BRIEF REPORTS

Long-term Outcome of a School-based, Universal Approach to Prevention of Depression in Adolescents

Susan H. Spence, Jeanie K. Sheffield, and Caroline L. Donovan

School of Psychology, University of Queensland

Brisbane, Australia

Keywords: Depression, prevention, adolescence, problem-solving

Abstract

This study examined the 2-, 3- and 4-year outcomes of a school-based, universal approach to the prevention of adolescent depression. Despite initial short-term positive effects these benefits were not maintained over time. Adolescents who completed the teacher-administered cognitive-behavioral intervention did not differ significantly from monitoring only controls in terms of changes in depressive symptoms, problem solving, attributional style or other indicators of psychopathology from pre-intervention to 4-year follow-up. Results were equivalent irrespective of initial level of depressive symptoms.

Long-term outcome of a School-based, Universal Approach to Prevention of Depression in Adolescents

To date, there has been minimal research into the long-term effectiveness of school-based approaches to the prevention of depression in young people, with follow-ups rarely extending beyond 12 months. The present study examined the 2-, 3- and 4-year follow-up results of a universal, school-based, cognitive-behavioral intervention designed to prevent the development of depression in youth. The original study was reported by Spence, Sheffield and Donovan (2003) for the post-intervention and 1-year follow-up results. It evaluated the impact of a brief cognitive-behavioral intervention that was administered by teachers on a whole-classroom basis. The original study comprised 1500 Grade 8 students, attending 16 high schools that were matched in pairs according to demographic features and randomly assigned to either the active intervention or a monitoring control condition. The effects of the program were examined separately for those with elevated depression scores (labelled “high risk”) and those whose initial depression scores fell below this cut-off (labelled “low risk”). The high-risk group who completed the preventive intervention demonstrated a significantly greater decline in depression scores and a greater increase in total problem solving scores over the course of the program compared to the monitoring control group. However, by 12-month follow-up there were no longer any significant differences between intervention and control groups on measures of depression, emotional well-being, social adjustment, attributional style or problem solving.

Results for the low-risk group suggested that while low-risk control participants demonstrated a slight increase in depressive symptom scores, those completing the preventive intervention demonstrated a slight decline from pre- to post-intervention. In addition, the low-risk intervention participants demonstrated significantly greater improvement in total problem-solving scores compared to students in the control condition. For both high- and low-risk participants, those who reported greater reductions in depression

scores tended to be those who reported greater improvements in problem-solving. Despite these positive short-term effects, the benefits were not maintained for the low-risk students at 12-month follow-up.

It was argued that the initial positive effects of the intervention, although not evident at 12-month follow-up, may emerge at a later stage in the adolescents' development when the skills and knowledge taught by the program became of considerable importance in helping young people to manage the life challenges of later adolescence. Indeed, such delayed effects have been noted in some longer-term evaluations of depression prevention programs. For example, Gillham et al. (1995) reported that while no positive effects were evident for an indicated prevention program at 12-month follow-up, the intervention group showed significantly lower levels of depressive symptoms at 2-year follow-up. This effect was not maintained however at the 3-year evaluation. Delayed benefits were also noted by Quayle et al. (2001) for a universal intervention for prevention of depression. Although no positive effects were found immediately after intervention, the intervention group reported significantly lower levels of depressive symptoms compared to controls at 6-month follow-up.

Although it is clearly important to examine whether delayed long-term effects were evident in the present study, the possibility must also be considered that this type of universal intervention is not effective in producing significant long-term preventive effects upon depression. Recent studies that have examined the impact of universal prevention strategies designed to prevent depression in young people (Cardemil, 2002; Clarke, Hawkins, Murphy, & Sheeber, 1993; Pattison & Lynd Stevenson, 2001; Quayle et al., 2001; Shatté, 1997; Shochet et al., 2001; Spence et al., 2003) have produced mixed findings. In a recent meta-analysis, Merry, McDowell, Hetrick, Bir and Muller (2003) concluded that effect sizes are very weak for universal preventive interventions for adolescent depression. Furthermore, the universal studies included in the meta-analysis generally failed to include follow-up periods

of 12 months or more. Those that have done so have failed to demonstrate longer-term preventive effects (Shatt, 1997; Spence et al., 2003). In addition, most studies suffered a range of methodological limitations, such as inadequate sample size, low participation rates, high drop out rates, and lack of randomisation of participants or schools to conditions.

