CRITICALLY EVALUATING SCHOOL MENTAL HEALTH: PILOT REPORT

Pilot report for the SOS Suicide Prevention Program and Yellow Ribbon Suicide Prevention Program

CRITICALLY EVALUATING SCHOOL MENTAL HEALTH: PILOT REPORT

Pilot Report on the Effectiveness, Cost-Effectiveness and Safety of the SOS Suicide Prevention Program and Yellow Ribbon Suicide Prevention Program

AUTHORS: Ms. Yifeng Wei (PhD Candidate) and Dr. Stan Kutcher

STEERING COMMITTEE: Dr. Connie Coniglio, Dr. Wendy Craig, Ms. Lisa Doucet, Dr. Ian Manion, Ms. Melissa Rowbotham, Dr. Kathy Short, Dr. Kimberly Schonert-Reichl, Ms. Margo Tait, Dr. Chris Wilkes and Mr. Ray Hughes

EDUCATOR ADVISORY GROUP: Ms. Lisa Doucet, Ms Margo Tait, Ms. France McClure, Dr. Carla Baetz, Mr. Joseph Damico, Ms. Lori Roe, Mr. Andrew Heaton, Ms. Roza Gray and Ms. Barb Isaac

INTRODUCTION

The majority of mental disorders have their onset prior to the age of 25 (Kessler et al., 2005), demonstrating the need for interventions designed to address mental health promotion, early identification and effective early interventions to increase the possibility of positive outcomes for young people, including better educational and social outcomes, reduced burden of illness, and improved physical health (Canadian Council on Learning, 2009; S. Kutcher, 2011; Waddell, Offord, Shepherd, Hua, & McEwan, 2002; Wei, Kutcher, & Szumilas, 2011).

Schools have been identified as appropriate institutions for addressing youth mental health, by internationally recognized agencies (U.S. Department of Health and Human Services, 1999; WHO Regional Office for Europe, 1996), educators and researchers (Farmer, E.M., Burns, B.J., Phillips, S.D., Angold, A., & Costello, E.J., 2003; Roeser, R.W., & Midgley, C., 1997; Rones, M., & Hoagwood, K.E., 2000; Wyn, J., Cahill, H., Holdsworth, R., Rowling, L., & Carson, S., 2000) and in national mental health policy frameworks such as Evergreen (S. P. Kutcher & McLuckie, A. for the Child and Youth Advisory Committee, Mental Health Commission of Canada., 2010). Currently, school administrators are faced with a plethora of programs purporting to address numerous aspects of "mental health". While the need to use best evidence-based interventions is recognized (Forman, S.G., Olin, S.S., Hoagwood, K.E., Crowe, M., & Saka, N., 2009), educators and schools continue to select and implement heavily marketed and costly programs that lack appropriate evidence for their effectiveness or safety (Hallfors, D., & Godette, D., 2002). This may be due in large part to: a) the lack of research evaluation capacity within schools/school boards to, identify and critically evaluate marketed programs; b) the lack of an independent best evidence based source of information on mental health promotion/prevention programs that are relevant for Canadian schools.

At present, there are no known national programs or services in Canada that partner mental health or education researchers with school professionals/schools/school boards for the explicit purpose of critically evaluating school based mental health promotion/prevention programs for safety, effectiveness and cost effectiveness. This report describes findings of a pilot project that may establish the foundation of an online repository of mental health promotion/prevention programs, designed to provide such a service to educators across Canada to assist them in selecting which of these programs they will choose to implement in schools.

The pilot project selects two school based suicide prevention programs, widely disseminated and heavily marketed to schools in North America: the Signs of Suicide (SOS) Prevention Program and the Yellow Ribbon Suicide Prevention Program, to assess their effectiveness, cost-effectiveness and safety, using a systematic review approach recommended by The Cochrane Collaboration (http://www.cochrane.org/). The choice of these two programs was the result of the discussion among the steering committee members of this pilot project that is composed of mental health experts, educators, school administrators, and researchers from across Canada. Findings of this review are reported based on the recommendations suggested by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (http://www.prisma-statement.org/index.htm) that is widely accepted in the health/public health research field.

METHODS

INCLUSION CRITERIA

We included research studies that specifically evaluate the effectiveness, cost-effectiveness, and safety of the SOS and Yellow Ribbon Suicide Prevention programs. We also included published systematic reviews or meta-analysis on suicide prevention since they may include studies addressing the two target programs. We only included studies published in peer-reviewed journals in English without time limits. We included studies of any type, and with any outcome as long as they evaluated the two target programs. Systematic reviews or meta-analysis of any type and level of suicide prevention were eligible for inclusion.

EXCLUSION CRITERIA

Non-English publications were excluded. Any systematic reviews or meta-analysis dealing with suicides related with physical illness were not eligible for inclusion. We did not include reports or studies from grey literature.

SEARCH METHODS

With the assistance of a health librarian, we searched a number of databases to identify relevant studies and reviews, including Medline (including HealthSTAR), PsycINFO (including Dissertation Abstracts International), EMBASE, CINAHL, the Cochrane Library, the Campbell Collaboration SPECTR database, SocIndex, Sociological Abstracts, Social Services Abstracts, ERIC, Social Work Abstracts, Research Library, and Web of Science. The search started with two sets of search terms to capture any studies related with suicide and intervention effectiveness, cost-effectiveness and program safety. Key search words for these two steps are: suicide, self-harm, self-injury/behavior/inflict/hatred, suicide attempt, prevention, crisis intervention, health services, health education, program development, program evaluation, experiments, trials, effectiveness, efficacious, cost-effectiveness, safety, and harm effects. The results from the above two steps were combined using the Boolean AND, which were further filtered by key terms "SOS suicide prevention" and "Yellow Ribbon suicide prevention". Meanwhile, we repeated the first two steps but filtered the results with key words "systematic review" and meta-analysis, to capture existing reviews/meta-analysis that could have included the two target programs. Appendix 1 includes detailed search strategies for this report.

STUDY SELECTION AND DATA EXTRACTION

We ran the search with the key terms in the identified databases and imported the results into the RefWorks 2.0 database management program. Duplicates were removed. We then screened titles and abstracts of imported studies to delete studies that are not relevant to the topic of interests. The remaining research studies were further assessed for inclusion/exclusion by reading the full text, and systematic reviews/meta-analysis were scanned to check whether the two target programs were included and investigated, which resulted in the final studies and systematic reviews/meta-analysis for inclusion for this report.

A data extraction form, developed a priori, to obtain information such as identifying information, study eligibility, study characteristics (baseline characteristics, location, timing, methods, and intervention type and duration), outcome measures and quantitative data.

