulveda, & Anashensel, 2004). When

depressive symptoms strike early in adolescence, they are likely to be chronic and relapse over

the life span, with reported recurrence rates ranging from 40 to 72 percent over three to seven

years (Fombonne, Wostear, Cooper, Harrington, & Rutter, 2001; Kovacs, Feinberg, Crouse-

Novak, Paulauskas, Pollock, & Finkelstein, 1984; Nolen-Hoeksema & Hilt, 2009). As such,

adolescent-onset depressive disorders often times foreshadow long-term conditions because of

their strong association with recurrent emotional problems in adulthood.

Consequences. The consequences of child and adolescent depressive symptomology can be devastating. One of the most shattering ramifications of depressive symptoms is its ability to affect every aspect of a person's life. For example, depressed adolescents tend to have poorer performance in work, school, sports, and socially they can become withdrawn or unable to maintain friendships; it affects them in their families, sometimes leaving them unable to engage in meaningful activities and relationships (Rudolph, Flynn, & Abaied, 2008). Furthermore, comorbidity is the norm among individuals with depression; it often occurs in combination with anxiety, poorer physical health, substance abuse, eating disorders, and externalizing behaviors (Gotlib & Goodman, 1999; Weissman, et al., 2005). As children, teens, or adults, depressed individuals tend to act in ways that elicit negative responses and generate stress and conflict in their relationships (Hammen, 2006; Joiner, Coyne & Blalock, 1999; Nolen-Hoeksema & Hilt, 2009), thereby perpetuating this downward depressive cycle. Consequently, children and adolescents exhibiting depressive symptoms tend to grow up to be depressed adults (Kim-Cohen, Caspi, & Moffitt, 2003). Thus, there is a need for children and adolescents to somehow inoculate themselves in preparation for these potentially arduous times, especially among those who are developmentally vulnerable.

Depression Prevention Programs

Theory. Given the pervasive consequences of depression, researchers have theorized and explored ways to lessen depressive symptoms. Theoretically, the positive psychology movement advocates that individuals have opportunities to increase or decrease depressive symptoms through what they do in their daily lives and through how they think (Sheldon & Lyubomirsky, 2006). Research conducted with identical and fraternal twins suggests that about fifty percent of the differences in people's depression or happiness levels can be accounted for

by their genetically determined set points, or in other words, their genetically determined temperament (Lykken & Tellegen, 1996; Sheldon & Lyubomirsky, 2006; Schulman, Keith, & Seligman; 1993). Only about ten percent of the variance in depression levels is explained by individual life circumstances or situations, such as income, health, or marital status, whether positive or negative (Diener, Horwitz, & Emmons, 1985; Sheldon & Lyubomirsky, 2006). It is believed that the remaining forty percent of depressive symptomology may be within one's ability to change and improve; this forty percent provides the room for opportunities to increase or decrease depressive symptoms through what individuals act upon in their daily lives and through how they think (Sheldon & Lyubomirsky, 2006). Hence, within the framework of positive psychology, the key to altering depressive symptoms lies not in changing genetic make-up, and not in changing individuals' unique personal circumstances, but rather in working with the forty percent that represents individuals' daily intentional activities and thoughts. This is a critical piece in the puzzle of depression and provides a possible way to help young people protect themselves from the growing epidemic of depression.

School-based depression prevention programs. School-based depression prevention programs are designed to help children and adolescents with that forty percent of intentional thoughts and behaviors that may make a difference in lessening depressive symptoms. With the increasing prevalence of mental health issues facing our young people, researchers have argued for health promotion programs that will reduce overall risk in universal and targeted populations, referring to the general population and at-risk student populations respectively (Cunningham et al., 2002; Rosenman, 1998). Some suggest that such programs should be implemented before adolescence as preventive programs in order to offset the increases in psychological problems after puberty (Cunningham et al., 2002; Roberts, 1999). Seligman (2007) suggests that

children's attributions and responses to life events and trials tend to be solidified and habitual by about nine years of age unless those thoughts are challenged, suggesting that elementary school would be an optimal time to begin implementing prevention programs. Schools are one of the most promising sites for delivery of programs, and universal implementation has much to offer. The majority of universal programs, and many of the targeted ones, are delivered in schools. Although there are some practical issues around adding anything extra into busy school schedules, schools provide the opportunity for depression prevention programs to reach the greatest amount of children and adolescents (Quayle et al., 2001; Merry et al., 2011).

Prevention science. There are three main approaches to depression prevention: universal, indicated, and selected, all of which aim to strengthen protective factors such as coping, self-esteem, cognitive skills, and problem solving, and which will hopefully help decrease anxiety, internalizing behaviors, and depressive symptoms. These prevention methods have advantages and disadvantages. Universal approaches are designed to reach a generally large audience, with varied levels of risk. The advantages to universal programs include avoiding labeling effects associated with being selected as at-risk, better participation rates, ability to deliver doses of the intervention over longer periods of time, and the ability to target a wide range of risk factors simultaneously (Shochet & Ham, 2004; Spence, 2008). Additionally, the public health approach of universal strategies accepts that a relatively small reduction in a risk factor with an entire population may produce a greater reduction in depressive symptoms than a large reduction for a relatively small number of individuals with initially high levels of depressive symptoms (Mason, Scott, Chapman, & Shihfen, 2000; Spence, 2008). Selective approaches address subgroups of the general population who are predicted to be at-risk, perhaps because of family instability, or their parents have suffered with depression. Indicated

prevention is aimed at those participants who are already showing early signs of depressive symptoms. These individuals have not yet met thresholds for clinical depression, but are experiencing the symptoms. The advantages of both indicated and selective approaches, sometimes collectively referred to as targeted approaches, are similar in that they can provide more in-depth and individualized interventions, often resulting in greater individual gains and improvement (Spence, 2008).

        Symptom assessment in school setting. Diagnosing depression usually requires a licensed clinician and involves a combination of self-report questionnaires, diagnostic interviews, and input from significant others. However, in a school setting, self-report questionnaires are often used alone to identify those who might show the beginnings of elevated symptoms, those who might have a probability of manifesting depressive symptoms, or those in need of professional evaluation (Clark, Hawkins, Murphy, Sheeber, Lewinsohn, & Seeley, 1995). Self-report questionnaires allow researchers to assess the number, severity, persistence, and impairment of symptoms a child or adolescent is experiencing. Assessing the severity of depressive symptomology is the most common goal in clinical and research settings because these evaluations are crucial when evaluating treatment success (Nezu, Nezu, Friedman, & Lee, 2009). In non-clinical settings, the self-report questionnaires serve as depression screening measures and are not used to diagnose the condition, rather they provide an indication of the level of symptoms for a given time period (Sharp & Lipsky, 2002). If depressive symptoms reach levels as described in the DSM-V, then those children and adolescents would be referred to professional clinicians. This step cannot be underestimated because only about 25% of depressed adolescents ever receive treatment, and this allows for earlier detection (Hirshfeld et al., 1997). Otherwise, the first assessment serves as a baseline of symptoms that can be examined as a pre-

test and then later compared to a post intervention assessment, thus evaluating the effectiveness of the program when compared to a control group. Researchers hope for post and follow-up reports of fewer, less severe, and less persistent depressive symptoms than before the program. They hope for improved cognitive attributions, anxiety, coping, self-esteem, and internalizing behaviors.

Program content and objectives. For the past several years, schools have tried to integrate various types of wellness programs into their curriculum as part of the school day to reduce depressive symptoms. For example, the University of Oregon has developed and implemented the Oregon Resiliency Project (ORP) throughout many K-12 school systems. This project features research, training, and outreach efforts aimed at social and emotional learning, mental health promotion, and social-emotional assessment. Known as the Strong Start, Strong Kids, and Strong Teens programs, their goals and objectives are to teach social and emotional skills, promote resilience, strengthen assets, and increase coping skills of children and adolescents (Oregon Resiliency Project, 2010).

Similarly, the University of Pennsylvania has developed the Penn Resiliency Program (PRP) with the purpose of reducing depressive symptoms. The PRP curriculum (Positive Psychology Center, 2007) teaches cognitive-behavioral and social problem-solving skills and is based in part on cognitive-behavioral theories of depression by Aaron Beck, Albert Ellis, and Martin Seligman (Abramson, Seligman, & Teasdale, 1978; Beck, 1976; Ellis, 1962). Central to PRP is Ellis' Adversity-Beliefs-Consequences (ABC) model, demonstrating the idea that our beliefs about events impact our emotions and behavior. The program goals and objectives are for students to learn to detect inaccurate thoughts, to evaluate the accuracy of those thoughts, and to challenge negative beliefs by considering alternative interpretations. PRP also teaches a

variety of strategies that can be used for solving problems and coping with difficult situations and emotions. Students learn techniques for assertiveness, negotiation, decision-making, social problem-solving, and relaxation. The skills taught in the program can be applied to many contexts of life, including relationships with peers and family members as well as achievement in academics, employment or other activities (Positive Psychology Center, 2007).

Overall program goals. Depression prevention programs aim to lessen depressive symptoms and to increase other protective factors and skills such as coping, self-esteem, and cognitive attributions. In universal populations, the main goal is to inoculate students from depressive symptomology. In targeted populations not only do researchers hope to prevent the onset of depressive symptoms, but they also hope to strengthen other skills that will help lessen the depressive symptoms that may already exist. This analysis will focus on the change in depressive symptoms because there were not enough school-based studies that reported on preventing the onset of depression among intervention participants relative to control participants.

Typically researchers want to know how a program is affecting critical outcomes such as depressive symptoms, coping, anxiety, self-esteem, cognitive thoughts, and internalizing behaviors, which can all be risk factors or protective factors for depressive symptomology (Abela & Hankin, 2009). For example, high levels of self-esteem can buffer individuals against experiencing increases in depressive symptoms following negative events, even if they possess cognitive vulnerabilities (Abela & Skitch, 2007). Additionally, cognitive vulnerabilities, such as a negative attributional style, seem to interact with negative events in life and predict increases in depressive symptoms in children and adolescents (Abela & Hankin, 2009). Researchers have found that a negative attributional style is positively associated with a maladaptive, emotion-

focused coping style, both of which are consistent predictors of increased depressive symptoms (Bruder-Mattson & Hovanitz, 1990; Cunningham et al., 2002). Hains and Ellmann (1994) found that their stress inoculation training program significantly lowered depressive symptoms in children with high emotional arousal, and depression prevention programs have been found to reduce both anxiety and depressive symptoms (Gillham et al., 2006; Roberts, Kane, Thomson, Bishop, & Hart, 2003). Programs are deemed most effective when depressive symptomology and anxiety decrease and coping, self-esteem, and cognitive skills increase.

Moderators of Intervention Effects

In school-based depression prevention programs, there are several variables that may differ from study to study and may have a significant impact on the effectiveness or ineffectiveness of a program. One of the strengths of a meta-analysis is the ability to examine methodological, participant, and programmatic moderators which help identify specific features that are associated with stronger effects (Lipsey & Wilson, 2001). When the effect size varies substantially from study to study an important goal of the meta-analysis could be to understand the reason for this variation. For example, as we try to understand if subgroups of individuals benefit more from the program, we can ask, "Is the treatment more effective for at-risk students than for students with no apparent risk?" Or to better explain differences in content and instruction we may ask, "Is a program that uses the PRP curriculum a more effective intervention than those that do not?" By doing this, information is obtained that can potentially indicate more effective ways to create and deliver depression prevention programs in the future. In the current study, several potential participant, programmatic, and methodological moderators will be examined that are based on literature reviews, previous meta-analytic findings, and theory.