Given the importance of follow-ups that are of sufficient duration to establish the long-term effectiveness of preventive efforts, the present study reports the 2-, 3- and 4-year follow-up effects of a universal approach to the prevention of depression in adolescence.

Method

Full details of the methodology, participants, measures, procedures and program content are provided in the original paper that outlined the short-term evaluation (Spence et al., 2003).

Participants

At the commencement of the study, participants were 1500 Grade 8 students, ranging in age from 12 to 14 years who attended one of 16 participating co-educational high schools. Schools were matched in pairs based on State versus private funding status and size of enrolment. All schools were urban, co-educational, and had approximately equal numbers of boys and girls. Of the 16 schools, matching was possible for 7 out of 8 pairs. Schools were then randomly assigned from each pair to either the universal prevention or a monitoring control condition. Written, informed consent was obtained from 66% (N = 1500) of 2272 students and their parents who were invited to participate in the evaluation component of the study. However, the universal program was completed by all students within participating classrooms as part of the regular school curriculum.

A total of 751 students (47.5% male and 52.5% female) took part in the universal prevention program and 749 students (49.4% male and 50.6% female) in the monitoring control condition. The average student age at commencement of the study was 12.82 years (SD = .54) and 12.90 years (SD=.53) for the intervention and control conditions respectively.

The High-Symptom Sample

The “universal” level impact of the intervention was evaluated for the full-sample of students using self-report questionnaires. However, the effects were also examined separately and in more depth using clinical interviews for those who reported elevated symptoms at the start of the study. It was proposed that the elevated symptom group would be at greater risk for the development of depression. This approach enabled channelling of limited project resources to more in-depth assessment and interviewing of the high-risk students as funds did not permit interviewing the full sample of students.

Students who reported initial elevated symptoms of depression were classified as “high-symptom” based on pre-test responses on the Beck Depression Inventory (BDI; Beck, Rush, Shaw, & Emery, 1979) or Dysthymia questions (see below). Specifically, students were categorized as high-symptom if they scored greater than or equal to 13 on the BDI, and/or endorsed the suicide question on the BDI and/or endorsed the frequency and duration Dysthymia questions. The BDI cut-off of 13 was used given that this score represents the cut-off for the top 20-25% of BDI scores for adolescents (Roberts, Lewinsohn, & Seeley, 1991) and approximately 25% of participants in the present study.

All high-symptom students were interviewed at each assessment occasion using the affective disorders sections of the Anxiety Disorders Interview Schedule for Children (ADIS-C; Silverman & Albano, 1996). In the intervention condition, 27.1% of the sample was categorized as high-symptom, compared to 26% of the control group. Clinical interviews revealed that 3.5% of the total and 12.7% of the high-symptom students in the intervention group and 2.3% of the total and 8.7% of the high-symptom students in the control group met diagnostic criteria for an affective disorder at pre-intervention.

Measures

Beck Depression Inventory (BDI; Beck et al., 1979). Several studies have shown the BDI to have acceptable internal consistency, test-retest reliability and construct validity with

adolescents (see Reynolds, 1994 for a review). In the present study, minor modifications were made to the original BDI, in accordance with other research studies that have used the BDI with adolescent samples (Bennett et al., 1997; Carter & Dacey, 1996; Reynolds, 1994). These changes included removal of the sexual interest item at the request of school ethics committees and simplification of terminology in item 11 (the word ‘annoyed’ was used to define irritable) to aid understanding. The internal consistency of the scale for the full-sample at entry to the study was .90, $N = 1500$.

Dysthymia. We were concerned that some adolescents may experience symptoms of dysthymia rather than depression that would not be identified by high scores on the BDI. Thus four questions were included in the questionnaire battery to reflect *DSM-IV* diagnostic criteria for dysthymia in adolescents. Students who indicated high frequency and duration of sadness and/or irritability over the past year, consistent with *DSM-IV* criteria for dysthymia were also categorized as high-symptom and selected for clinical interview along with those who reported elevated symptoms of depression on the BDI. The dysthymia questions were included only as a screen for initial high symptom risk and were not used as an outcome measure.

Social Problem-Solving Inventory-Revised Short Form (SPSI-R; D’Zurilla, Nezu, & Maydeu-Olivares, 1997). The total problem solving score has been shown to have strong internal consistency and test-retest reliability, with higher SPSI-R scores being indicative of higher problem-solving ability. The internal consistency of the scale for the full-sample at entry to the study was .81, $N = 1500$.