CRITICAL APPRAISAL

We applied The Office of Justice Program (OJP) What Works Repository (National Criminal Justice Reference Services, 2005) to assess the quality of evidence of the included studies. The OJP What Works Repository classifies programs into 6 levels of evidence of effectiveness: effective, effective with reservation, promising, inconclusive evidence, insufficient evidence, and ineffective; and 3 levels of readiness for dissemination: fully prepared for widespread dissemination, fully prepared for limited dissemination, and not ready for dissemination. Programs are critically appraised against the following 5 criteria to determine its level of evidence: randomized controlled trials, replication with different population and contexts, focus on socially important behavior outcomes, identification of evidence of enduring effects, and dissemination capacity. Further, programs are assessed against 4 indicators to determine levels of readiness for dissemination, including information on training and related support materials, technical assistance support, informational materials, and quality control for implementation. The OJP What Works Repository enables us to determine both the internal and external validity of the included studies.

DATA ANALYSIS

We decided the method to conduct meta-analysis (quantitative) to produce estimates of the effectiveness of the target programs would be to analyze data based on the quality of the studies from critical appraisal and numbers of studies as follows: if the included studies are ranked as "promising" and above based on The OJP What Works Repository and there are at least two studies for each intervention with homogeneity in study design, study duration, outcome measures, population sizes, and timeframe. Otherwise, we will apply a narrative approach (qualitative).

RESULTS

STUDY CHARACTERISTICS

Figure 1 indicates the process of how included studies were obtained, applying the template of flow diagram recommended by PRISMA statement (http://www.prisma-statement.org/statement.htm). Three studies evaluating the SOS program (Aseltine, 2003; Aseltine & DeMartino, 2004; Aseltine, James, Schilling, & Glanovsky, 2007) and 1 study evaluating the Yellow Ribbon program (Freedenthal, 2010) were eligible for inclusion for this report. Two systematic reviews were identified to include one ((M. D. Cusimano & Sameem, 2011b) or two studies (Mann, Apter, Bertolote, Beautrais, Currier, Haas, Hegerl, Lonnqvist, Malone, Marusic, Mehlum, Patton, Phillips, Rutz, Rihmer, Schmidtke, Shaffer, Silverman, Takahashi, Varnik, Wasserman, Yip, & Hendin, 2005b) investigating the effectiveness of the SOS program. However, neither reviews analyzed the overall quality of evidence of effectiveness of the included studies, therefore keeping unknown the level of evidence of both studies. No systematic reviews have been located so far to synthesize the evidence of the Yellow Ribbon program. This finding indicates that it is necessary to conduct a systematic review that critically observes the two target programs in order to justify their wide dissemination in schools.

Figure 1: Flow chart of included studies

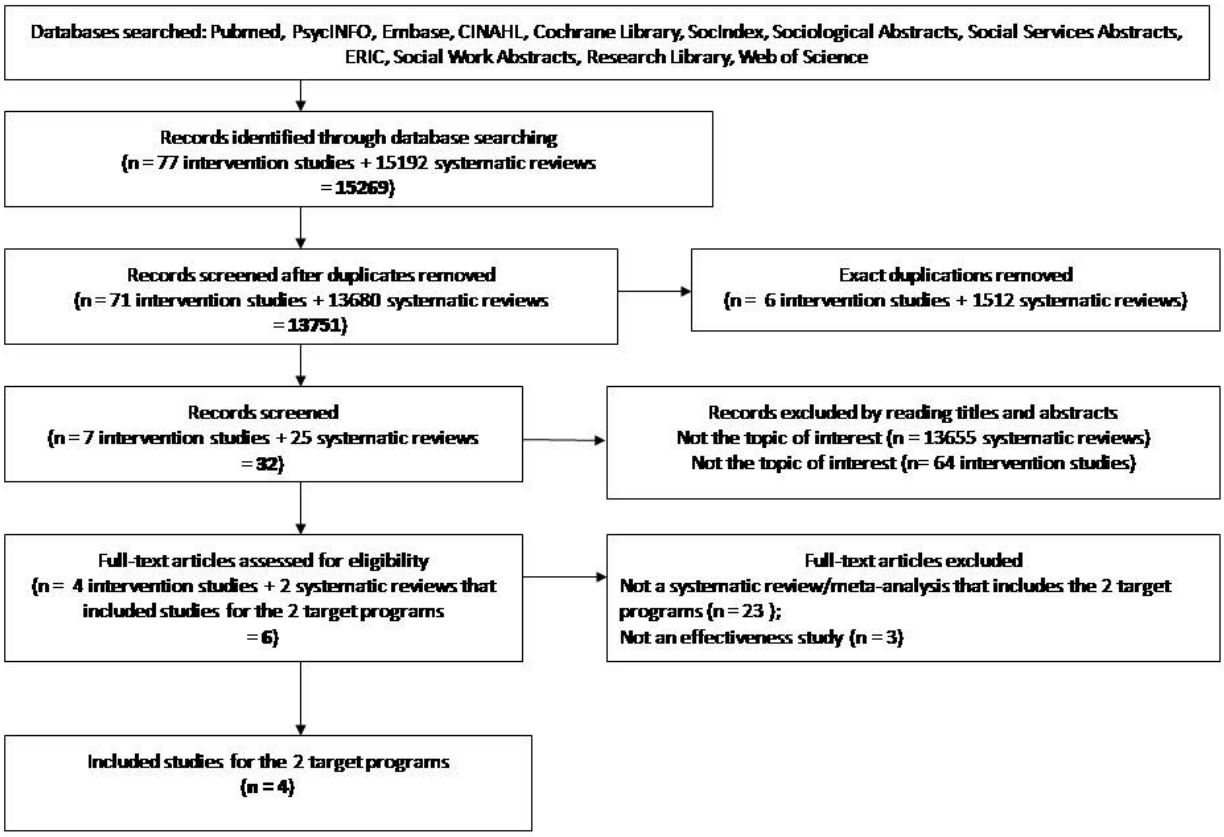

Table 1 summarizes the characteristics of included studies. Of three studies evaluating the SOS program, there are two Randomized Controlled Trial (RCT) studies (Aseltine & DeMartino, 2004; Aseltine et al., 2007) evaluating the effectiveness of the SOS program among high school students in five aspects: knowledge, attitudes, help-seeking behaviors, suicide attempts and suicide ideation. One survey study (Aseltine, 2003) conducted the process evaluation, student help-seeking behaviors, and reported on the perspectives of school staff on cost and benefits of the program. All three studies addressed only short term impact of the program and no follow-up data were collected. The only one study (Freedenthal, 2010) on the Yellow Ribbon program is the first ever study in school based suicide prevention research to investigate the effectiveness of the program on students' help-seeking behaviors. This study also only observed the short-term impact of the program.

Although one study (Aseltine, 2003) demonstrated participants' opinions on cost and benefits of the SOS program, no specific cost-effectiveness studies have been conducted on the SOS program or the Yellow Ribbon program. No studies have been conducted to address whether these two programs will cause harmful or negative effects. There are no reports regarding cost effectiveness. All four studies were carried out in the US.