Methodological moderators. Study design and depression measure selection are both methodological moderators that may reveal differences between more or less effective programs. The primary interest of this meta-analysis is the efficacy of depression prevention programs, which is addressed by effect sizes representing the difference between intervention and no intervention. Thus, only studies that used control groups were included. This study examines both experimental and quasi-experimental studies because quasi-experimental studies are common in school-based programs. Often times the classroom or school becomes the unit of randomization, which is not a truly randomized experiment. Additionally, depression prevention programs utilize a combination of about fifteen different measures to assess changes in actual depressive symptoms, so in this meta-analysis, the most frequently used depression measures will be evaluated as moderators, to determine if effect sizes are independent of the instruments used to measure outcomes.

Participant moderators. Participant moderators are characteristics about the subjects that could reveal subgroups of individuals for whom the program has a stronger or weaker effect. For example, examining sample participant school level may reveal that younger children and younger adolescents have a more difficult time understanding and incorporating new skills. Previous meta-analyses support the notion that larger effects emerge for older adolescents (Brunwasser et al., 2009; Horowitz & Garber, 2006; Stice et al., 2009). However, some studies have shown greater effects for primary school children than high school students, theorizing that earlier preventive intervention was perhaps more advantageous than later intervention (Lock & Barrett, 2003). An unexplored participant moderator is sample location, but now there are a sufficient amount of studies from North America, Europe, Australia, and a few from Asia, which

allow for moderator analyses that could possibly indicate whether programs have stronger effects by region.

Certain risk factors may also serve as important moderators. Participant ethnicity may moderate the effectiveness of depression prevention programs because ethnic minority youth have reported more depressive symptoms than White youth in the past (Cuffe, Waller, Cuccaro, & Pumariega, 1995; Siegel, Aneshensel, Taub, Cantwell, & Driscoll, 1998). Generally, studies have shown that depression prevention programs often have a significantly greater effect on high-risk participants, such as targeted samples, than universal samples that include all individuals in a population (Horowitz et al., 2006; Stice et al., 2009), possibly because at-risk youth have a greater opportunity to show symptom reduction (Stice & Shaw, 2004). However, one meta-analysis on preventing the onset of depressive disorders (Cuijpers et al., 2008) revealed no differences between targeted and universal populations. At-risk samples in depression prevention studies may include individuals with elevated depressive symptoms, including children and adolescents who have a negative attributional style, or youth whose parents suffer from depressive symptoms, or who experience family instability.

Programmatic moderators. Programmatic moderators help us identify the program components under which optimal prevention effects occur, and the length and span of the program could yield significant differences. For example, the length of treatment, in hours, and the span of the treatment, over how many weeks it lasted, could be very important. It seems that a longer program would allow for greater opportunity to present material, practice new skills, and allow for attitudinal and behavioral change, and yet not be so long that young people lose interest or even drop out. Stice and colleagues (2009) found that programs with a shorter duration produced larger effect sizes, and yet Jane-Llopis and colleagues (2003) found that

programs with more than eight sessions that were each 60-90 minutes long had the larger effect sizes.

Similarly, there are other types of program components that could also generate differences in effects. For example, the program facilitator, or the person who delivers the content and instruction, could have an impact on how effective the program is. Whether the facilitators are trained research team members and professionals or school staff, such as teachers or counselors, could make a difference in the success of the program. Some studies are beginning to show that depression prevention programs can be effective when delivered by school personnel; this provides greater ease and lower cost of delivery, which might hold promise for wide-scale dissemination (Arnarson & Craighead, 2011; Gillham et al., 2012).

Another possibly relevant programmatic moderator is whether a program includes homework or parental involvement components. Programs that include homework or parental involvement components that reinforce and practice what was taught in class could produce larger effects than programs without them. One meta-analytic review suggested that those programs with homework assignments produced significantly larger effects than those without (Stice et al., 2009). Lastly, the types of curricula used in depression prevention programs could moderate effects. The most commonly used program is the Penn Resiliency Program (PRP), and so the curriculum will be an additional moderator to evaluate any significant differences between PRP and all others.

Summary

The current meta-analysis will make a contribution to the field in several ways. First, this meta-analysis will extend the reviews by including a greater number of depression prevention studies up to the most recent work, therefore giving the results greater statistical power for group

comparisons and moderator effects. Second, it will analyze effect sizes for the secondary

outcomes such as anxiety, self-esteem, coping, cognitive attributions, and internalizing behavior

which all contribute to the course and outcome of depressive symptoms. Third, it will analyze

effect size differences according to which depression measure was used. Fourth, it will examine

a full range of school-aged children and adolescents from elementary school through high

school, to determine whether school-based programs are effective methods of delivery to

decrease depressive symptoms. As a result, this study of the effectiveness of depression

prevention programs for children and adolescents should make a significant contribution to the

existing literature and suggest future directions improving school-based program

implementation. In summary, this study addresses the following general research question:

- Do school-based depression prevention programs significantly decrease overall
  depressive symptoms?

Specifically, it will also address the following:

1. Do school-based depression prevention programs impact some secondary
   depression-related outcomes more than others?

2. Do specific methodological features (i.e., random assignment, depression
   measure choice) moderate the effect sizes?

3. Do specific participant features (i.e., risk status, ethnicity) moderate the effect
   sizes?

4. Do specific programmatic features (i.e., program duration, program
   curriculum, program facilitator, homework/parental component) moderate the
   effect sizes?

Method

Search Procedures

Studies were identified from the 1980s, when serious depression prevention evaluation research began, through 2013. Four methods of obtaining relevant studies were used in this meta-analysis. First, a search was conducted for previously published depression prevention and well-being meta-analyses, and the reference lists for reports were examined. Second, to find more recent articles, academic databases were searched using the keywords *depression, school-based, intervention, mental health programs, well-being education, depression prevention,* and *resilience.* PsycInfo, PsycExtra, Medline, and Dissertation and Theses databases were used in the search process. Third, the reference list of each included report was examined. Fourth, researchers active in depression prevention were contacted over the past year to find any missing or unpublished reports. Unpublished materials that were found relevant were included in the list of reports selected for further screening. A specific search for studies in non-English languages was not conducted, but three were seen in the online search procedures.

Inclusion Criteria

Studies initially identified were further screened using inclusion and exclusion criteria. Studies were included for the final analysis if they met the following criteria: (a) an assessment of a depression prevention program was conducted; (b) if it did not evaluate a specific depression prevention program, but one of its primary or secondary goals was to prevent depression; (c) quantitative pre- and post-program outcome evaluation scores were reported as a treatment and control-group design; (d) the study targeted a school-based, K-12 program, (e) the study measured at least one depression outcome, (f) if attrition rates exceeded 33 percent, then that

particular assessment would be excluded, and (g) programs that targeted extreme circumstances were excluded.

The online database and manual searches yielded 230 possible manuscripts that matched the purpose of this meta-analysis. Upon closer evaluation, 40 were not empirical studies, 28 were not program evaluation studies, 48 were not targeting depression, 32 had unusable data, 24 were not school-based, and 2 included extreme circumstances such as orphans with AIDS in an under-developed country. These exclusions resulted in 56 manuscripts that matched the inclusion criteria and were used in this analysis. Of these 56 manuscripts, twelve of them contained 2 studies, four contained four studies, and one of them contained 3 studies, for a total of 82 studies (See Figure 1).

Both experimental (randomized controlled trials) and quasi-experimental (i.e., non-randomized groups) studies were included. Those studies that used classrooms or schools as the unit of randomization were coded as quasi-experimental. Those that randomly assigned students to treatment and control groups were coded as experimental designs. All the studies included a comparison group, most of which conducted classroom lessons as usual. Both published and unpublished studies were included to minimize the possibility of publication bias. It is possible that studies with non-significant results are less likely to be submitted or accepted for publication, thus upwardly biasing effect sizes, although Stice and colleagues (2009) did not find this to be a significant difference in their meta-analysis.

Coding

A 30-item codebook was created to code for all studies. Outcome measures, study identification variables, study moderators, and effect size coding information were recorded. Data were independently coded by two coders, the author and another graduate student, and they compared answers. When there were questions about a particular study, coders consulted together. Thus, inter-rater reliability was not calculated since all coding disparities were resolved through discussion and further examination of the study.

The studies examined a wide array of depression-related measures. After analyzing dozens of different measures that were used to study various depression-related outcomes in the intervention programs, and with the help of a leading researcher in depression prevention, I divided all of those measures into six main outcomes: depression, anxiety, self-esteem, coping, cognitive and attributional style, and internalizing behaviors. This was a new addition to meta-analytic work with depression prevention programs, and it was a necessary step that could lead to a deeper understanding of how these programs affect a multitude of depression-related outcomes. This provided the ability to examine the program's impact on all of the combined outcomes, or combinations of outcomes such as internalizing behaviors with depression and anxiety combined, or individual outcomes. Without this list of depression-related outcomes, there would have been one study that measured self-confidence, one that measured self-image, and one that measured self-esteem, for example. However, these measures were closely related, and so it was decided to combine them into one outcome under the construct of self-esteem. The same procedure was used to group the forty-nine different measures into the six main outcomes. Thus, scales that measured depression and hopelessness were categorized under the construct, or outcome, of depression; anxiety measures were labeled as anxiety; coping, solving, and working

with problems were categorized as coping; and the depression, anxiety, and stress measures were combined for an internalizing behaviors outcome. See Table 1 for the specific measures that were categorized under each outcome.

Methodological moderators were examined for differences in effect sizes, including: (a) study design: randomized controlled trial or quasi-experimental design, and (b) depression measure used: BDI, CDI, CES-D, RADS, and other.

Potential moderators of program and participant effectiveness were examined and coded. Possible differences due to six program moderators were explored: (a) length of treatment in total hours: less than 12 hours, greater or equal to 12 hours, (b) span of treatment in total weeks: less than five weeks, five to nine weeks, 10-12 weeks, more than 12 weeks, (c) program name: Penn or other, (d) program facilitator: non-school personnel (such as research team/professionals), school personnel, or a combination of both, (e) program timing: during or after school, (f) homework/parent component: none, homework, or parent. The following four participant moderators were also analyzed to better understand whether certain subgroups of participants benefited more: (a) school level: elementary, middle, or high school, (b) ethnicity: percentage of White participants, (c) subject risk factor: risk or no risk, and (d) location of sample: North America, Australia, Europe, Asia, or other.

Computing and Reporting of Effect Sizes

The effect sizes were analyzed with Comprehensive Meta-Analysis II (Biostat, 2006). The standardized mean gain score difference effect between pre- and post-scores were analyzed to compute the effect size, using reported means and standard deviations. It was advantageous to use the standardized mean gain score difference instead of the simple standardized difference at post test because with quasi-experimental designs, the lack of true randomization can result in

pre-test differences in groups on key variables (Lipsey & Wilson, 2001). The standardized mean gain score takes into account pre-test scores; it creates a standardized change difference score so that any initial group differences do not misconstrue the difference between treatment and control groups. Otherwise it would be possible for the standardized mean difference score to be under or overestimated when considering the true change of quasi-experimental studies because the treatment groups in quasi-experimental studies could potentially have lower or higher scores on key outcome variables than those in the control groups. Calculating this effect size required information about the pre-post correlation of outcome measures to precisely compute the effect size statistic. That correlation was not reported in the studies I examined, so I estimated it to be .50, which has been found to be a robust estimate in previous meta-analytic studies (Nowak & Heinricks, 2008). This effect size statistic describes the difference between program gain scores for control and treatment groups divided by the pooled standard deviation of the two groups. A small effect size is considered to be .2 or smaller, a moderate effect is .3 to .8, and a large effect is greater than .8 (Lipsey & Wilson, 2001). Random effects models were utilized to estimate effect sizes to allow for the possibility that the differences in effect sizes were not only due to participant sampling error but also variations in study and intervention methods (Lipsey & Wilson, 2001). The traditional 95% confidence interval was implemented to determine the significance of effect sizes.