Children’s Attributional Style Questionnaire-Revised (CASQ-R; Seligman et al., 1984). The total score of the 24-item short-form CASQ-R was used to assess explanatory style, with higher scores indicating more optimistic attributional style. The psychometric properties of the CASQ-R have been shown to be acceptable, but not strong, with moderate internal consistency and fair test-retest reliability (Thompson, Kaslow, Weiss, & Nolen-

Hoeksema, 1998). The internal consistency of the scale for the full-sample at entry to the study was .65, $N = 1500$.

Child and Adolescent Social and Adaptive Functioning Scale (CASAFS; Price, Spence, Sheffield, & Donovan, 2002). The CASAFS was completed as an indicator of social functioning. The scale has high internal consistency and acceptable test-retest reliability, with confirmatory factor analysis supporting its construct validity (Price et al., 2002). The internal consistency of the scale for the full-sample at entry to the study was .81, $N = 1500$.

Youth Self-Report form of the Child Behavior Checklist (YSR; Achenbach, 1991). The YSR was used to assess general child psychopathology. It has been widely used in the literature and has well-established psychometric properties. The internal consistency of the scale for the full-sample at entry to the study was .91 for the internalising and .89 for the externalising subscales, $N = 1500$.

Anxiety Disorders Interview Schedule for Children (ADIS-C; Silverman & Albano, 1996). The ADIS-C is a semi-structured interview based on DSM-IV diagnostic criteria. The interview protocol has a separate section relating to affective disorders and includes diagnostic criteria for major depressive disorder, dysthymia and bipolar disorders, based on DSM-IV criteria. Inter-rater reliability of ADIS-C diagnoses and test-retest reliability has been found to be good (Silverman & Eisen, 1992; Silverman & Nelles, 1988; Silverman, Saavedra, & Pina, 2001). Although a parallel parent version of the interview is available, resources allowed only for the child interview. Details relating to training of interviewers and inter-rater reliability are reported in Spence et al. (2003).

Survival Analysis. To track new or continuing episodes of depression during the follow-up periods, high-symptom students were interviewed at 6-months, 1-year, 18-months, 2-year, 30-month, 3-year and 4-year follow-ups using the Longitudinal Follow-up Evaluation (LIFE; Shapiro & Keller, 1979) in conjunction with the diagnostic interview. The LIFE interview comprises a series of structured prompts to identify episodes of depression over the

previous 6 or 12 months. This technique has been used in previous longitudinal research with adolescents and has been shown to have high inter-assessor reliability (Clarke et al., 1995). A more detailed description of the LIFE is provided by Shapiro and Keller (1979).

The Intervention

The Problem Solving for Life (PSFL) program focused on cognitive restructuring and problem-solving skills training (Beck et al., 1979; D'Zurilla & Nezu, 1980). Teaching methods included didactic sessions; cartoons; individual, small group, and whole class interactive exercises and activities; home-tasks; and diary keeping. The curriculum involved 8 self-contained sessions, each lasting a class period of 45-50 mins that were conducted once per week over 8 weeks. Teachers were provided with pre-prepared curriculum materials designed to teach life problem-solving skills, positive problem-solving orientation and optimistic thinking styles. The content was developed to be “user” friendly, relatively culture-free, age-appropriate, interesting, and relevant. Implementation required minimal preparation by teachers, as supporting materials such as resource books, overheads, background notes, handouts, cartoons, puzzle pieces and posters were provided for each session. Further details regarding content and student and teacher evaluations and feedback about the program can be found in Spence et al. (2003).

Teachers attended training sessions of approximately 6 hours duration. Training covered the theory underlying the program, program content, and program implementation techniques. In order to enhance teacher adherence to program fidelity, they completed a qualitative evaluation form after each session, recording the usefulness, organization and degree of implementation of that program component. Teachers were encouraged to contact the research team at any time regarding problems or questions. However, in general, they required minimal support and reported a high level of compliance with program implementation.

Monitoring Control Condition

Students in the monitoring control (MC) condition completed all assessments according to the same protocol as the intervention group. However, they did not participate in the preventive intervention, and can be regarded as a “school as usual” group. However, for ethical reasons, the school guidance officer/counsellor was informed if a student was identified during the assessment process as meeting criteria for a depressive disorder on the clinical interview or reported suicidal thoughts/intent. The same procedure for notification of clinically depressed students was followed in the intervention and control conditions.