PROGRAM EFFECTIVENESS

Two RCT studies (Aseltine & DeMartino, 2004; Aseltine et al., 2007) on the SOS program applied similar measures for program effectiveness. The third study (Aseltine, 2003) on the SOS program is a process evaluation survey as well as an evaluation of self-reported help-seeking behaviors. All three studies applied measurement tools specifically developed for the target intervention and the reliability and validity of the measurement tool were unknown except the two RCT studies reported the internal consistency (Cronbach α=.74) of the attitude measurement tool (Aseltine & DeMartino, 2004; Aseltine et al., 2007). The study on the Yellow Ribbon did not report the reliability and validity of its measurement tool either.

Mixed results were reported across the four studies. Table 2 is a summary of results of included studies. Both the RCT studies (Aseltine & DeMartino, 2004; Aseltine et al., 2007) indicated that the SOS program was effective in increasing student knowledge of and attitudes towards depression and suicide and reduced self-reported suicide attempts among students (Table 2, $p<.05$). No effect sizes were reported. However, the SOS program didn't show impact on student help-seeking behaviors in either study and in one study (Aseltine & DeMartino, 2004) that further measured suicidal ideation, the SOS program didn't decrease suicide thoughts among students, making the above self-reported impact on suicide attempts difficult to understand. In the survey study (Aseltine, 2003), although it reported a nearly 60% increase of self-reported student help-seeking behaviors, it remained unknown whether this increase was statistically significant or could have been obtained from other methods of providing information to students. This study further indicated that teachers rated the SOS program as either very or somewhat effective, and most teachers (81%) did not think the program will have adverse effects. No objective measures of these domains were provided.

The study on the Yellow Ribbon program (Freedenthal, 2010) collected data from both school staff and students. School staff data showed that there was a statistically significant decrease of student disclosed suicide attempts ($p<.001$) and thoughts ($p<.05$). However, student data showed that student help-seeking behaviors decreased significantly ($p<.05$) following the intervention although school staff data demonstrated a non-significant small increase of help-seeking behaviors, leading to concerns about the inconsistencies of the findings. School staff data further indicated that the program did not have impact on the number of student-reported peer suicide attempts or thoughts.

No study demonstrated a decrease in completed suicide and no study demonstrated independent evaluation of self-harm or suicide attempts (such as parental report or health record analysis).

CRITICAL APPRAISAL OF INCLUDED STUDIES

Two studies (Aseltine & DeMartino, 2004; Aseltine et al., 2007) were ranked as having "inconclusive evidence" based on the OJP What Works Repository, indicating the SOS programs with experimental or quasi-experimental designs lack sustained effects. One study on SOS (Aseltine, 2003) was ranked as having "insufficient evidence" since it is a purely descriptive evaluation. And one study (Freedenthal, 2010) on Yellow Ribbon was ranked as "ineffective" (A program with an experimental or quasi-experimental research design that in an initial study and at least one

replication failed to demonstrate a significant effect). None of the studies reached to the level of "promising", cut-off criteria for the readiness of program dissemination. The quality of evidence of the internal validity of these studies were undermined by a number of methodological issues, such as the lack of sustained effects, lack of at least one external replication, lack of independent replication, lack of statistical significance (effect sizes or odds ratios) , lack of validity and reliability of the measurement tools used, randomization in only half of the studies, information about intention-to-treat in only half of the studies, lack of information about cost-benefits and information about program safety.

Neither the SOS program nor the Yellow Ribbon program were considered ready for dissemination when assessing the information provided in the 4 included studies against the criteria set by the OJP What Works Repository to determine levels of readiness for dissemination. Appendix 2, 3, 4 and 5 provide detailed quality assessment information about the included studies.

Due to the relatively low level of evidence of effectiveness of all included 4 studies, and variety of study design (RCT and descriptive evaluation), variability and lack of validation of measurement tools, insufficient number of studies (the Yellow Ribbon Program), it is inappropriate to conduct a meta-analysis that combine the results of included studies to produce general estimates of the effectiveness of the two target programs. Therefore, we provide a narrative analysis of results that informs our conclusions

DISCUSSION

We applied systematic and methodologically clear criteria to search for relevant studies on the SOS and the Yellow Ribbon program and we are confident the 4 included studies are the exhausted list of related studies as reported in the literature. The findings that SOS program is effective in improving knowledge of and attitudes towards suicide and depression are consistent with other similar studies evaluating various types of school or community based universal suicide prevention program in children and adolescents (M. D. Cusimano & Sameem, 2011b; Gould, Greenberg, Velting, & Shaffer, 2003; Mann, Apter, Bertolote, Beautrais, Currier, Haas, Hegerl, Lonnqvist, Malone, Marusic, Mehlum, Patton, Phillips, Rutz, Rihmer, Schmidtke, Shaffer, Silverman, Takahashi, Varnik, Wasserman, Yip, & Hendin, 2005a; Ploeg et al., 1996; van et al., 2011; York et al., 2013) and are not unique to the SOS program. Furthermore, it is unknown if the SOS or Yellow Ribbon program contributions to improved knowledge and attitudes are significantly greater than, equal to or less than traditional educational interventions such as embedding information about depression and suicide in usual school curriculum. This information is essential for educators and school administrators, for if similar results can be obtained without the purchase of costly add-on programs, there may be less interest in using either program in the school setting.

While improvements in knowledge and attitudes are useful, none of the four included studies showed significant increase of help-seeking behaviors among participants. This adds confusion to the existing conflicting evidence on the effectiveness of universal suicide prevention programs for this outcome in schools. For instance, a review of youth suicide risk and preventive interventions (Gould et al., 2003) demonstrates conflicting evidence for help-seeking behaviors, and claims that improvement in knowledge and attitudes may not necessarily enhance help-seeking behaviors. Another review (Mann, Apter, Bertolote, Beautrais, Currier, Haas, Hegerl, Lonnqvist, Malone, Marusic, Mehlum, Patton, Phillips, Rutz, Rihmer, Schmidtke, Shaffer, Silverman, Takahashi, Varnik, Wasserman, Yip, & Hendin, 2005a) further stresses that public education and awareness program demonstrates no detectable effects on treatment seeking. It is thus unlikely that either program may have an impact on prevention of suicide if these interventions do not bring young people into treatment (Gould et al., 2003; Mann, Apter, Bertolote, Beautrais, Currier, Haas, Hegerl, Lonnqvist, Malone, Marusic, Mehlum, Patton, Phillips, Rutz, Rihmer, Schmidtke, Shaffer, Silverman, Takahashi, Varnik, Wasserman, Yip, & Hendin, 2005a). We also recognize that help-seeing behavior is a proxy measure for completed suicide and that even with substantive changes in rates of help seeking behavior there may be no positive impact on suicide rates. However, without substantive evidence that these programs have a robust impact on increasing help seeking behavior we are not even able to suggest that they might have an impact on suicide prevention, let alone that they are suicide prevention programs.

Similarly, the relationship of other proxy measures, such as self-reports of suicide related ideation or suicide attempts, to suicide behaviors/acts is not robust. Epidemiological data demonstrates very high rates of self-reported suicide ideation and significantly high rates of suicide attempts but very low rates of completed suicide

(http://www.cdc.gov/violenceprevention/pdf/Suicide_DataSheet-a.pdf). Thus even significant changes in these proxy measures may have little or no impact on the outcome of interest – completed suicide.