Furthermore, this meta-analysis tested whether there were significant differences, or heterogeneity, amongst participant subgroups and programmatic moderators or whether the variation in findings was compatible with chance alone. As an indicator of homogeneity, the *Q* statistic was calculated. A significant *Q* rejects the null hypothesis of homogeneity and indicates

that the variability between groups is greater than what is likely to have resulted from subject-level sample error alone (Cuijpers et al., 2008).

Plan for Analyses

This meta-analysis examined the overall effect size for depression prevention programs, as well as the effect size for each of the six distinct outcomes. In addition to reporting on multiple outcomes, most studies reported scores for these secondary outcomes at different time points. For studies that utilized a follow-up assessment, the immediate post and follow-up scores were computed as two separate effect sizes. While this strategy might contribute to the variation in the overall effect size, it is important to test for possible deterioration or gain over time. Follow-up scores provide valuable information as they allow for some time for the changes to take place. Some researchers found that the effects of depression prevention programs for universal samples actually strengthened as time passed (Gillham et al., 2006).

One potential concern with follow-up data is attrition bias which could falsely inflate effect sizes. To help with this matter, the acceptable attrition rate was set to 33%, so that any study with more than 33% attrition would be dropped from follow-up effect size analyses in order to minimize bias due to high attrition. This study examined how methodological, participant, and program specific variables moderated the overall effect and individual outcome effects. Each effect's direction was coded accurately, and double checked, such that a positive program effect was coded as a positive effect size, and the same being true for negative effects.

<div align="center">

Results

</div>

Descriptive Statistics

This meta-analytic study included 82 studies from 56 reports. A review of these studies yielded over 550 effect sizes which enabled the examination of the effects of depression

prevention programs on several child and adolescent outcomes. The majority of the studies included predominantly White (k =39), universal programs (k = 52), and about half were from outside of the United States (k = 40). The participants were from K-12 schools, with reported mean ages between eight and 16 years old. There were 13 studies from elementary schools, 38 from middle schools, and 31 from high schools. Thirty-nine of the studies examined evenly mixed male and female samples, while 19 were predominantly female (60% or more) and 10 were predominantly male.

Effects on Depressive Outcomes

Overall depressive symptoms. A primary research question was whether school-based depression prevention programs significantly decreased overall depressive symptoms. According to our data, when all outcomes and data collection times were combined, these programs had an overall significant effect (d = .151, p < .001, k = 82). When analyzed at the various data collection times, the programs had overall significant effects at post-test (d = .174, p < .001, k = 80), one- to six-month follow-ups (d = .169, p < .001, k = 50), and at seven- to 12-month follow-ups (d = .108, p < .001, k = 28). Then the effects were not significant during 18-month to 36-month follow-ups (d = .051, p = .097, k = 13).

Depression-related outcomes. Another primary research question was whether depression prevention programs impacted various depression-related outcomes more than others. It is important to note that the Q and I^2 are values that describe the distribution of effect sizes to determine whether there is significant heterogeneity, and systemic or non-systemic variation that justifies moderator analyses. When analyzing only the depression measures for all time points, there was a significant effect (d = .154, p < .001, k = 80; Q = 148.16***, I^2 = 47.36***). The Q and I^2 are significant for depression and indicate that it makes sense to explore systematic

heterogeneity in further moderator analyses. Similarly, all the other outcomes also had positive and significant effects, including anxiety ($d = .132$, $p < .001$, $k = 26$; $Q = 31.51$, $I^2 = 23.84$), as well as self-esteem ($d = .251$, $p < .001$, $k = 12$; $Q = 10.02$, $I^2 = .00$), cognitive attributional style ($d = .080$, $p < .05$, $k = 30$; $Q = 56.21$**, $I^2 = 48.40$**), internalizing symptoms ($d = .164$, $p < .001$, $k = 80$; $Q = 132.79$***, $I^2 = 41.26$***), and coping ($d = .125$, $p < .01$, $k = 13$; $Q = 32.21$**, $I^2 = 62.74$**; See Table 2). However, since there was no significant heteregeneity (Q and I^2) in the anxiety and self-esteem outcomes, they will not be explored in the following moderator analyses.

Moderator Analyses

In meta-analysis, moderator analyses separate out the effect sizes, or find the differences between them, while getting closer to a true effect size. This can be understood in the context of moderator and mediator differences. For example, moderator analyses are conducted when there is an unexpectedly inconsistent relation between a predictor and an outcome. In this case, depression prevention programs have a lot of variability in effect size. Some programs report a .69 effect size while others report .08. Moderator analyses help to untangle the differences and explain when the differences occur. In contrast, mediation is best used in the case of a strong relation between the predictor and the outcome variable and in the process explains how or why such effects occur (Baron & Kenny, 1986). As an example, we might discover that a non-school facilitator moderates the effect of the program. Thus, when non-school personnel deliver the program, the effect size tends to be stronger. This does not explain why or how the facilitator accomplishes that, which would be a mediating effect, but it does separate out the effect that the facilitator has on the outcome.

The following section provides moderator analyses that were conducted using all time points combined unless specifically stated otherwise. Also, overall effects for all combined outcomes are reported first, followed by effects for specific secondary outcomes when there are significant differences between groups for a particular outcome and a sufficient number of studies to be analyzed.

Methodological features. The next research question was whether any specific methodological features such as experimental or quasi-experimental design, or the choice of depression measure used, moderated the effect sizes (See Table 3).

Experimental v. quasi-experimental. Studies were coded as either a true experimental design or quasi-experimental. In school-based programs, often times researchers randomize treatment and control groups by school or by classroom. In these cases, they were coded as quasi-experimental because there can be systematic differences among participants that remain unrandomized. True experimental designs randomly assigned students to treatment and control groups. Experimental designs had significant effects ($d = .246$, $p < .001$, $k = 44$), as did quasi-experimental designs ($d = .064$, $p < .001$, $k = 38$), but experimental designs seemed to have a stronger impact than quasi-experimental designs, ($Q = 25.378$, $p < .001$), indicating that with these data, the type of randomization did significantly affect the outcome.

Depression measures. Depression prevention programs all utilize a variety of different depression measures to capture depressive symptoms. In this analysis, the CDI was used 45 times, the CESD 16 times, the RADS 11 times, and the BDI 6 times. A wide variety of other measures were also used, but these four were used with the most frequency. In ten studies, a combination of the CDI and one other of these measures were used, in which case I excluded the CDI for this moderator analysis in order to allow the other measures to come through. This did

not significantly alter the effect size of the CDI. All of them had significant effect sizes, BDI (d = .227, $p < .001$, $k = 6$), CDI ($d = .122$, $p < .001$, $k = 45$), and RADS ($d = .214$, $p < .01$, $k = 11$), except the CESD was only trending towards significance ($d = .099$, $p = .058$, $k = 16$). However, no significant difference was found between the groups ($Q = 4.281$, $p = .233$).

Participant features. Another research question was whether specific participant features would moderate the effectiveness of the depression prevention program. School level, ethnicity, risk status, and sample location were examined (See Table 4).

School level. Participants were coded by age and level of school they were in. Programs had overall significant effects on elementary ($d = .129$, $p = < .001$, $k = 13$), middle school ($d = .135$, $p = < .001$, $k = 38$), and high school students ($d = .185$, $p < .001$, $k = 31$). However, differences between them were not significant ($Q = 1.490$, $p = .475$). Further analyses were conducted to see if elementary, middle school, or high school students differed significantly on secondary outcomes, and they did not. They all had significant effect sizes, but they were not statistically different (See Table 4).

Ethnicity. Depression prevention programs had an overall significant effect on both predominantly White ($d = .136$, $p < .001$, $k = 39$) and non-White ($d = .268$, $p < .001$, $k = 18$) samples. The results identified a significant difference between the groups ($Q = 4.849$, $p < .05$). Thus, the programs had a stronger effect, almost double, on predominantly non-White samples. A pattern of stronger effects emerged when analyzing specific outcomes; for depression, both White ($d = .141$, $p < .001$, $k = 39$) and non-White groups ($d = .279$, $p < .001$, $k = 16$) had significant effects and the difference between them was significant ($Q = 3.774$, $p < .05$); for cognitive thinking, non-White groups had significant effects ($d = .233$, $p < .01$, $k = 5$), while White samples did not ($d = .042$, $p = .425$, $k = 10$), and the difference between groups was

significant ($Q = 7.004$, $p < .05$); for internalizing symptoms, both White ($d = .152$, $p < .001$, $k = 39$) and non-White ($d = .275$, $p < .001$, $k = 16$) groups had positive effects, but the difference between groups was only a trend ($Q = 3.082$, $p = .079$).

Targeted v. universal. The analysis for targeted and universal samples showed similar results to those of predominantly White and non-White samples. There was an overall significant effect for both targeted groups ($d = .309$, $p < .001$, $k = 30$), and universal samples ($d = .071$, $p < .001$, $k = 52$), and a significant difference was found between the groups ($Q = 27.683$, $p < .001$). Targeted groups seem to benefit about four times more than universal samples. When analyzed by specific outcomes, this pattern of stronger effects for targeted groups was evident; for depression, both targeted ($d = .346$, $p < .001$, $k = 28$) and universal groups ($d = .089$, $p < .001$, $k = 52$) had significant effects, but effects for targeted groups were significantly stronger ($Q = 25.094$, $p < .001$); for internalizing symptoms, both targeted ($d = .360$, $p < .001$, $k = 28$) and universal ($d = .095$, $p < .001$, $k = 52$) groups had positive effects, but effects for targeted were significantly stronger ($Q = 28.672$, $p < .001$). Also, targeted samples showed improved cognitive thinking ($d = .168$, $p < .01$, $k = 9$), while universal samples did not ($d = .043$, $p = .262$, $k = 21$), although the difference was only a trend ($Q = 3.229$, $p = .072$).

Sample location. The studies came from thirteen different countries. Forty-two came from North America (only one from Canada), 20 from Australia (one New Zealand study was included), 3 from Asian countries, and 15 from Europe (including Germany, Spain, England, Belgium, and the Netherlands). The remaining countries only produced one study from their continent so I was unable to compare them to other regions. Although there were only three studies from Asia, they were kept in this analysis to show emerging trends from that region. Moderator analyses showed that there were overall moderate effects for countries in Asia ($d =$

.429, $p = .053$, $k = 3$), but that effect should be interpreted with caution due to small k. Small effects were found for samples in North America ($d = .153$, $p < .001$, $k = 42$), Australia ($d = .103$, $p < .01$, $k = 20$), and Europe ($d = .160$, $p < .001$, $k = 15$), but the difference between the groups was not significant ($Q = 3.623$, $p = .305$).

Programmatic features. Another research question was whether specific programmatic features would moderate the effectiveness of the depression prevention programs. The length and span of the programs, curriculum, type of facilitator, during or after school timing, and homework or parental involvement additions were examined (See Table 5).

Length and span of program. The studies in this meta-analysis had a span between one and 20 weeks in length. Groups were formed to try and capture differences between shorter, average, and longer spans of time. Programs that were implemented in less than five weeks ($d = .143$, $p < .05$, $k = 8$), between five and nine weeks ($d = .222$, $p < .001$, $k = 17$), ten to twelve weeks ($d = .145$, $p < .001$, $k = 54$), and more than twelve weeks ($d = .079$, $p = .374$, $k = 3$) were all analyzed. The differences were not statistically different ($Q = 2.437$, $p = .487$), and there were only 3 studies in the more than twelve week group. Similarly, the analysis of the length of program time (in hours) did not reveal any significant differences ($Q = 1.720$, $p = .19$) between those programs that were twelve or more hours in length ($d = .135$, $p < .001$, $k = 45$) and those that were less than twelve hours ($d = .188$, $p < .001$, $k = 37$).