Results

Means and standard deviations for study variables at each occasion are shown in Tables 1 and 2. A MANOVA confirmed that intervention and control groups did not differ on any demographic or outcome measures at pre-intervention. Over the course of the study, a proportion of students were not available for reassessment. Efforts were made at each assessment point to contact all of the original participating students in order to maximise retention rates. Retention rates at follow-up were 71.6% ($n = 499$, 66.4% intervention; $n = 575$, 76.9% control) at 2 years, 60.6% ($n = 389$, 51.8% intervention; $n = 520$, 69.4% control) at 3 years, and 60.6% ($n = 441$, 58.7% intervention; $n = 468$, 62.5% control) at 4 years. The difference in drop-out rates between intervention and control conditions at 4-year follow-up was not statistically significant. However, there were significantly more missing students in the intervention compared to control conditions at 2- and 3-year follow-ups ($\chi^2 = 48.80$, $p < .001$; $\chi^2 = 19.68$, $p < .001$, respectively, $df=1$, $N=1500$). A variable of missingness (number of missing time points) correlated significantly with several pre-test measures including depression, $r = .12$, problem solving scores $r = -.09$; and total number of negative life events $r = .14$. The issue of missing data is a problem in many long-term evaluations and traditional data analysis methods such as repeated measures MANOVA cannot deal well with partially missing data given list-wise deletion of cases. Hierarchical linear modeling (HLM) has the

advantage of allowing retention of cases with partially missing data in longitudinal research and also gives higher weighting to those cases with complete data sets compared to those with missing time points. It also allows comparison of the time trajectories of those who were present for all assessments with those who showed varying levels of missingness and comparison of the effect of differences in missingness across groups. The data were analysed using HLM Version 5 (Bryk & Raudenbush, 1992). As this is a brief report, the analyses are reported in detail only for depression scores, with brief summaries for the remaining outcome measures.

HLM allows mixed, within- and between-subjects models, and enables examination of the impact of between subject factors (such as experimental group) upon changes in an outcome variable over time. The first step was to determine the number of levels of analysis that were most appropriate for the data. If school level factors significantly influenced the outcome variable (such as depression scores) then it would be appropriate to evaluate outcomes with 3 levels (Arnold, 1992; Bryk & Raudenbush, 1992; Lee, 2000). In this context, Level 1 represents a within-subject model in which change in individuals on the outcome measure over time is investigated. Level 2 examines how change on the outcome variable varies as a function of characteristics of the individual (in this case, experimental condition). Level 3 explores how change varies as a function of school factors, taking individual characteristics into account (Raudenbush, Bryk, Cheong, & Congdon, 2000).

However, the 3-level unconditional model revealed that the proportion of total variance in depression scores lying systematically between schools (the intraclass correlation) accounted for less than 1% of variance in depression scores or change in depression scores for both the full and high-symptom samples, with the remainder being explained by the levels 1 and 2 student effects. Thus, further analyses were limited to the first two levels.

Full Sample

The unconditional model for the 2-level analysis for the full sample is shown in Table 3. At level 1 time was coded zero for pre-test and then by month of reassessment. There were six occasions of measurement for depression scores (pre- and post-intervention, 1-, 2- 3- and 4- year follow-ups). The unconditional model showed that there was no significant linear trajectory in mean depression scores over time, $\beta_{10} = -.003$, $p = .612$. However, the random effects co-efficients indicated that there was significant individual variation in depression score trajectories, $\tau_1 = .011$, $\chi^2 = 2232$, $p < .001$, justifying examination of whether intervention significantly predicted this variation.

The 2-level conditional model examined the influence of experimental group (at Level 2) upon individuals' linear growth curves for depression scores at level 1, with group being dummy coded as 0 (control) and 1 (intervention). Only linear time effects are reported here as no significant quadratic time effects were evident. There was no significant effect of group on either the intercept (initial depression scores), $\beta_{01} = -.42$, $p = .29$, or linear time trajectories for depression scores, $\beta_{11} = -.11$, $p = .49$. Thus, there were no significant differences between groups in depression scores prior to intervention and participation in the intervention did not significantly explain individual variations in changes in depression scores over time.

In order to examine the relationship between levels of missing data and outcome, a further analysis was conducted using a dummy coded variable for pattern of missing values as outlined by Hedeker and Gibbons (1997). There was no significant influence of level of missingness upon the time trajectories for depression scores and no significant interaction between missingness and group upon time.