Given this context the impact of decreases in self-reported suicide attempts (Aseltine & DeMartino, 2004; Aseltine et al., 2007) and ideation (Aseltine & DeMartino, 2004) on completed suicide may be minimal if any. Additionally, and of concern, these findings have not been replicated by other researchers, albeit evaluating different interventions. Instead, numerous studies and systematic reviews/meta-analysis have concluded that there is a lack of evidence that universal suicide prevention programs/interventions are able to decrease suicide attempts or ideation (Mann, Apter, Bertolote, Beautrais, Currier, Haas, Hegerl, Lonnqvist, Malone, Marusic, Mehlum, Patton, Phillips, Rutz, Rihmer, Schmidtke, Shaffer, Silverman, Takahashi, Varnik, Wasserman, Yip, & Hendin, 2005a; Ploeg et al., 1996; Robinson, Hetrick, & Martin, 2011; York et al., 2013). It is unclear as to what relationship if any the authors of the SOS intervention studies have to the developers or purveyors of the SOS program. Regardless of that relationship, there is a lack of comparative data carried out by investigators independent of the existing SOS and Yellow Ribbon researchers. The inconsistency of findings between the studies on these programs and the other studies on similar types of programs/interventions warrants further in-depth investigation in both research and practice.

None of the four included studies conducted a thorough analysis on whether the two target programs could cause harmful or negative effects except one methodologically weak study (Aseltine, 2003) which reported that most teachers did not feel the SOS program would cause harm to students. In addition to being non-objective and highly sensitive to confirmation bias, this result raises red flags especially when there is abundant research showing awareness and education may have detrimental effects and serious negative consequences, particularly among children and adolescents (M. D. Cusimano & Sameem, 2011a; Gould et al., 2003; Ploeg et al., 1996; van et al., 2011). This includes more hopelessness and maladaptive coping strategies following the intervention (Kalafat, J., Elias, M., 1994); more negative reactions among high risk youth (Shaffer, D., Vieland, V., Garland, A., Rojas, M., Underwood, MM., Busner, C., 1990), and normalization of suicide behavior upon receiving the intervention (Beautrais et al., 2007). Some studies indicate that this is especially the case among male participants (Lester, 1992; Ploeg et al., 1996; Pompili et al., 2010).

Additionally, none of the available research has evaluated cost/benefit impact of these programs. There is simply no data on this, but this kind of data is very important for educators and school administrators to have in order to make informed decisions about whether to purchase and implement any program. Hence, more research is needed in terms of both the safety and cost effectiveness of both the SOS program and the Yellow Ribbon program to justify their wide dissemination.

It is also very important to note that all the evaluated studies have comparatively poor quality of evidence of effectiveness, meaning there are significant methodological issues embedded in the studies that prevent us from being confident in their conclusions and in recommending the programs for implementation in the school setting. Of particular concern is that none of the four studies demonstrate long-term impact of the intervention or successful external, independent replication. Furthermore, all four studies applied self-report questionnaires without established reliability and validity of the measurement tools. This is especially problematic when using self-report to evaluate suicide attempts or ideation because of participants' tendency to choose socially desirable answers or "fake good" responses. The difference in findings related to self-reports of suicide ideation and suicide attempts can be simply the result of this bias, an example of the Hawthorne effect (McCarney et al., 2007). At the very least, studies of interventions purporting to prevent suicide must use outcome measures independent of self-report of suicide ideation and suicide attempts. Another significant methodological limitation is that the failure to collect baseline pre-intervention data in the two RCT studies may result in bias of research findings caused by the significant differences between the experimental and control group at baseline. In addition, all four studies were conducted in US school settings, and it remains unknown whether the findings can be generalized to settings outside of the US.

Finally, the conclusions of the study authors that either of the programs is an effective suicide prevention strategy is just problematic. No study demonstrated a significant decrease in actual suicide rates related to the implementation of either program. At best, un-validated proxy measures for completed suicide (such as self-report of suicide ideation and suicide attempts) were used. The relationship of changes in these parameters (even if not the result of attributional or confirmation bias) to actual decreases in suicide rates has never been demonstrated and is

statistically unlikely to be robust. Thus the marketing of these programs as universal youth suicide prevention programs is at best questionable.

LIMITATIONS

These considerations notwithstanding, there are a number of limitations to our study. We were unable to conduct meta-analysis to generate effect sizes of both the SOS program and the Yellow Ribbon program because of insufficient number of included studies and poor quality of the studies. Therefore we only did a narrative synthesis of the evidence. We only searched the academic databases for data collection and did not include grey literature, and there may be information missing from unpublished data.

IMPLICATION FOR PRACTICE AND RESEARCH

The SOS program is the only school-based suicide prevention program listed on the Substance Abuse and Mental Health Services Administration's National Registry of Evidence-based Programs and Practices that addresses suicide risk and depression, and reduces suicide attempts. This information as well as highly selective marketing promotion driven advertising is used to "sell" the program and various training interventions to educators and school administrators (for example: http://www.mentalhealthscreening.org/programs/youth-prevention-programs/sos/; accessed April 25, 2013) The Yellow Ribbon program is listed on the Best Practices Registry of National Suicide Prevention Resource Center in US and is marketed to educators and schools as a best practice suicide prevention program (http://www.yellowribbon.org/; accessed April 25, 2013). In Canada it is marketed by the Light for Life Foundation which advertises that the Yellow Ribbon Program follows Canadian National Suicide Prevention Strategy (http://www.yellowribbon.ca/gatekeeper.html). Both programs are adopted and implemented by schools across North America for the purposes of suicide prevention. Yet, neither program has demonstrated that it prevents suicide or that it effectively, safely and cost-effectively addresses numerous other aspects of self-harm behaviors.

The strengths of the promising findings from the included studies are weakened by a great number of inherent methodological issues. Therefore, more advanced research, especially well-designed RCT studies, is needed on these two programs to address their long-term impact and to investigate their application in settings outside of US. More reliable and validated measurement tools should be developed for outcome evaluation, especially for the evaluation of suicide attempts, suicide ideation to mitigate measurement bias. Furthermore, measurement of completed suicides should be considered as the robust measure for the effectiveness of suicide prevention programs for further research.

Youth suicide is a complex phenomenon and it seems unlikely that one single intervention will tackle the problem. It needs a comprehensive approach that involves various domains and different types of preventions, and links the school with families, the wider community, and the health system for better outcomes. Current research in suicide prevention to address suicide attempts, suicide ideation, and completed suicides in children and youth is still extremely limited, which warrants action and investment in this field a top priority.

In conclusion, we do not recommend the SOS program or the Yellow Ribbon program to schools or the community to reduce suicide attempts, suicide ideation and completed suicides; or to promote help-seeking behaviors at this stage due to the inconsistency of research findings in this field and methodological limitations of the included studies. While it may be appropriate to apply these two programs to educate youth about depression and suicide, there is no evidence that these programs are better, worse or the same compared to other less expensive interventions such as embedding information about depression and suicide in regular school curriculum.