Program curriculum. There were 23 studies that reported using some type of Penn prevention program curriculum, and a wide variety of programs that operated under different curricula ($k = 59$) that are referred to as non-Penn programs. For overall effectiveness, the Penn programs had positive effects ($d = .203$, $p < .001$, $k = 23$) as well as the non-Penn programs ($d = .143$, $p < .001$, $k = 59$). Moderator analyses indicated that Penn programs generally yielded

stronger effect sizes, but there was no significant difference between the Penn and non-Penn programs in overall outcomes ($Q = 1.561$, $p = .211$) or individual outcomes.

School v. non-school facilitator. In overall program effectiveness, with all outcomes combined, there were significant differences between programs ($Q = 15.517$, $p < .001$) that were administered by school personnel only ($d = .087$, $p < .001$, $k = 24$), non-school personnel only ($d = .244$, $p < .001$, $k = 38$), or a combination of both ($d = .114$, $p < .01$, $k = 20$). Further tests of specific differences showed that the differences were significant between school and non-school personnel ($Q = 15.284$, $p < .001$). Programs that were facilitated by non-school personnel, such as graduate students, research team members, or psychologists, seemed to yield significantly higher effect sizes not only in overall depression as mentioned, but also in depression, and internalizing symptoms. For depression, there were significant effects with non-school facilitators ($d = .269$, $p < .001$, $k = 36$) and school facilitators ($d = .104$, $p < .001$, $k = 24$) with a significant difference between groups ($Q = 10.769$, $p < .01$), and the internalizing outcome with non-school personnel ($d = .274$, $p < .001$, $k = 36$) and school personnel ($d = .101$, $p < .001$, $k = 24$), with a significant difference between the groups ($Q = 10.956$, $p < .01$). Differences were not significant for cognitive skills or coping.

During v. after school programs. School-based programs are typically held either during school or after school. The analysis indicated that both after school programs ($d = .246$, $p < .001$, $k = 22$) and during-school programs ($d = .119$, $p < .001$, $k = 56$) were effective, but there was a significant difference between these two groups ($Q = 7.943$, $p < .05$). Thus, the after school programs have a significantly greater effect size, more than double that of the during school programs.

Homework and parental components. Moderator analyses were conducted to determine if adding a homework or parental involvement component into the programs would yield greater benefits. According to our samples, there are significant effects for both programs with the added homework component ($d = .183$, $p < .001$, $k = 27$) and the parental involvement component ($d = .146$, $p < .001$, $k = 10$), as well as for those studies with no homework or parental component ($d = .134$, $p < .001$, $k = 37$), and the difference between them was not significant ($Q = 1.905$, $p = .386$).

Discussion

Effects on Depressive Symptoms

The primary goal of this meta-analysis was to determine whether school-based depression prevention programs have been effective in reducing overall depressive symptoms (combination of all outcomes). The overall results are similar to previous meta-analytic studies on depression prevention programs that addressed depressive symptoms in youth which reported overall low to moderate effect sizes ranging from .11 to .22 (Brunwasser, et al., 2009; Horowitz & Garber, 2006; Jane-Llopis et al., 2003; Stice et al., 2009) and .16 at post-intervention and .11 at follow-up (Horowitz & Garber, 2006). The data in this analysis yielded an overall effect size of .151, and .174 at post-intervention, .169 at one to six month follow-ups, and .108 at seven to 12 month follow-ups. Overall the programs are having a significant, but small, effect up to a year after the intervention.

Another goal of this meta-analysis included an examination of secondary depression-related outcomes, and depression, anxiety, coping, cognitive attributions, self-esteem, and internalizing behaviors all yielded significant effect sizes. There was a gap in the research and a need to examine effects on specific types of symptoms which would help identify where

programs were having their strongest influence and identify symptoms and challenges for which other skills and interventions may be needed (Gillham et al., 2012). The school-based programs appear to be decreasing depressive symptoms (as analyzed by depressive measures), decreasing child and adolescent anxiety, increasing self-esteem, improving cognitive attributional style, decreasing internalizing symptoms, and enhancing the ability of children and teens to cope with their difficulties, with significant effect sizes from .08 to .25. However, there was substantial heterogeneity in the distribution of most of these effect sizes, thus moderator analyses were analyzed to reveal some potential sources of this heterogeneity.

Methodological Moderators

Another objective in this analysis was to determine whether specific methodological features would moderate the effect sizes. Study design was significantly related to the magnitude of the effect sizes. Forty-four studies with an experimental design yielded an effect size of .246, while thirty-eight quasi-experimental programs produced an effect size of .064. This may be confounded by an effect of targeted versus universal samples, because quasi-experimental studies in schools generally assign entire classes or schools to be the unit of randomization, and are therefore indicative of universal studies that tend to yield lower but significant effect sizes. The quasi-experimental studies in this analysis included 27 universal programs and 11 targeted programs, so this could support the theory that we are getting a confounding effect of universal programs. In contrast, experimental designs typically go beyond classroom randomization to target more of the at-risk students, therefore yielding stronger effect sizes, and the majority of the experimental designs in this analysis were targeted programs. Previous meta-analytic work did not find any significant differences between experimental and quasi-experimental designs (Stice et al., 2009), but no other meta-analysis has analyzed only K-

12 schools where there are potentially more studies with quasi-experimental designs to power the analysis.

Participant Moderators

Analyzing participant features was an important purpose in this study and they revealed some expected and some new findings. As shown in Table 2, there was some heterogeneity to explore in most of the outcomes, so moderator analyses were conducted to help explain some of that variance. I expected the targeted samples to have greater effect sizes than universal, as well as the predominantly non-White samples to have greater effect sizes than the predominantly White, and this analysis confirmed these expectations. Previous meta-analyses have found similar results (Horowitz & Garber, 2006; Stice et al., 2009; Merry et al., 2011). Studies have shown that ethnic minorities tend to report higher levels of depressive symptoms which could possibly be explained by lower socio-economic status, anticipation of possible unfair treatment, higher levels of promoted mistrust, and socialization barriers (Liu & Lau, 2013). These possible reasons imply that being a minority can be a risk factor at times. As such, ethnic minorities tend to benefit with stronger effects similar to the targeted, at-risk populations. However, it is critical not to overlook that the universal samples and predominantly White samples generated significant and ongoing effect sizes too, up to the twelve month follow-up. There were significant effect sizes at post-intervention for both targeted ($d = .35$, p $<.001$, $k = 29$) and universal studies ($d = .12$, p $<.001$, $k = 51$), at one- to six- month follow-up for targeted ($d = .40$, p $<.001$, $k = 17$) and universal studies ($d = .11$, p $<.001$, $k = 33$), and at 7-12 month follow-up for targeted ($d = .24$, p $<.001$, $k = 5$) and universal studies ($d = .09$, p $<.01$, $k = 23$). Some have speculated as to whether or not universal programs were effective (Spence, 2008). These data

confirm that school-based universal programs are having a positive effect, even though the effect is not as strong as the targeted samples which reached a moderate overall effect size of .40.

This was the first meta-analysis to examine differences between countries or regions, namely North America, Australia, Europe, and Asia. The studies from China, Hong Kong, and Korea were grouped together as the Asian studies in order to examine some preliminary effects. The North American region included studies from the U.S. and one from Canada, and the Australia group included one study from New Zealand. Europe included studies from Germany, Spain, England, Belgium, and the Netherlands. Sample sizes surely inhibited the findings for Asia, but I felt they were worth analyzing to capture emerging effects. There were no significant differences between regions, implying that the depression prevention programs are having similar effects across cultural, ethnic, and regional differences. The studies from North America, Australia, and Europe yielded positive overall effects on depression, and studies from North America and Australia also produced positive effects impacting depressive symptoms, and internalizing symptoms.

A main focus of this meta-analysis is the criteria for being a school-based program. This allowed the examination of the effects by school level, which also implies age groups. Previous meta-analyses have suggested that prevention programs would produce larger effects for older adolescents relative to younger adolescents and children (Horowitz & Garber, 2006; Stice et al., 2009). The reasoning behind this notion was that risk for depression increases during adolescence and that perhaps older adolescents may better understand and apply the concepts taught in the programs. However, when Horowitz and Garber (2006) removed the two college studies from their analysis, the results were no longer significant. This analysis revealed that there was no significant difference between elementary, middle, and high school outcomes which

supports the notion that depression prevention programs can be effective at all school levels. So, perhaps the original thought that younger students may not be able to understand and apply the concepts does not hold true with these data. Seligman (2007) suggests that children's attributions and responses to life events and trials tend to be solidified and habitual by about nine years of age unless those thoughts are challenged, suggesting that elementary school would be an optimal time to begin implementing prevention programs. This analysis suggests that the programs are equally effective for elementary, middle, and high school-aged students.

Programmatic Moderators

Programmatic characteristics were analyzed to better understand which features seemed to have the greatest impact on children and adolescents. Program span in weeks and length of program in hours did not seem to have a significant impact on which programs were more effective. The differences between groups were not significant. However, programs that were more than twelve weeks in length had smaller effect sizes, and programs that were more than twelve hours long also generated smaller effect sizes, but they were not statistically different. Stice and colleagues (2009) reported that relatively shorter prevention programs produced larger intervention effects than did the longer programs, but this meta-analysis cannot confirm that.

This analysis considered the moderating effects of program timing, before or after school, and homework and parental components in interventions. After school programs had significantly greater effects than during school programs. This is not surprising since many targeted programs tend to be held after school, rather than during the school day as would typically be done for universal programs. After school programs seem to be an effective option for meeting the needs of smaller groups of students. Additionally, more programs are including homework components or added parental involvement features. These data revealed no

significant difference between interventions that did or did not include extra components. The effect size was slightly larger for those with homework components, but the difference between no components, homework, and parental involvement was not significant. This would indicate that homework and parental involvement do not seem to make much of a difference in the effectiveness of programs and the strength of effect sizes. This is surprising since homework entails thinking about the concepts outside of class or at home with family members, and similarly the parental component would add parental awareness and involvement. However, it is difficult to get consistent parental involvement, such as attending an evening meeting or a monthly group session, and to ensure that students thoughtfully complete outside of class assignments. So, perhaps this finding is not so surprising. In contrast, Stice and colleagues (2009) found that programs with homework assignments produced significantly larger effects than those without.

Another feature that differs across studies is the type of curriculum that is implemented. Stice and colleagues (2009) assessed studies by program content by evaluating whether the interventions contained components of cognitive change, behavioral activation, problem solving, and social skills. However, the analysis did not reveal an association between content and effect sizes. In this study, the most frequently used program curriculum in school-based programs, the Penn Resiliency Programs, was compared to the other prevention programs being used. It was the only program used with enough frequency to test against the other programs. The Penn program yielded higher effect sizes overall and on most outcomes, but the differences were not found to be significant. Stice and colleagues (2009) examined program content, and this meta-analysis examined a commonly used program curriculum, and yet no differences were found between content type and program curriculum. These are critical findings because it seems to

imply that it is not the content or the program that produces the greatest impact, but rather the people, the human connection between facilitators and students that makes the biggest difference.