This same set of analyses was then conducted for problem solving, attributional style, social functioning, internalising and externalising scores. There were 6 measurement occasions for problem solving scores and 5 occasions for the remaining variables for which

immediate post-intervention data were not collected. The unconditional models revealed no significant linear time trajectories (β_{10}) for these outcome measures, except for a decrease in internalizing problems over time. However, significant variation in individual growth trajectories (τ_1) was evident for all outcome measures.

The conditional models revealed no significant effects of group upon the intercept (pre-intervention levels) for the full sample of students for problem solving, $\beta_{01} = .139$, $SE = .450$, $t\text{-ratio} = .309$, $p = .757$, attributional style, $\beta_{01} = -.042$, $SE = .169$, $t\text{-ratio} = -.246$, $p = .806$, social functioning, $\beta_{01} = .139$, $SE = .451$, $t\text{-ratio} = .309$, $p = .757$, internalising, $\beta_{01} = .002$, $SE = .458$, $t\text{-ratio} = .005$, $p = .996$, and externalising scores, $\beta_{01} = -.239$, $SE = .415$, $t\text{-ratio} = -.576$, $p = .564$ ($df=1498$ in all cases). This indicates equivalence between groups at baseline for these outcome measures.

When the effect of group upon the linear trajectory was examined, there was a significant effect for problem solving skills, $\beta_{11} = .055$, $SE = .017$, $t\text{-ratio} = 3.246$, $p = .002$. However, this result was explained by the increase in problem solving scores in the intervention group from pre-test to post-intervention. When post-intervention was deleted from the time sequence, the group effect upon the linear trajectory was no longer significant, $\beta_{11} = -.028$, $SE = .017$, $t\text{-ratio} = -1.66$, $p = .10$. Thus, participation in the intervention did not influence the long-term trajectory for problem solving scores from pre-intervention through to 4-year follow-up. There were no significant effects of group upon the linear trajectory for any of the other outcome variables including attributional style, $\beta_{11} = -.006$, $SE = .005$, $t\text{-ratio} = -1.231$, $p = .219$, social functioning, $\beta_{11} = .002$, $SE = .011$, $t\text{-ratio} = .190$, $p = .850$, internalising, $\beta_{11} = .003$, $SE = .013$, $t\text{-ratio} = .272$, $p = .796$, and externalising scores, $\beta_{11} = .011$, $SE = .011$, $t\text{-ratio} = 1.025$, $p = .306$ ($df=1498$ in all cases).

High-Symptom Sample

The unconditional model (see Table 3) revealed a significant linear time effect, $\beta_{10} = -.11$, $p < .001$, suggesting that mean depression scores for the high-risk students tended to decrease over time, consistent with the mean values shown in Table 1. There was also evidence of significant individual variation in both initial depression scores, $\tau_0 = 42.42$, $\chi^2 = 844.84$, $p < .001$, and depression trajectories over time, $\tau_1 = .10$, $\chi^2 = 558.60$, $p < .001$. However, it was clear from the conditional model that group membership (intervention vs control) did not significantly influence either the intercept, $\beta_{01} = .21$, $p = .804$, or linear time slope $\beta_{11} = -.01$, $p < .623$. These results indicate that there were no significant differences between groups for the high-symptom students in terms of depression scores at baseline or in changes in depression scores over time. Subsidiary analysis revealed no significant relationship between level of missingness and time trajectories, nor for the interaction between missingness and group and time.

A similar pattern of results was found for the high-symptom students on the remaining outcome variables. The unconditional models revealed significant time trajectories and significant variance in individual growth patterns for problem solving, attributional style, social functioning, internalizing and externalizing scores.

The conditional models revealed that participation in the intervention (group effect) did not significantly impact upon the intercept for problem solving, $\beta_{01} = -.780$, $SE = 1.18$, $t\text{-ratio} = -0.661$, $p = .508$, attributional style, $\beta_{01} = -.214$, $SE = .346$, $t\text{-ratio} = -0.620$, $p = .508$, social functioning, $\beta_{01} = -.414$, $SE = .867$, $t\text{-ratio} = -0.477$, $p = .633$, internalising, $\beta_{01} = .648$, $SE = .954$, $t\text{-ratio} = .680$, $p = .497$, and externalising scores, $\beta_{01} = .422$, $SE = .947$, $t\text{-ratio} = .446$, $p = .655$ ($df=397$ in all cases). This indicates equivalence between experimental groups at entry to the study for these variables. There were also no significant effects for the influence of intervention (group) upon the time trajectories for problem solving, $\beta_{11} = .005$,

SE = .035, t-ratio = .151, $p = .881$, attributional style, $\beta_{11} = -.004$, SE = .011, t-ratio = -.398, $p = .691$, social functioning, $\beta_{11} = -.019$, SE = .025, t-ratio = -.797, $p = .431$, internalising, $\beta_{11} = -.049$, SE = .028, t-ratio = -1.702, $p = .088$, and externalising scores, $\beta_{11} = -.010$, SE = .024, t-ratio = -.418, $p = .676$ ($df=397$ in all cases).