References

Aseltine, R. H.,Jr. (2003). An evaluation of a school based suicide prevention program. Adolescent & Family Health, 3(2), 81-88.

Aseltine, R. H.,Jr, & DeMartino, R. (2004). An outcome evaluation of the SOS Suicide Prevention Program. American Journal of Public Health, 94(3), 446-451.

Aseltine, R. H.,Jr, James, A., Schilling, E. A., & Glanovsky, J. (2007). Evaluating the SOS suicide prevention program: a replication and extension. BMC Public Health, 7, 161.

Beautrais, A., Fergusson, D., Coggan, C., Collings, C., Doughty, C., Ellis, P., et al. (2007). Effective strategies for suicide prevention in New Zealand: a review of the evidence. The New Zealand Medical Journal, 120(1251), U2459.

Canadian Council on Learning. (2009). A barrier to learning: Mental health disorders among Canadian youth. Retrieved January/25th, 2010, from http://www.ccl-cca.ca/CCL/Reports/LessonsInLearning/LinL200900415MentalhealthBarrier.htm

Cusimano, M. D., & Sameem, M. (2011a). The effectiveness of middle and high school-based suicide prevention programmes for adolescents: a systematic review. Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention, 17(1), 43-49.

Cusimano, M. D., & Sameem, M. (2011b). The effectiveness of middle and high school-based suicide prevention programmes for adolescents: A systematic review. Injury Prevention, 17(1), 43-49.

Farmer, E.M., Burns, B.J., Phillips, S.D., Angold, A., & Costello, E.J. (2003). Pathways into and through mental health services for children and adolescents. Psychiatric Services, 54(1), 60-66.

Forman, S.G., Olin, S.S., Hoagwood, K.E., Crowe, M., & Saka, N. (2009). Evidence-based intervention in schools: Developers' views on implementation barriers and facilitators. School Mental Health, 1, 26-36.

Freedenthal, S. (2010). Adolescent help-seeking and the Yellow Ribbon Suicide Preention Program: An evaluation. Suicide & Life-Threatening Behavior, 40(6), 628-639.

Gould, M. S., Greenberg, T., Velting, D. M., & Shaffer, D. (2003). Youth suicide risk and preventive interventions: A review of the past 10 years. Journal of the American Academy of Child & Adolescent Psychiatry, 42(4), 386-405.

Hallfors, D., & Godette, D. (2002). Will the 'principles of effectiveness' improve prevention practice? Early findings from a diffusion study. Health Education Research: Theory & Practice, 17, 461-470.

Kalafat, J., Elias, M. (1994). An evaluation of a school-based suicide awareness intervention. Suicide & Life-Threatening Behavior, 24, 224-233.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 593-602.

Kutcher, S. (2011). Facing the challenge of care for child and youth mental health in Canada: A critical commentary, five suggestions for change and a call to action. Health Care Quarterly, 14, 15-21.

Kutcher, S. P., & McLuckie, A. for the Child and Youth Advisory Committee, Mental Health Commission of Canada. (2010). Evergreen: A child and youth mental health framework for Canada. Calgary, AB: Mental Health Commission of Canada.

Lester, D. (1992). Religiosity, suicide and homicide: a cross-national examination. Psychol Rep, 71, 1282.

Mann, J. J., Apter, A., Bertolote, J., Beautrais, A., Currier, D., Haas, A., et al. (2005a). Suicide prevention strategies: a systematic review. JAMA : The Journal of the American Medical Association, 294(16), 2064-2074.

Mann, J. J., Apter, A., Bertolote, J., Beautrais, A., Currier, D., Haas, A., et al. (2005b). Suicide prevention strategies: a systematic review. JAMA : The Journal of the American Medical Association, 294(16), 2064-2074.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. BMC Medical Research Methodology, 7, 30.

National Criminal Justice Reference Services. (2005). The Office of Justice Program (OJP) What Works Repository. Retrieved from: https://www.ncjrs.gov/pdffiles1/nij/220889.pdf

Ploeg, J., Ciliska, D., Dobbins, M., Hayward, S., Thomas, H., & Underwood, J. (1996). A systematic overview of adolescent suicide prevention programs. Canadian Journal of Public Health.Revue Canadienne De Sante Publique, 87(5), 319-324.

Pompili, M., Innamorati, M., Vichi, M., Masocco, M., Vanacore, N., Lester, D., et al. (2010). Suicide prevention among youths. Systematic review of available evidence-based interventions and implications for Italy. Minerva Pediatrica, 62(5), 507-535.

Robinson, J., Hetrick, S. E., & Martin, C. (2011). Preventing suicide in young people: systematic review. The Australian and New Zealand Journal of Psychiatry, 45(1), 3-26.
Roeser, R.W., & Midgley, C. (1997). Teachers' views of issues involving students' mental health. The Elementary School Journal, 98(2), 115-133.
Rones, M., & Hoagwood, K.E. (2000). School-based mental health services: A review. Clinical Child and Family Psychology Review, 3(4), 223-241.
Shaffer, D., Vieland, V., Garland, A., Rojas, M., Underwood, MM., Busner, C. (1990). Adolescent suicide attempters: response to suicide-prevention programs. JAMA : The Journal of the American Medical Association, 264, 3151-3155.
U.S. Department of Health and Human Services. (1999). Mental Health: A report of the surgeon general. Rockville, MD: U.S. Department of Health and Human Services.
van, d. F., Sarchiapone, M., Postuvan, V., Volker, D., Roskar, S., Grum, A. T., et al. (2011). Best practice elements of multilevel suicide prevention strategies: A review of systematic reviews. Crisis: The Journal of Crisis Intervention and Suicide Prevention, 32(6), 319-333.
Waddell, C., Offord, D. R., Shepherd, C. A., Hua, J. M., & McEwan, K. (2002). Child psychiatric epidemiology and Canadian public policy-making: The state of the science and the art of the possible. The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie, 47(9), 825-832.
Wei, Y., Kutcher, S., & Szumilas, M. (2011). Comprehensive School Mental Health: An Integrated "School-Based Pathway to Care" Model for Canadian Secondary Schools. McGill Journal of Education, 46(2), 213-229.
WHO Regional Office for Europe. (1996). Regional guidelines: development of health-promoting schools: a framework for action. Manila, WHO Regional Office for the Western Pacific: Retrieved from http://whqlibdoc.who.int/wpro/1994-99/a53203.pdf
Wyn, J., Cahill, H., Holdsworth, R., Rowling, L., & Carson, S. (2000). MindMatters, a whole-school approach promoting mental health and wellbeing. Australian and New Zealand Journal of Psychiatry, 34, 594-601.
York, J., Lamis, D. A., Friedman, L., Berman, A. L., Joiner, T. E., Mcintosh, J. L., et al. (2013). A systematic review process to evaluate suicide prevention programs: A sample case of community-based programs. Journal of Community Psychology, 41(1), 35-51.