The variation and inconsistencies of effects sizes could not be explained by program content and program curriculum, however, according to the data in this analysis, a critical indicator of effectiveness was determined by who taught and delivered the program, whether by school personnel, such as teachers and school counselors, or non-school personnel, such as graduate students, researchers, and psychologists. These data yielded significant differences between the two types of delivery. Non-school personnel helped produce significantly stronger effect sizes on overall and secondary outcomes when compared to school personnel. When analyzed by risk factor, non-school personnel produced a .39 effect size for targeted samples, and .17 for universal samples, which was approximately three times stronger than school personnel's impact on universal programs. Another meta-analysis produced a similar finding (Stice et al., 2009). This could be due to the motivation, investment, experience, and expertise of non-school personnel, in combination with the stress of time constraints and other demands of teacher and school counselor responsibilities, and we do not know how most schools select their facilitators. Do all third grade teachers get to sit through professional development on how to implement a new mental health program into their already tight schedules? Are teachers asked who feels qualified and who feels passionate about mental health? Is teacher rapport with students taken into consideration? Is the sports coach who teaches health given the task of also implementing this depression prevention program into his curriculum? Are school personnel compensated for their additional responsibilities? There are many facilitator selection criteria to consider when launching a school-based depression prevention program.

Facilitator competence is not a new issue specific to only the field of depression prevention programs. Anytime a large scale social program is launched, facilitator quality is a critical issue. In a communication skills training program, researchers assessed the quality of facilitators and concluded that modest training and practice does not result in facilitator competence, and only 18 of the 32 facilitators reached at least a basic level of competence (Bylund, Brown, di Ciccone, Diamond, Eddington, & Kissane, 2009). Researchers who conducted a school-based cognitive behavioral intervention found that even when school clinicians were trained in evidence-based-practices, the training did not result in the implementation of those practices (Langley, Nadeem, Kataoka, Stein, & Jaycox, 2010). And yet most mental health program analyses, including depression prevention program meta-analyses, conclude with the same statement that program facilitators require greater support, resources, and training (Merry et al., 2011; Stice et al., 2009), but even support and training does not seem to ensure implementation of those practices.

Future Directions

Because time and money are important and limited resources, it may not be practical for schools to implement large scale depression prevention programs, and when it comes to schools, one size does not fit all. For those schools within a 30 mile radius of proactive mental health research teams at universities, they can benefit from the effectiveness of non-school personnel coming in to help implement effective programs. There might even be some schools who have a strong partnership with neighboring community mental health professionals who may want to team up to implement universal or targeted programs in neighborhood schools. But, the majority of schools probably do not have either of these options.

In light of the findings presented above, I offer the following suggestions in how to continue addressing school-based depression prevention efforts. Level one intervention would be for every school to at least screen students during the first month or two of school with a standardized depression measure, similar to the routine scoliosis and vision screenings that currently take place in schools. Those students who score in the clinical range should be referred to mental health professionals, and each school should have a clear and identifiable procedure for facilitating that referral process. This step would address the concern of clinically depressed children and adolescents going undetected. They could receive the professional help and support they need. Level one intervention requires minimal time, effort, and money.

Level two intervention requires schools to take the data from their screening assessment, and look at the subclinical level of students, those who are experiencing depressive symptoms, but not quite at the severity of clinical samples. This is a subgroup who could benefit the most from a prevention program. School personnel can collaborate with each other and with non-school personnel, when available, to determine how to best meet students' needs, but remembering that the facilitator selection is critical to the effectiveness of the program.

School personnel need to brainstorm and determine what resources they have to address the needs of the subclinical group. Based on the studies reviewed here, I recommend school administrators ask the following questions as a way to begin the dialogue about how to provide these much needed programs and services. They need to ask important questions to address who, where, when, and how questions. According to the data in this meta-analysis, the who question seems to be critical. Does this group consist of ten students or 200 students? Does the school have a teacher or counselor who already has some training and experience in mental health, who would be invested and motivated, who already has natural rapport with students?

Does the school have access to someone like that at the district level or in the community?

Could they provide an after school program once or twice a week, or a pull out program once or

twice a week at the elementary level? Could a teacher or counselor be paid to conduct such a

program during one or two of their prep periods each week at the middle and high school levels?

Could one period of a life skills class or a Health 102 class be added where these subclinical

students enroll and receive the support, care, and skills of a depression prevention program?

Maybe non-school personnel could be paid a stipend or part-time salary to come in and deliver

the program once or twice a week if the right facilitator cannot be found within the school.

These are the types of questions that will guide the team in selecting the facilitator and

determining the logistics of how, where, and when to implement the program. The answers will

be unique to each school and district.

Level two requires the time and money equivalent to a part-time employee. To pay a

teacher for a prep period, additional period, or after school period would require between $3,000

and $10,000 a year, depending on the pay scale, and that number could fluctuate depending on

how many students fall into the subclinical category.

Level three intervention would entail examining the depression screening results and

determining whether a universal program would best fit the needs of the school population. This

is the most costly type of program, as it would require the equivalent of one to two full-time

positions at the school, ranging approximately between $40,000 and $75,000 per facilitator each

year, depending on the pay scales. This would be appropriate for schools, districts, and

communities who have extra money, who highly value mental health and its importance in

boosting well-being, student learning, and academic performance.

There are many ways to incorporate universal depression prevention programs, so I recommend that each school team brainstorm how to incorporate it best into their school day. Some ways to do this are through a full-time facilitator who visits classrooms in an elementary setting on a rotating schedule once to twice a week, or a full-time facilitator who goes into every health, social science, or P.E. class at the middle and high school levels on a rotating basis, or offer a life skills/well-being class as part of the school curriculum. But the caution still remains, that people, not programs, are what seems to make the greatest impact, so selecting the right facilitator is crucial. If the right facilitator cannot be found, it would be best to only fulfill level one recommendations while continuing to find the best facilitator.

Given the cost and the effects of universal programs, I would recommend that school personnel and researchers consider level one for all schools, and level two for those schools who can afford it, depending on the population of subclinical students. Level three is the caviar of programs, and its overall small effects may not justify the cost in time, effort, and money. But neither Level 2 or 3 are recommended if motivated, invested, competent, supportive facilitators cannot be ensured.

These level one, two, and three recommendations provide a framework for implementing depression prevention programs in K-12 schools. Researchers have a good idea about what tends to contribute to effective depression prevention programs, and the general effects they will produce. Now a cost/benefit decision must be made based on the needs and resources of each school. Shemilt and colleagues (2012) recognize this and have built a profile of resources to more accurately account for the expenses of each type of depression prevention program. They call for researchers to include formal economic evaluations in their reports to assess and account for the cost and efficiency of programs and alternatives. The estimated costs for level two and

three implementation presented earlier are based on the expected need of part-time or full-time help. Now it is a matter of finding what each school can afford in time, effort, and money to best meet the needs of their student populations. School personnel first need to determine what the mental health need is for their students by assessing them, and that will help determine what they can justify in limited resources to best meet those needs.

Limitations

One methodological limitation from this meta-analysis is whether the participants' risk factors persisted from the time they started in the program until the last day of assessment. While we wanted to account for this factor, we were not able to do so because it was almost always unreported and so we could not draw conclusions as to whether or not the children had transitioned from a stressful environment (e.g., poverty) to more wellness-fostering circumstances.

Another limitation at times was a lack of statistical power for a few of the moderator analyses and secondary outcome effects. For depression, anxiety, and internalizing behavior there were typically enough studies for each group, but fewer studies examined cognitive attributions, self-esteem, and coping, so the under-powered analyses need to be interpreted with caution.

Furthermore, this meta-analysis was enhanced by the dozens of measures that were utilized within studies to report certain outcomes. For example, with the help of experienced researchers in the field of adolescence and depression prevention, we were able to categorize numerous scales that assessed depression, anxiety, coping, cognitive attributions, self-esteem, and internalizing behaviors in order to analyze them separately to see how they were being impacted by the interventions. Nevertheless, we relied on the title of the measure, the author's

description of what the scale would be measuring, the items in the measure, and our experience with the measure to create the categories of outcomes. Perhaps, a more formal analysis could be conducted to see if the various measures load onto the same factors before they are categorized together, thus creating a statistical foundation for the measures in each outcome.

Summary

These data revealed several important highlights that seem to contribute to effective implementation of depression prevention programs in the schools. However, there were a few key components that did not yield significantly different results, but these findings are helpful as well for researchers to consider when designing and implementing programs. For example, in this analysis the programs seemed to be equally effective regardless of length and span of treatment, homework and parental components, school level, geographic location, and selection of depression measures. Nonetheless, this analysis found that when compared with no intervention, depression prevention programs were effective in decreasing overall depressive symptoms, as well as secondary outcomes of depression, anxiety, and internalizing behaviors, and increasing coping, cognitive attributions, and self-esteem, and these effects remained significant up to a year after the program.

Participant moderators yielded significant findings. For example, this analysis yielded data to support both targeted and universal programs, which is important since universal programs are likely to be easier to implement. As expected, effect sizes were small for universal programs, and moderate for targeted programs. They were effective at a universal level, demonstrated by a smaller, significant, effect size ($d = .07$, $p < .001$) but reaching larger samples of students. They were effective in targeted samples ($d = .31$, $p < .001$), with a stronger effect but reaching a smaller number of students. Effects were stronger for predominantly non-White

and targeted samples, but both universal and targeted samples maintained significant effect sizes up through one year of follow-up assessments (See Tables 3 and 4). Elementary, middle, and high school levels were all benefiting; they all had significant effect sizes in overall depression and secondary, depression-related outcomes, which implies that all age groups are gaining positive effects. Therefore, this means that depending on the school administrators' and researchers' goals, they should continue with universal and targeted programs, at all school levels, knowing that they can have a stronger effect on fewer students by targeting at-risk students, or a small but still significant effect on more students by implementing a universal program.

Programmatic moderators produced important elements to consider in program implementation. Curriculum did not seem to make a difference in the effectiveness of depression prevention programs. While Penn programs tended to have larger effect sizes on most outcomes, they were not significantly different from those of non-Penn programs. The difference between Penn and non-Penn may be difficult to fully capture until there are more commonly used programs that can be compared and analyzed. This supports the notion that the strongest effects seem to be from people effects and not program effects. Another novel finding was that the program facilitator seemed to be critical in the success of universal and targeted programs. Non-school personnel, such as research teams and psychologists, helped produce significantly stronger effect sizes on overall and secondary outcomes when compared to school personnel, such as teachers and school counselors. When analyzed by risk factor, non-school personnel produced a .39 effect size for targeted samples, and .17 for universal samples, which was approximately three times stronger than school personnel's impact on universal programs. Thus, for more widely available programs, school personnel will need more time to implement

the program, training, and supervision if they will be the facilitators. This is an important next step for program implementation.

In conclusion, researchers and school administrators decide whether their goal is to address universal or targeted samples, but in that decision, the key factors to consider are the same. There are moderators that schools cannot always control, such as geographic location or ethnicity of the sample, even though predominantly non-White samples yield greater effect sizes. There are moderators that did not seem to account for the differences in program effectiveness, such as school level, length and span of treatment, depression measure selection, program curriculum, and homework/parental components. They each yielded significant results, but were not significantly different within groups. Then, there are moderators that researchers and school administrators can control, such as the type of facilitator, whether that be school personnel or non-school personnel. Non-school personnel had significantly greater success in targeted and universal samples, and with depression, anxiety, and internalizing outcomes. So, this is an area where the field needs to improve. If programs are going to be implemented by school teachers and counselors, we need to analyze where the weaknesses lie and provide more careful selection, time, better training, and compensation to ensure the greatest quality of program delivery. Quasi-experimental designs and after school programs also generated stronger effect sizes than experimental designs and during school programs, however those seemed to overlap with the effects of targeted samples; often times targeted programs are held after school and can use a randomized controlled design more easily than universal groups, which are typically randomized by classroom or school. Merry and colleagues (2011) advocated for identification of the more promising programs, and research to confirm effectiveness in formats suitable for implementation, and perhaps the facilitator is a part of that identification.