APPENDIX 1: SEARCH STRATEGIES

1. SUICIDE WILL BE SEARCHED USING THE TERMS

Pubmed:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*)) OR "Suicide"[Mesh] OR "Suicide, Attempted"[Mesh]

PsycINFO:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

EMBASE:
suicid* OR (“self harm”) OR (self NEAR/3 (injur* OR behav*)) OR (self NEAR/3 (hatred OR mutilate* OR injur* OR inflict*))

CINAHL:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Cochrane:
suicid* OR “self harm” OR (self NEAR/3 (injur* OR behav*)) OR (self NEAR/3 (hatred OR mutilate* OR injur* OR inflict*))

SocIndex:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Sociological Abstracts:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Social Services Abstracts:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

ERIC:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Social Work Abstracts
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Research Library:
Suicid* OR “self harm” OR (self N3 (injur* OR behav*)) OR (self N3 (hatred OR mutilate* OR injur* OR inflict*))

Web of Science:
suicid* OR “self harm” OR (self NEAR/3 (injur* OR behav*)) OR (self NEAR/3 (hatred OR mutilate* OR injur* OR inflict*))

2. INTERVENTIONS WILL BE SEARCHED USING THE TERMS:

Pubmed and PsycINFO:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

EMBASE:
Prevent* OR (interven* NEAR/3 (crisis OR crises) OR health NEAR/3 (service* OR educat*) OR program* NEAR/3 (develop* OR evaluat*) OR experiment* OR trial* OR effective* OR efficac*

CINAHL:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

The Cochrane Library:
Prevent* OR (interven* NEAR/3 (crisis OR crises)) OR (health NEAR/3 (service* OR educat*)) OR (program* NEAR/3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

SocIndex:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Sociological Abstracts:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Social Services Abstracts:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

ERIC:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Social Work Abstracts
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Research Library:
Prevent* OR (interven* N3 (crisis OR crises)) OR (health N3 (service* OR educat*)) OR (program* N3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Web of Science:
Prevent* OR (interven* NEAR/3 (crisis OR crises)) OR (health NEAR/3 (service* OR educat*)) OR (program* NEAR/3 (develop* OR evaluat*)) OR experiment* OR trial* OR effective* OR efficac*

Table 1: Study characteristics

Author/study	Aseltine 2003	Aseltine & DeMartino 2004	Aseltine et al. 2007	Freedenthal 2010
Sample	68	2100	4133	70 (teachers); 210 (students)
Country	US	US	US	US
Population	School staff	High school students	High school students	High school students and teachers
Design	Survey	RCT*	RCT*	Quasi-experimental
Intervention	SOS program combines curricula to raise awareness of suicide and its related issues with a brief screening for depression and other risk factors associated with suicidal behavior.	SOS program combines curricula to raise awareness of suicide and its related issues with a brief screening for depression and other risk factors associated with suicidal behavior.	SOS program combines curricula to raise awareness of suicide and its related issues with a brief screening for depression and other risk factors associated with suicidal behavior.	Yellow Ribbon Program includes school wide assemblies, peer leadership training for students, staff training for adult gatekeepers such as high school teachers, community presentations, and local chapters that provide outreach and education.
Duration	5.3 days on average	Not described	2 days	60 minutes (student training); 1.5 hours (staff training); 50 minutes (school assembly)
Comparison	No	Yes	Yes	Yes
Follow-up	Post intervention	3 month post intervention	3 months post intervention	Post intervention
Outcome	Overview of Implementation; Ratings of and Reactions to the Program; help-seeking behaviors; Costs and benefits of the program	Knowledge; Attitudes; Help-seeking behaviors: Suicide attempts; Suicide ideation	Knowledge; Attitudes; Help-seeking behaviors: Suicide attempts; Suicide ideation	Help-seeking behaviors

RCT*: Randomized Controlled Trial

Table 2: Summary of study results

Author/study	Aseltine 2003	Aseltine & DeMartino 2004	Aseltine et al. 2007	Freedenthal 2010
Knowledge	Not assessed	M*=6.49 (C); M=7.18* (T)	M=4.36 (C); M=5.00** (T)	Not assessed
Attitudes	Not assessed	M=3.80 (C); M=4.05** (T)	M=3.83 (C); M=3.99** (T)	Not assessed
Help-seeking	60% (increase of help-seeking compared to the previous year)	Asked help for friends; for oneself: 13%; 18.7% (C) 11.9%; 15.9% (T)	12.4% (C); 12.8% (T)	Staff data: +2.3% (within-school); +2.6% (between-school) Student data: M=12.0 (pre-intervention) M=10.0** (post-intervention)
Suicide attempts	Not assessed	5.4% (C); 3.6%** (T)	4.5% (C); 3.0%** (T)	Student disclosed suicide attempts (Staff data): 7.7% (within-school); -15.6%*** (between-school) Student reported peer suicide attempts (Staff data): 7.8% (with-school); -21.1% (between-school)
Suicide ideation	Not assessed	12.2% (C); 10.1% (T)	Not assessed	Student disclosed suicidal thoughts (staff data): +4.5% (within-school); -4.9%** (between-school) Student reported peer suicidal thoughts (staff data): -0.9% (within-school); -2.5% (between-school)
Other	Implementation: 74% (in health class); 40%-45% students receiving the entire program Program rating: 64%-87% (very or somewhat effective) Costs and benefits: 19% (adverse effects) 88% (somewhat or very helpful)	Not assessed	Not assessed	

*: M=mean, C=control, T=treatment; **: p<.05, indicating statistically significant results; ***: p<.001, indicating statistically significant results.

| APPENDIX 2: CRITICAL APPRAISAL (ASELTINE, 2003)

Levels of evidence of effectiveness

Criteria	Criteria Description	Judgment
Significant effect	Rigorous statistical evidence of a change (effect size ≥ .20; or statistical significance $p \leq .05$)	Y ☐ N ☐ U √
Sustained effect	A prevention effect that endures beyond the end of the intervention for at least one year, or a treatment effect that endures for at least two years after entering the program)	Y ☐ N ☐ U √
≥ 1 successful external replication	The program was found effective in randomized controlled trials conducted in at least two implementation sites by different implementation teams	Y ☐ N √ U ☐
Study design and execution	1. random assignment	Y ☐ N √ U ☐
	2. large, representative sample that minimizes selection bias (300-400)	Y ☐ N √ U ☐
	3. description of intervention, including clear explanation of scientific background, the logic of the intervention, specific objectives/hypotheses and primary secondary outcome measures	Y √ N ☐ U ☐
	4. independent evaluation	Y ☐ N ☐ U √
	5. adequate outcome measurement (i.e. outcomes are behavioral or distal in nature and are consistent across multiple measures)	Y ☐ N √ U ☐
	6. description of differences	Y ☐ N ☐ U √

	7. statistical significance ($p \leq .05$)	Y ☐ N √ U ☐
	8. identification of important adverse effects/events	Y ☐ N ☐ U √
	9. modest attrition (≤ 25%)	Y √ N ☐ U ☐
	10. intent-to-treat analytic approach	Y ☐ N √ U ☐
	11. accurate interpretation of the results	Y √ N ☐ U ☐
Other important factors	1. Evidence of change in risk/protective factors	Y ☐ N ☐ U √
	2. cost information	Y ☐ N ☐ U √
	3. cost-benefit estimates	Y ☐ N ☐ U √
	4. measurement of potential side-effects or negative effects	Y ☐ N ☐ U √

Ratings: Effective ⟶ Effective with Reservation ⟶ Promising ⟶ Inconclusive Evidence ⟶ Insufficient Evidence ⟶ Ineffective

Final Rating: Insufficient evidence (A program with a quasi-experimental design that lacks sufficient methodological rigor, a pre-post test design, or a purely descriptive evaluation.)