It is recommended that school personnel form mental health teams who can discuss the three levels of intervention as it pertains to their school. Level one ensures that students are screened for depressive symptoms so that the clinically depressed do not go undetected. Level two begins the dialogue and implementation of a targeted program that will meet the needs of the subclinical population, and level three begins the efforts for implementing a universal program. The caution remains that perhaps the strongest effects are people effects, and not necessarily program effects, so facilitator selection is an essential component of meeting students' needs. So, first school personnel determine what the mental health needs are for their students by assessing them, then they can determine what they can justify in limited resources of time, effort, and money to best meet those needs.

References

Abela, J. R. Z., & Hankin, B. L. (2009). Cognitive vulnerability to depression in adolescents: A developmental psychopathology perspective. In S. Nolen-Hoeksema & L. M. Hilt (Eds.), *Handbook of depression in adolescents* (pp. 335-376). New York, NY US: Routledge/Taylor & Francis Group.

Abela, J. R. Z., & Skitch, S. A. (2007). Dysfunctional attitudes as a cognitive vulnerability factor for depression in children of affectively-ill parents: A multi-wave longitudinal study. *Behaviour Research and Therapy, 45,* 1127-114.

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*(1), 49-74.

Arnarson, E. O., & Craighead, W. E. (2011). Prevention of depression among Icelandic adolescents: A 12-month follow-up. *Behaviour Research and Therapy, 49*(3), 170-174.

*Balle, M.-F. M. (2010). Efficacy of a brief school-based program for selective prevention of childhood anxiety. *Anxiety, Stress & Coping, 23*(1), 71-85.

*Barrett, P. M., Farrell, L. J., Ollendick, T. H., & Dadds, M. (2006). Long-term outcomes of an Australian universal prevention trial of anxiety and depression symptoms in children and youth: An evaluation of the Friends program. *Journal of Clinical Child and Adolescent Psychology, 35*(3), 403-411.

*Barrett, P., & Turner, C. (2001). Prevention of anxiety symptoms in primary school children: Preliminary results from a universal school-based trial. *British Journal of Clinical Psychology, 40*(4), 399-41.

Beck, A. T. (1963). Thinking and depression: Idiosyncratic content and cognitive distortions. *Archives of General Psychiatry, 9,* 324–333.

Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. Oxford England: International Universities Press.

Beck, A. T., Steer, R. A., &. Brown, G. K. (1996). *Manual for the Beck Depression Inventory II*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–10.

Beck, A. T.,Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

*Berry, K., & Hunt, C. J. (2009). Evaluation of an intervention program for anxious adolescent boys who are bullied at school. *Journal of Adolescent Health, 45*(4), 376-382.

Biostat (2006). *Comprehensive meta-analysis* (Version no. 2.2). Englewood, NJ: Author.

Bruder-Mattson, S. F., & Hovanitz, C. A. (1990). Coping and attributional styles as predictors of depression. *Psychological and Mood Disturbance, 46,* 557-565.

Brunwasser, S. M., Gillham, J. E., & Kim, E. S. (2009). A meta-analytic review of the Penn Resiliency Program's effect on depressive symptoms. *Journal of Consulting and Clinical Psychology, 77*(6), 1042-1054.

*Bursuk, L. I. (1998). *The effects of a school-based cognitive-behavioral intervention program on the depression scores of sixth-grade students: A comparison outcome study.* ProQuest Information & Learning, US.

Bylund, C. L., Brown, R. F., di Ciccone, B. L., Diamond, C., Eddington, J., & Kissane, D. W. (2009). Assessing facilitator competence in a comprehensive communication skills training programme. *Medical Education, 43*(4), 342-349.

*Cardemil, E. V., Reivich, K. J., & Seligman, M. E. P. (2002). The prevention of depressive symptoms in low-income minority middle school students. *Prevention & Treatment, 5*(1): ArtID8.

*Castellanos, N. P. & Conrad (2006). Brief interventions targeting personality risk factors for adolescent substance misuse reduce depression, panic and risk-taking behaviours. *Journal of Mental Health, 15*(6), 645-658.

*Chaplin, T. M., Gillham, J. E., Reivich, K., Elkon, A. G. L., Samuels, B., Freres, D. R., et al. (2006). Depression prevention for early adolescent girls: A pilot study of all girls versus co-ed groups. *The Journal of Early Adolescence, 26*(1), 110-126.

Cicchetti, D., & Curtis, W. J. (2006). The developing brain and neural plasticity: Implications for normality, psychopathology, and resilience. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol 2: Developmental neuroscience (2nd ed.).* (pp. 1-64). Hoboken, NJ US: John Wiley & Sons Inc.

*Clarke, G. N., Hawkins, W., Murphy, M., & Sheeber, L. (1993). School-based primary prevention of depressive symptomatology in adolescents: Findings from two studies. *Journal of Adolescent Research, 8*(2), 183-204.

*Clarke, G. N., Hawkins, W., Murphy, M., Sheeber, L., Lewinsohn, P. M., & Seeley, J. R. (1995). Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: A randomized trial of a group cognitive intervention. *Journal of the American Academy of of Child & Adolescent Psychiatry, 34,* 312-321.

Corrieri, S., Heider, D., Conrad, I., Blume, A., König, H.-H., & Riedel-Heller, S. G. (2013). School-based prevention programs for depression and anxiety in adolescence: A systematic review. *Health Promotion International.* doi:10.1093/heapro/dat001

Cuffe, S. P., Waller, J. L., Cuccaro, M. L., & Pumariega, A. J. (1995). Race and gender differences in the treatment of psychiatric disorders in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 1536–1543.

Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C., & Beekman, A. (2008). Preventing the onset of depressive disorders: A meta-analytic review of psychological interventions. *The American Journal of Psychiatry, 165*(10), 1272-128.

Cunningham, E. G., Brandon, C. M., & Frydenberg, E. (2002). Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. *Anxiety, Stress and Coping, 15,* 369-381.

Department of Health and Human Services. (1991). *Healthy people 200.* Washington, D.C.: Government Printing Office, DHHS Pub. No. (PHS) 91-50212.

Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). (2013). Arlington, VA US: American Psychiatric Publishing, Inc.

Diener, E., Horwitz, J., & Emmons, R. A. (1985). Happiness of the very wealthy. *Social Indicators Research, 16,* 263-274.

*Dobson, K. S., Hopkins, J. A., Fata, L., Scherrer, M., & Allan, L. C. (2010). The prevention of depression and anxiety in a sample of high-risk adolescents: A randomized controlled trial. *Canadian Journal of School Psychology, 25*(4), 291-31.

Ellis, A. (1962). *Reason and emotion in psychotherapy*. Oxford England: Lyle Stuart.

Essau, C. A., & Ollendick, T. H. (2009). Diagnosis and assessment of adolescent depression. In S. Nolen-Hoeksema & L. M. Hilt (Eds.), *Handbook of depression in adolescents* (pp. 33-52). New York, NY US: Routledge/Taylor & Francis Group.

Fombonne, E., Wostear, G., Cooper, V., Harrington, R., & Rutter, M. (2001). The Maudsley
    long-term follow-up of child and adolescent depression. 1. Psychiatric outcomes in
    adulthood. *The British Journal of Psychiatry: The Journal of Mental Science, 179*, 210-
    217.

*Gallegos, J. (2009). *Preventing childhood anxiety and depression: Testing the effectiveness of a
    school-based program in Mexico.* ProQuest Information & Learning, US.

Gans, J. (1990). *America's adolescents: How healthy are they?* Chicago: American Medical
    Association.

*Gilboy, S. (2005). Students' optimistic attitudes and resiliency program: Empirical validation of
    a prevention program developing hope and optimism. *Dissertation Abstracts
    International: Section B. The Sciences and Engineering, 66*(6), 3434.

*Gillham, J. E. (1995). *Preventing depressive symptoms in school children.* ProQuest
    Information & Learning, US.

*Gillham, J. E., Reivich, K. J., Brunwasser, S. M., Freres, D. R., Chajon, N. D., Kash-
    MacDonald, V. M., et al. (2012). Evaluation of a group cognitive-behavioral depression
    prevention program for young adolescents: A randomized effectiveness trial. *Journal of
    Clinical Child and Adolescent Psychology, 41*(5), 621-639.

*Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatté, A. J., Samuels, B., et al.
    (2007). School-based prevention of depressive symptoms: A randomized controlled study
    of the effectiveness and specificity of the Penn Resiliency Program. *Journal of
    Consulting and Clinical Psychology, 75*(1), 9-19.

*Gillham, J. E., Reivich, K. J., Freres, D., Lascher, M., Litzinger, S., Shatte, A., & Seligman, M.
    E. (2006). School-based prevention of depression and anxiety symptoms in early

adolescence: A pilot of a parent intervention component. *School Psychology Quarterly, 21,* 323–348.

Gotlib, I. H., & Goodman, S. H. (1999). Children of parents with depression. In W. K. Silverman & T. H. Ollendick (Eds.), *Developmental issues in clinical treatment of children* (pp. 415-432). Needham Heights, MA: Allyn & Bacon.

Gotlib, I. H., & Hammen, C. L. (2009). *Handbook of depression (2nd ed.)*. New York, NY US: Guilford Press.

Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment, 4*(1).

Gueldner, B. A. (2008). *The effectiveness of a social-emotional learning program with middle school students in a general education setting and the impact of consultation support using performance feedback.* ProQuest Information & Learning, US.

*Gueldner, B. K. (2011). Evaluation of a social-emotional learning program in conjunction with the exploratory application of performance feedback incorporating motivational interviewing techniques. *Journal of Educational & Psychological Consultation, 21*(1), 1-27.

*Hains, A. A. (1992). A stress inoculation training program for adolescents in a high school setting: A multiple baseline approach. *Journal of Adolescence, 15*(2), 163-175.

*Hains, A. A., & Ellmann, S. W. (1994). Stress inoculation training as a preventative intervention for high school youths. *Journal of Cognitive Psychotherapy, 8*(3), 219-232.

*Hains, A. A., & Szyjakowski, M. (1990). A cognitive stress-reduction intervention program for adolescents. *Journal of Counseling Psychology, 37*(1), 79-84.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23*, 56-61.

Hammen, C. (2006). Stress generation in depression: Reflections on origins, research, and future directions. *Journal of Clinical Psychology, 62,* 1065-1082.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology, 107*(1), 128-14.

Hirshfeld, R., Keller, M., Panico, S., Arons, B., Barlow, D., Davidoff, F., et al. (1997). The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *Journal of the American Medical Association, 277,* 333-34.

*Horn, A. B., Pössel, P., & Hautzinger, M. (2011). Promoting adaptive emotion regulation and coping in adolescence: A school-based programme. *Journal of Health Psychology, 16*(2), 258-273.

Horowitz, J. & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 74,* 401-415.

*Horowitz, J. L., Garber, J., Ciesla, J. A., Young, J. F., & Mufson, L. (2007). Prevention of depressive symptoms in adolescents: A randomized trial of cognitive-behavioral and interpersonal prevention programs. *Journal of Consulting and Clinical Psychology, 75*(5), 693-706.

Jané-Llopis, E., Hosman, C., Jenkins, R., & Anderson, P. (2003). Predictors of efficacy in depression prevention programmes: Meta-analysis. *The British Journal of Psychiatry, 183*(5), 384-397.