Levels of readiness for dissemination

Criteria	Criteria Description	Judgment
Training and Related Support Materials	Detailed curriculum	Y ☐ N √ U ☐
	Prepared trainers and technical experts	Y √ N ☐ U ☐
	Supportive informational materials	Y ☐ N √ U ☐
	Operations manuals	Y √ N ☐ U ☐
	Implementation guides	Y ☐ N √ U ☐
	Case studies	Y ☐ N √ U ☐
	Evidence of change in risk/protective factors	Y ☐ N ☐ U √
	Cost information and cost- benefit estimate	Y ☐ N ☐ U √
	Effectiveness indicators	Y √ N ☐ U ☐
	Other support materials employing a variety of educational mediums, such as videotapes, audiotapes, or interactive Web-based programs, all of which are field tested	Y ☐ N √ U ☐
Technical Assistance Support	Following the provision of training experts are available on-site or online to provide specific guidance related to the implementation of the intervention techniques, problem solving, and modifications as necessary and appropriate	Y ☐ N √ U ☐
Informational Materials	Supplemental guidance provided over time through newsletters, Web sites, and other mediums to inform regarding innovations	Y ☐ N √ U ☐

	made in other sites, methods to enhance implementation, operations management and assessment procedures and practices	
Quality Control for Implementation	Procedures for ensuring that the intervention is implemented with fidelity to the original design. These may include clinical supervision, review of tape recordings of intervention sessions, or other methods	Y ☐ N √ U ☐

Ratings: Fully Prepared for Widespread dissemination ⟶ Fully Prepared for Limited Dissemination ⟶ Not Prepared for Widespread Dissemination

Final rating: Not Prepared for Widespread Dissemination

| APPENDIX 3: CRITICAL APPRAISAL (ASELTINE & DeMARTINO, 2004)

Levels of evidence of effectiveness

Criteria	Criteria Description	Judgment
Significant effect	Rigorous statistical evidence of a change (effect size ≥ .20; or statistical significance $p \leq .05$)	Y √ N ☐ U ☐
Sustained effect	A prevention effect that endures beyond the end of the intervention for at least one year, or a treatment effect that endures for at least two years after entering the program)	Y ☐ N √ U ☐
≥ 1 successful external replication	The program was found effective in randomized controlled trials conducted in at least two implementation sites by different implementation teams	Y ☐ N √ U ☐
Study design and execution	12. random assignment	Y √ N ☐ U ☐
	13. large, representative sample that minimizes selection bias (300-400)	Y √ N ☐ U ☐
	14. description of intervention, including clear explanation of scientific background, the logic of the intervention, specific objectives/hypotheses and primary secondary outcome measures	Y √ N ☐ U ☐
	15. independent evaluation	Y ☐ N ☐ U √
	16. adequate outcome measurement (i.e. outcomes are behavioral or distal in nature and are consistent across multiple measures)	Y √ N ☐ U ☐
	17. description of differences	Y ☐ N ☐ U √
	18. statistical significance ($p \leq .05$)	Y √ N ☐ U ☐
	19. identification of important adverse effects/events	Y ☐ N ☐ U √

	20. modest attrition (≤ 25%)	Y √ N ☐ U ☐
	21. intent-to-treat analytic approach	Y √ N ☐ U ☐
	22. accurate interpretation of the results	Y √ N ☐ U ☐
Other important factors	5. Evidence of change in risk/protective factors	Y ☐ N ☐ U √
	6. cost information	Y ☐ N √ U ☐
	7. cost-benefit estimates	Y ☐ N √ U ☐
	8. measurement of potential side-effects or negative effects	Y ☐ N √ U ☐

Ratings: Effective → Effective with Reservation → Promising → Inconclusive Evidence → Insufficient Evidence → Ineffective

Final Rating: Inconclusive evidence (Programs with adequately rigorous experimental or quasi-experimental designs that lack sustained effects.)

Levels of readiness for dissemination

Criteria	Criteria Description	Judgment
Training and Related Support Materials	Detailed curriculum	Y ☐ N √ U ☐
	Prepared trainers and technical experts	Y √ N ☐ U ☐
	Supportive informational materials	Y ☐ N √ U ☐
	Operations manuals	Y √ N ☐ U ☐
	Implementation guides	Y ☐ N √ U ☐
	Case studies	Y ☐ N √ U ☐
	Evidence of change in risk/protective factors	Y √ N ☐ U ☐
	Cost information and cost- benefit estimate	Y ☐ N ☐ U √
	Effectiveness indicators	Y √ N ☐ U ☐
	Other support materials employing a variety of educational mediums, such as videotapes, audiotapes, or interactive Web-based programs, all of which are field tested	Y ☐ N √ U ☐
Technical Assistance Support	Following the provision of training experts are available on-site or online to provide specific guidance related to the implementation of the intervention techniques, problem solving, and modifications as necessary and appropriate	Y ☐ N √ U ☐
Informational Materials	Supplemental guidance provided over time through newsletters, Web sites, and other mediums to inform regarding innovations made in other sites, methods to enhance implementation,	Y ☐ N √ U ☐

	operations management and assessment procedures and practices	
Quality Control for Implementation	Procedures for ensuring that the intervention is implemented with fidelity to the original design. These may include clinical supervision, review of tape recordings of intervention sessions, or other methods	Y ☐ N √ U ☐

Ratings: Fully Prepared for widespread Dissemination ⟶ Fully Prepared for Limited Dissemination ⟶ Not Prepared for Widespread Dissemination

Final rating: Not Prepared for Widespread Dissemination

| APPENDIX 4: CRITICAL APPRAISAL (ASELTINE ET AL., 2007)

Levels of evidence of effectiveness

Criteria	Criteria Description	Judgment

Significant effect	Rigorous statistical evidence of a change (effect size ≥ .20; or statistical significance $p \leq .05$)	Y ☐ N ☐ U √
Sustained effect	A prevention effect that endures beyond the end of the intervention for at least one year, or a treatment effect that endures for at least two years after entering the program)	Y ☐ N √ U ☐
≥ 1 successful external replication	The program was found effective in randomized controlled trials conducted in at least two implementation sites by different implementation teams	Y ☐ N √ U ☐
Study design and execution	23. random assignment	Y √ N ☐ U ☐
	24. large, representative sample that minimizes selection bias (300-400)	Y √ N ☐ U ☐
	25. description of intervention, including clear explanation of scientific background, the logic of the intervention, specific objectives/hypotheses and primary secondary outcome measures	Y √ N ☐ U ☐
	26. independent evaluation	Y ☐ N ☐ U √
	27. adequate outcome measurement (i.e. outcomes are behavioral or distal in nature and are consistent across multiple measures)	Y √ N ☐ U ☐
	28. description of differences	Y ☐ N ☐ U √
	29. statistical significance ($p \leq .05$)	Y ☐ N ☐ U √
	30. identification of important adverse effects/events	Y ☐ N ☐ U √