*Jaycox, L. H., Reivich, K. J., Gillham, J., & Seligman, M. E. (1994). Prevention of depressive symptoms in school children. *Behaviour Research and Therapy, 32*(8), 801-816.

Joiner, T., Coyne, J. C., & Blalock, J. (1999). On the interpersonal nature of depression: Overview and synthesis. In T. Joiner & J. C. Coyne (Eds.), *The interactional nature of depression: Advances in interpersonal approaches* (pp. 3-19). Washington, DC US: American Psychological Association.

Kazdin, A. E. (1995). *Conduct disorders in childhood and adolescence* (2nd ed.)*.* Thousand Oaks, CA: Sage.

Kessler, R. C., Avenei, S., & Merikangas, K. R. (2001). Mood disorders in children and adolescents: An epidemiologic perspective. *Biological Psychiatry, 49,* 1002-1014.

Kim-Cohen, J., Caspi, A., & Moffitt T. E. (2003). Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry, 60,* 709-717.

Kovacs, M., Feinberg, T. L., Crouse-Novak, M., Paulauskas, S. L., Pollock, M., & Finkelstein, R. (1984). Depressive disorders in childhood. II. A longitudinal study of the risk for a subsequent major depression. *Archives of General Psychiatry, 41,* 653-649.

*Kowalenko, N., Rapee, R. M., Simmons, J., Wignall, A., Hoge, R., Whitefield, K., et al. (2005). Short-term effectiveness of a school-based early intervention program for adolescent depression. *Clinical Child Psychology and Psychiatry, 10*(4), 493-507.

*Kraag, G., Van Breukelen, G. J. P., Kok, G., & Hosman, C. (2009). 'Learn Young, Learn Fair',
a stress management program for fifth and sixth graders: Longitudinal results from an
experimental study. *Journal of Child Psychology & Psychiatry, 50*(9), 1185-1195.

Langley, A. K., Nadeem, E., Kataoka, S. H., Stein, B. D., & Jaycox, L. H. (2010). Evidence-
based mental health programs in schools: Barriers and facilitators of successful
implementation. *School Mental Health, 2*(3), 105-113.

Lipsey, M. & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA. Sage
Publications.

Liu, L. L., & Lau, A. S. (2013). Teaching about race/ethnicity and racism matters: An
examination of how perceived ethnic racial socialization processes are associated with
depression symptoms. *Cultural Diversity and Ethnic Minority Psychology, 19*(4), 383-
394.

*Lock, S., & Barrett, P. M. (2003). A longitudinal study of developmental differences in
universal preventive intervention for child anxiety. *Behaviour Change, 20*(4), 183-199.

*Lowry-Webster, H. M., Barrett, P. M., & Lock, S. (2003). A universal prevention trial of
anxiety symptomology during childhood: Results at 1-year follow-up. *Behaviour
Change, 20*(1), 25-43.

Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological
Science, 7,* 186-189.

Lyubomirsky, S. (2007). *The how of happiness.* New York, NY: The Penguin Press.

*Manassis, K., Wilansky-Traynor, P., Farzan, N., Kleiman, V., Parker, K., & Sanford, M.
(2010). The feelings club: Randomized controlled evaluation of school-based CBT for
anxious or depressive symptoms. *Depression and Anxiety, 27*(10), 945-952.

*Marchand, E., Ng, J., Rohde, P., & Stice, E. (2010). Effects of an indicated cognitive-behavioral depression prevention program are similar for Asian American, Latino, and European American adolescents. *Behaviour Research and Therapy, 48*(8), 821-825.

Mason, C. A., Scott, K. G., Chapman, D. A., & Shihfen, T. (2000). A review of some individual and community-level effect size indices for the study of risk factors and child and adolescent development. *Educational and Psychological Measurement, 60*(3), 385-41.

*McCarty, C. A., Violette, H. D., Duong, M. T., Cruz, R. A., & McCauley, E. (2013). A randomized trial of the positive thoughts and action program for depression among early adolescents. *Journal of Clinical Child and Adolescent Psychology, 42*(4), 554-563.

Merry, S. N., Hetrick, S. E., Cox, G. R., Brudevold-Iversen, T., Bir, J. J., & McDowell, H. (2011). Psychological and educational interventions for preventing depression in children and adolescents. *Cochrane Database of Systematic Reviews (Online)*(12), CD003380.

Merry, S., McDowell, H., Hetrick, S., Bir, J., & Muller, N. (2004). Psychological and/or educational interventions for the prevention of depression in children and adolescents. *Cochrane Database of Systematic Reviews (Online)*(1), CD00338.

*Merry, S., McDowell, H., Wild, C. J., Bir, J., & Cunliffe, R. (2004). A randomized placebo-controlled trial of a school-based depression prevention program. *Journal of The American Academy of Child And Adolescent Psychiatry, 43*(5), 538-547.

Naify, A. M. (2008). Effectiveness of a brief strength-based intervention on depression in at-risk adolescent girls. *Dissertation Abstracts International: Section B. The Sciences and Engineering, 70*(6), 3791.

National Institute of Mental Health (2013, August 20). *Depression in adolescents and children fact sheet.* Retrieved from http://www.nimh.nih.gov/health/publications/depression-in-children-and-adolescents/index.shtml

Nezu, A. M., Nezu, C. M., Friedman, J., & Lee, M. (2009). Assessment of depression. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression (2nd ed.)* (pp. 44-68). New York, NY US: Guilford Press.

Nolen-Hoeksema, S., & Hilt, L. M. (2009). *Handbook of depression in adolescents*. New York, NY US: Routledge/Taylor & Francis Group.

Oregon Resiliency Project, (2010). *Strong kids*, retrieved Aug. 30, 2013. http://strongkids.uoregon.edu/index.html

*Pattison, C., & Lynd-Stevenson, R. M. (2001). The prevention of depressive symptoms in children: The immediate and long-term outcomes of a school based program. *Behaviour Change, 18*(2), 92-102.

Peterson, C. & Seligman, M. E. (2004). *Character strengths and virtues.* New York, NY: Oxford University Press.

Positive Psychology Center, (2007). *Resilience research in children*, retrieved Aug. 30, 2013. http://www.ppc.sas.upenn.edu/prpsum.htm

*Pössel, P., Adelson, J. L., & Hautzinger, M. (2011). A randomized trial to evaluate the course of effects of a program to prevent adolescent depressive symptoms over 12 months. *Behaviour Research and Therapy, 49*(12), 838-851.

*Pössel, P., Baldus, C., Horn, A. B., Groen, G., & Hautzinger, M. (2005). Influence of general self-efficacy on the effects of a school-based universal primary prevention program of

depressive symptoms in adolescents: A randomized and controlled follow-up study. *Journal of Child Psychology and Psychiatry, 46*(9), 982-994.

*Pössel, P., Horn, A. B., Groen, G., & Hautzinger, M. (2004). School-based prevention of depressive symptoms in adolescents: A 6-month follow-up. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(8), 1003-101.

*Pössel, P., Martin, N. C., Garber, J., & Hautzinger, M. (2013). A randomized controlled trial of a cognitive-behavioral program for the prevention of depression in adolescents compared with nonspecific and no-intervention control conditions. *Journal of Counseling.* Advance online publication. doi: 10.1037/a0032308

*Quayle, D., Dzuirawiec, S., Roberts, C., Kane, R., & Ebsworthy, G. (2001). The effect of an optimism and lifeskills program on depressive symptoms in preadolescence. *Behaviour Change, 18*(4), 194-203.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

*Raes, F., Griffith, J. W., Gucht, K., & Williams, J. M. G. (2013). School-based prevention and reduction of depression in adolescents: A cluster-randomized controlled trial of a mindfulness group program. *Mindfulness*, *3,* 2-11.

*Rivet-Duval, E., Heriot, S., & Hunt, C. (2011). Preventing adolescent depression in Mauritius: A universal school-based program. *Child and Adolescent Mental Health, 16*(2), 86-91.

Roberts, C. M. (1999). The prevention of depression in children and adolescents. *Australian Psychologist, 34,* 49-57.

*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.

*Roberts, C., Kane, R., Bishop, B., Matthews, H., & Thomson, H. (2004). The prevention of depressive symptoms in rural school children: A follow-up study. *International Journal of Mental Health Promotion, 6*(3), 4-16.

Roberts, C., 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.

*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.

*Rooney, R., Roberts, C., Kane, R., Pike, L., Winsor, A., White, J., et al. (2006). The prevention of depression in 8- to 9-year-old children: A pilot study. *Australian Journal of Guidance and Counselling, 16*(1), 76-9.

Rudolph, K. D., Flynn, M., & Abaied, J. L. (2008). A developmental perspective on interpersonal theories of youth depression. In J. R. Z. Abela & B. L. Hankin (Eds.), *Handbook of depression in children and adolescents* (pp. 79-102). New York, NY US: Guilford Press.

Santor, D. A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives, 4*(3), 135-155.

*Sawyer, M. G., Pfeiffer, S., Spence, S. H., Bond, L., Graetz, B., Kay, D., Patton, G., Sheffield, J. (2010). School-based prevention of depression: A randomised controlled study of the

beyond blue schools research initiative. *Journal of Child Psychology & Psychiatry, 51*(2), 199-209.

Schulman, P., Keith, D., & Seligman M. (1993). Is optimism heritable? A study of twins. *Behaviour Research and Therapy, 6,* 569-574.

Seligman, M. E. (2007). *The optimistic child*. New York, NY: Houghton Mifflin.

Seligman, M. E. (2011). *Flourishing.* New York, NY: Houghton Mifflin.

Sharp, L. K., & Lipsky, M. S. (2002). Screening for depression across the lifespan: A review of measures for use in primary care settings. *American Family Physician, 66*(6), 1001-1008.

*Shatte, A. J. (1997). *Prevention of depressive symptoms in adolescents: Issues of dissemination and mechanisms of change.* ProQuest Information & Learning, US.

*Sheffield, J. K., Spence, S. H., Rapee, R. M., Kowalenko, N., Wignall, A., Davis, A., et al. (2006). Evaluation of universal, indicated, and combined cognitive-behavioral approaches to the prevention of depression among adolescents. *Journal of Consulting and Clinical Psychology, 74*(1), 66-79.

Sheldon, K. M., & Lyubomirsky, S. (2006). Achieving sustainable gains in happiness: Change your actions, not your circumstances. *Journal of Happiness Studies, 7*(1), 55-86.

Shemilt, I., Valentine, J. C., Pössel, P., Mugford, M., & Wooldridge, D. T. (2012). Costing program implementation using systematic reviews: Interventions for the prevention of adolescent depression. *Research Synthesis Methods, 3*(3), 191-201.

*Shochet, I. M., Dadds, M. R., Holland, D., Whitefield, K., Harnett, P. H., and Osgarby, S. M. (2001). The efficacy of a universal school-based program to prevent adolescent depression. *Journal of Clinical Child Psychology, 30*(3), 303-315.

Shochet, I. M., & Ham, D. (2004). Universal school-based approaches to preventing adolescent depression: Past findings and future directions of the resourceful adolescent program. *International Journal of Mental Health Promotion, 6*(3), 17-25.

Short, J. L., & Russell-Mayhew, S. (2009). What counselors need to know about resiliency in adolescents. *International Journal for the Advancement of Counselling, 31,* 213-227.

*Simpson, A. T. (2008). *The roles of self-regulation and coping in a preventative cognitive-behavioural intervention for school-age children at-risk for internalizing disorders.* ProQuest Information & Learning, US.

Snaith, P. (1993). What do depression rating scales measure? *British Journal of Psychiatry, 163,* 293–298.