	31. modest attrition (≤ 25%)	Y √ N ☐ U ☐
	32. intent-to-treat analytic approach	Y √ N ☐ U ☐
	33. accurate interpretation of the results	Y √ N ☐ U ☐
Other important factors	9. Evidence of change in risk/protective factors	Y ☐ N ☐ U √
	10. cost information	Y ☐ N √ U ☐
	11. cost-benefit estimates	Y ☐ N √ U ☐
	12. measurement of potential side-effects or negative effects	Y ☐ N √ U ☐

Ratings: Effective ⟶ Effective with Reservation ⟶ Promising ⟶ Inconclusive Evidence ⟶ Insufficient Evidence ⟶ Ineffective

Final Rating: Inconclusive evidence (Programs with adequately rigorous experimental or quasi-experimental designs that lack sustained effects.)

Levels of readiness for dissemination

Criteria	Criteria Description	Judgment
Training and Related Support	Detailed curriculum	Y ☐ N √ U ☐

Materials		
	Prepared trainers and technical experts	Y √ N ☐ U ☐
	Supportive informational materials	Y ☐ N √ U ☐
	Operations manuals	Y √ N ☐ U ☐
	Implementation guides	Y ☐ N √ U ☐
	Case studies	Y ☐ N √ U ☐
	Evidence of change in risk/protective factors	Y √ N ☐ U ☐
	Cost information and cost- benefit estimate	Y ☐ N ☐ U √
	Effectiveness indicators	Y √ N ☐ U ☐
	Other support materials employing a variety of educational mediums, such as videotapes, audiotapes, or interactive Web-based programs, all of which are field tested	Y ☐ N √ U ☐
Technical Assistance Support	Following the provision of training experts are available on-site or online to provide specific guidance related to the implementation of the intervention techniques, problem solving, and modifications as necessary and appropriate	Y ☐ N √ U ☐
Informational Materials	Supplemental guidance provided over time through newsletters, Web sites, and other mediums to inform regarding innovations made in other sites, methods to enhance implementation, operations management and assessment procedures and practices	Y ☐ N √ U ☐

Quality Control for Implementation	Procedures for ensuring that the intervention is implemented with fidelity to the original design. These may include clinical supervision, review of tape recordings of intervention sessions, or other methods	Y ☐ N √ U ☐

Ratings: Fully Prepared for widespread Dissemination ⟶ Fully Prepared for Limited Dissemination ⟶ Not Prepared for Widespread Dissemination

Final rating: Not Prepared for Widespread Dissemination

| APPENDIX 5: CRITICAL APPRAISAL (FREEDENTHAL 2010)

Levels of evidence of effectiveness

Criteria	Criteria Description	Judgment
Significant effect	Rigorous statistical evidence of a change (effect size ≥ .20; or statistical	Y ☐ N √ U ☐

	significance $p \leq .05$)	
Sustained effect	A prevention effect that endures beyond the end of the intervention for at least one year, or a treatment effect that endures for at least two years after entering the program)	Y ☐ N √ U ☐
≥ 1 successful external replication	The program was found effective in randomized controlled trials conducted in at least two implementation sites by different implementation teams	Y ☐ N √ U ☐
Study design and execution	34. random assignment	Y ☐ N √ U ☐
	35. large, representative sample that minimizes selection bias (300-400)	Y ☐ N √ U ☐
	36. description of intervention, including clear explanation of scientific background, the logic of the intervention, specific objectives/hypotheses and primary secondary outcome measures	Y √ N ☐ U ☐
	37. independent evaluation	Y ☐ N ☐ U √
	38. adequate outcome measurement (i.e. outcomes are behavioral or distal in nature and are consistent across multiple measures)	Y √ N ☐ U ☐
	39. description of differences	Y ☐ N ☐ U √
	40. statistical significance ($p \leq .05$)	Y ☐ N √ U ☐
	41. identification of important adverse effects/events	Y ☐ N ☐ U √
	42. modest attrition (≤ 25%)	Y ☐ N ☐ U √

	43. intent-to-treat analytic approach	Y ☐ N √ U ☐
	44. accurate interpretation of the results	Y √ N ☐ U ☐
Other important factors	13. Evidence of change in risk/protective factors	Y ☐ N √ U ☐
	14. cost information	Y ☐ N √ U ☐
	15. cost-benefit estimates	Y ☐ N √ U ☐
	16. measurement of potential side-effects or negative effects	Y ☐ N √ U ☐

Ratings: Effective ⟶ Effective with Reservation ⟶ Promising ⟶ Inconclusive Evidence ⟶ Insufficient Evidence ⟶ Ineffective

Final Rating: Ineffective (A program with an experimental or quasi-experimental research design that in an initial study and at least one replication failed to demonstrate a significant effect.)

Levels of readiness for dissemination

Criteria	Criteria Description	Judgment
Training and Related Support Materials	Detailed curriculum	Y ☐ N √ U ☐
	Prepared trainers and technical experts	Y √ N ☐ U ☐

	Supportive informational materials	Y ☐ N √ U ☐
	Operations manuals	Y ☐ N √ U ☐
	Implementation guides	Y ☐ N √ U ☐
	Case studies	Y ☐ N √ U ☐
	Evidence of change in risk/protective factors	Y ☐ N √ U ☐
	Cost information and cost- benefit estimate	Y ☐ N √ U ☐
	Effectiveness indicators	Y ☐ N √ U ☐
	Other support materials employing a variety of educational mediums, such as videotapes, audiotapes, or interactive Web-based programs, all of which are field tested	Y ☐ N √ U ☐
Technical Assistance Support	Following the provision of training experts are available on-site or online to provide specific guidance related to the implementation of the intervention techniques, problem solving, and modifications as necessary and appropriate	Y ☐ N √ U ☐
Informational Materials	Supplemental guidance provided over time through newsletters, Web sites, and other mediums to inform regarding innovations made in other sites, methods to enhance implementation, operations management and assessment procedures and practices	Y ☐ N √ U ☐
Quality Control for Implementation	Procedures for ensuring that the intervention is implemented with fidelity to the original design. These may include clinical	Y ☐ N √ U ☐

	supervision, review of tape recordings of intervention sessions, or other methods	

Ratings: Fully Prepared for widespread Dissemination ⟶ Fully Prepared for Limited Dissemination ⟶ Not Prepared for Widespread Dissemination

Final rating: Not Prepared for Widespread Dissemination