Spence, S. H. (2008). Integrating individual and whole-school change approaches in the prevention of depression in adolescents. In J. R. Z. Abela & B. L. Hankin (Eds.), *Handbook of depression in children and adolescents.* (pp. 333-353). New York, NY US: Guilford Press.

*Spence, S. H., Sheffield, J. K., & Donovan, C. L. (2005). Long-term outcome of a school-based, universal approach to prevention of depression in adolescents. *Journal of Consulting and Clinical Psychology, 73*(1), 160-167.

Steinberg, L. (2008). *Adolescence.* New York, NY: McGraw-Hill.

*Stice, E., Rohde, P., Gau, J. M., & Wade, E. (2010). Efficacy trial of a brief cognitive–behavioral depression prevention program for high-risk adolescents: Effects at 1- and 2-year follow-up. *Journal of Consulting and Clinical Psychology, 78*(6), 856-867.

Stice, E. & Shaw, H. (2004). Eating disorder prevention programs: A meta-analytic review. *Psychological Bulletin, 130,* 206-227.

Stice, E., Shaw, H., Bohon, C., Marti, C. N., & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting and Clinical Psychology, 77*(3), 486-503.

*Sung, K. M. (2012). Effects of a school-based intervention program for middle school adolescent girls with depression: As part of the school health services. *Journal of Korean Academy of Nursing, 42*(7), 984-991.

Weissman, M. M., Wickramaratne, P., Nomura, Y., Warner, V., Verdeli, H., & Pilowsky, D. J., et al. (2005). Families at high and low risk for depression: A 3-generation study. *Archives of General Psychiatry, 62,* 29-36.

Wight, R. G., Sepulveda, J. E., & Aneshensel, C. S. (2004). Depressive symptoms: How do adolescents compare with adults? *Journal of Adolescent Health, 34,* 314-323.

*Wijnhoven, L. A. M. W., Creemers, D. H. M., Vermulst, A. A., Scholte, R. H. J., & Engels, R. C. M. E. (2013). Randomized controlled trial testing the effectiveness of a depression prevention program ('op volle kracht') among adolescent girls with elevated depressive symptoms. *Journal of Abnormal Child Psychology*. Advance online publication. doi 10.1007/s10802-013-9773-5

*Wong, P. W. C., Fu, K.-W., Chan, K. Y. K., Chan, W. S. C., Liu, P. M. Y., Law, Y.-W., et al. (2012). Effectiveness of a universal school-based programme for preventing depression in Chinese adolescents: A quasi-experimental pilot study. *Journal of Affective Disorders, 142*(1-3), 106-114.

*Young, J. F., Mufson, L., & Davies, M. (2006). Efficacy of Interpersonal Psychotherapy-

    Adolescent Skills Training: An indicated preventive intervention for depression. *Journal*

    of Child Psychology and Psychiatry, 47(12), 1254-1262.

*Yu, D. L., & Seligman, M. E. P. (2002). Preventing depressive symptoms in Chinese children.

    Prevention & Treatment, 5(1): ArtID9.

Appendix

Table 1
Depression Prevention Outcomes and Measures – Assessments of Overall Depressive Symptoms

Depression and Secondary Outcomes with Corresponding Measures

Depression Outcome

        Beck Depression Inventory (BDI)
        Brief Symptom Inventory for depression (BSI)
        Child Behavior Checklist - depression subscale (CBCL/D)
        Children's Depression Inventory (CDI)
        Center for Epidemiologic Studies Depression Index (CES-D)
        Depression Anxiety Stress Scales – depression subscale (DASS21-D)
        Depression Self Rating Scale (DSRS)
        Hamilton Depression Rating Scale (HDRS)
        Kiddie-Schedule for Affective Disorders and Schizophrenia (KSADS)
        Moods and Feelings Questionnaire (MFQ)
        Mood and Anxiety Symptom Questionnaire for depressed symptoms (MASQ-D)
        Positive and Negative Affect Schedule (PANAS)
        Reynolds Adolescent Depression Scale (RADS)
        Reynold's Children's Depression Scale (RCDS)
        Self Report Questionnaire – depression subscale (SBB-DES)
        Short Depression Inventory for Children (SDIC)

Anxiety Outcome

        Anxiety Self Statements Questionnaire (ASSQ)
        Beck Anxiety Inventory (BAI)
        Childhood Anxiety Sensitivity Index – AS subscale (CASI)
        Depression Anxiety Stress Scales – anxiety subscale (DASS21-A)
        Multi-dimensional Anxiety Scale for Children (MASC)
        Mood and Anxiety Symptom Questionnaire for anxiety (MASQ-A)
        Perceived Stress Scale (PSS)
        Revised Children's Manifest Anxiety Scale (RCMAS)
        Screen for Child Anxiety Relational Emotional Disorders (SCARED)
        Spence Children's Anxiety Scale (SCAS)
        State Trait Anxiety Inventory for Children (STAI-C)

Coping Outcome

        Adolescent Coping Scale (ACS)
        Brief Coping Orientations to Problems Experienced (COPE)

Children's Coping Scale adapted from the ACS (CCS)
Children's Coping Strategies Checklist-Revision (CCSC-R)
Coping Action Scale (CAS)
Social Problem Solving Inventory – NPO and AS subscales (SPSI)
Youth Coping Index (YCI)

Self-Esteem Outcome

Behavior Assessment Scale for Children – self-esteem subscales (BASC2)
Contingencies of Self Worth Scale (CSW)
Coopersmith Self-Esteem Inventory (CSE)
General Self-efficacy (GSE)
Rosenberg Self-Esteem Scale (RSES)
Self-Perception Profile for Children or Adolescents (SPPC/A)
What I Am Like/Perceived Self Competence Scale (WIAL)

Cognitive/Attributional
Style Outcome

Automatic Thoughts Questionnaire (ATQ)
Beck Hopelessness Scale (BHS)
Children's Attributional Style Questionnaire (CASQ)
Children's Automatic Thoughts Scale (CATS)
Children's Hope Scale (CHS)
Cognitive Triad Inventory (CTI-C)
Hopelessness Scale (H-Scale)
Hopelessness Scale for Children (HSC)
Optimistic Thinking Style (OTS)

Stress

Adolescent Perceived Events Scale – stress (APES)
Perceived Stress Scale (PSS)
Maastricht University Stress Instrument for Children (MUSIC)

Internalizing

Depression Outcome Measures
Anxiety Outcome Measures
Stress Measures
Behavior Assessment Scale for Children – internalizing subscale
(BASC2)
Child Behavior Checklist – internalizing subscale (CBCL)
Internalizing Symptoms Scale for Children (ISSC)
Teacher Report Form for internalizing behavior – (TRF)
Youth Self Report for internalizing (YSR)

Table 2

Overall Effect Sizes by Outcome

Outcome	Effect Size	k	p	Q	I^2
Overall Combined	.151	82	< .001	108.96*	26.58*
Depression	.154	80	< .001	148.16***	47.36***
Anxiety	.132	26	< .001	31.51	23.84
Cognitive	.080	30	< .05	56.21**	48.40**
Self-Esteem	.251	12	< .001	10.02	.00
Coping	.125	13	< .01	32.21**	62.74**
Internalizing	.164	80	< .001	132.79***	41.26***

*Note: *$p < .05$, ***$p < .01$, ****$p < .001$.*

Table 3

Moderator Analyses for Methodological Comparisons and Assessment Timing Effect Sizes

Moderators	Group Comparisons	k	Effect Size	p	Q	p
Study Design	Experimental	44	.246	< .001	25.378	< .001
	Quasi-Experimental	38	.064	< .001		
Depression Measures	BDI	6	.227	< .001	4.281	.233
	CDI	45	.122	< .001		
	CESD	16	.099	.058		
	RADS	11	.214	< .01		
Data Collection Times	Post	80	.174	< .001	NR	NR
	Follow-up	63	.144	< .001		
Timing of Follow-ups	1-6 months	50	.169	< .001	NR	NR
	7-12 months	28	.108	< .001		
	13-36 months	13	.051	.097		

Table 4
Moderator Analyses for Sample Participant Characteristics

Moderators	Group Comparisons	Overall Effect Size (k)	Q	Depression Effect Size (k)	Depression Q	Cognitive Effect Size (k)	Cognitive Q	Coping Effect Size (k)	Coping Q	Internalizing Effect Size (k)	Internalizing Q
School Level	Elementary	.13*** (13)	1.49	.14*** (13)	.19	.10 (8)	1.40	.15** (3)	1.97	.12*** (13)	1.99
	Middle	.14*** (38)		.14*** (35)		.03 (14)		.06 (3)		.18*** (36)	
	High School	.19*** (31)		.16*** (30)		.119 (8)		.16* (7)		.16*** (31)	
Ethnicity	Predominantly non-White	.27*** (18)	4.85*	.28*** (16)	3.77*	.23** (5)	7.00*	.28** (3)	2.00	.28*** (16)	3.08
	Predominantly White	.14*** (39)		.14*** (39)		.04 (10)		.12 (5)		.15*** (39)	
Risk Status	Targeted	.31*** (30)	27.68***	.35*** (28)	25.09***	.17*** (9)	3.23	.46* (1)	2.27	.36*** (28)	28.67***
	Universal	.07*** (52)		.09*** (52)		.04 (21)		.12** (12)		.10*** (52)	
Geographic Location	Asia	.43 (3)	3.62	.42 (3)	3.47	.30* (1)	2.99	NR		.43 (3)	1.88
	Australia	.10** (20)		.10* (20)		.09 (14)				.13*** (20)	
	Germany	.16*** (15)		.18** (15)		.10 (4)				.18** (15)	
	U.S.	.15*** (40)		.12*** (40)		.04 (11)				.16*** (40)	

Note: *p* < .05, **p* < .01, ***p* < .001. NR indicates there was no, or too little, data to report. ^ indicates a significant difference between marked groups.

Table 5
Moderator Analyses for Programmatic Characteristics

Moderators	Group Comparisons	Effect Size (k)	Effect Size Q	Depression (k)	Depression Q	Cognitive (k)	Cognitive Q	Coping (k)	Coping Q	Internalizing (k)	Internalizing Q
Curriculum	Non-Penn	.14*** (59)	1.56	.14*** (59)	3.08	.07 (17)	.31	.13** (11)	2.14	.14*** (59)	2.93
	Penn	.20*** (23)		.22*** (21)		.11* (13)		.10 (2)		.24*** (21)	
Facilitator	Combination~	.11** (20)	15.52***	.08 (20)	1.77**	-.15 (3)	1.42	.16 (1)	.07	.11** (20)	1.96**
	Non-school^~	.24*** (38)		.27*** (36)		.11 (15)		.14 (3)		.27*** (36)	
	School^	.088*** (24)		.10*** (24)		.07 (14)		.13** (9)		.10*** (24)	
Length of Treatment (hours)	Less than 12 hours	.19*** (37)	1.72	NR		NR		NR		NR	
	More than 12 hours	.14*** (45)									
Span of Treatment (weeks)	1 - 4 weeks	.14* (8)	2.44	NR		NR		NR		NR	
	5 - 9 weeks	.22*** (17)									
	10 - 12 weeks	.15*** (54)									
	15 - 20	.08 (3)									
During or After School	After School	.25*** (22)	7.94*	NR		NR		NR		NR	
	During School	.12*** (56)									
Homework/Parental Component	No	.13*** (37)	2.32	NR		NR		NR		NR	
	Homework	.18*** (27)									
	Parent Involvement	.15*** (10)									

Note: *p < .05, **p < .01, ***p < .001. NR indicates there was no, or too little, or non-significant data to report. ^ and ~ indicate a significant difference between marked groups.

Figure 1
Flow of studies included and excluded from the meta-analysis.