The absence of significant group differences in changes over time was confirmed when the data were analysed using repeated measures MANOVA. There were no significant group X time or group effects from pre-test to any of the long-term follow-ups for either the full or high-symptoms samples.

Diagnosis of Depression for the High-Symptom Students: Survival Analysis

The purpose of conducting a survival analysis in the current study was to identify the percentage of original high-symptom students (who did not already meet diagnostic criteria for a depressive disorder) in each condition who remained diagnosis free over the 4-year follow-up. The survival analysis method calculates successive interval-specific (conditional) probabilities of onset of a depressive episode based on the number of participants remaining at risk for the outcome event at successive time points. The intervention and control groups were compared at all points simultaneously using the Kaplan-Meier test. Consistent with the conditions of survival analysis, high-symptom participants who were already depressed at pre-assessment were excluded from the analyses. Those participants who dropped out of the study prior to 1-year follow-up were also excluded. For those participants who were lost to follow-up after one year, the conservative approach of assuming non-depressed status over the remaining time period was employed. As a result, a total of 282 high-symptom participants were included in the survival analysis. Of these, 143 participated in the PSFL condition, while 139 participated in the MC condition.

The Kaplan-Meier test revealed no significant difference in depressive disorder incidence rates over the four-year period between the PSFL and MC conditions, *log rank statistic* = .04, $p=.85$. While 74.83% of the high-symptom PSFL group remained diagnosis-

free over the four-year follow-up, 75.54% of the MC high-symptom condition did so. The corollary is that around 25% of high-symptom students in both conditions experienced a clinically diagnosable depressive disorder over the course of the 4-year follow-up.

Discussion

The results revealed no significant differences between students who completed the preventive intervention compared to those who did not in terms of time trajectories for any of the evaluation measures from pre-intervention to 4-year follow-up. This pattern of results was found for both the sample as a whole and those students who had initially been identified as showing elevated symptoms of depression and thus at risk for depression in the future. In the initial study, both high-symptom and low-symptom students had shown positive short-term benefits after participating in the preventive intervention in comparison to the control condition. However, these benefits were no longer evident by 1-year follow-up (Spence et al., 2003). The present study indicates that the lack of difference between conditions remained throughout the 4-year follow-up period. It was clear that no delayed or “sleeper” effects were evident.

Even the high-symptom students who completed the intervention did not respond differently to their control counterparts over the 4 years of follow-up on any of the evaluation measures. Over time, the high-symptom sample showed significant decreases in depressive, internalising and externalising problems, and increases in social problem-solving and social functioning, irrespective of experimental condition. However, their level of functioning at 4-year follow-up was still clearly below those of the sample as a whole. This finding suggests that the high-symptom group were indeed “at risk” and also demonstrates the stability of such problems over prolonged periods. Of those students who were identified as showing elevated symptoms of depression (but not criteria for a depressive disorder) at baseline, around 25% reported experiencing a diagnosable depressive disorder over the following 4-year period, irrespective of whether they participated in the preventive intervention.

The lack of long-term benefits from the intervention emphasizes the importance of long-term follow-ups in evaluations of programs designed to prevent the development of depression. To date, such extended follow-ups have rarely been conducted. It is important to explore possible reasons why the short-term benefits were not maintained. Some possible explanations that we considered were that (i) the teachers may not have had the background skills and knowledge to apply a cognitive-behavioral intervention effectively, (ii) the intervention may have been of insufficient duration to produce lasting acquisition in cognitive coping and problem solving skills, and (iii) the large group, classroom format of delivery may not provide sufficient individualized learning opportunities for long-term skill acquisition. In relation to teacher skills, although the present study collected qualitative data that indicated teacher adherence to the program, we were unable to observe teachers in the classroom. Thus we do not have an objective measurement of the degree to which teachers were able to deliver the program effectively. However, interestingly, Shatté (1997) found no difference in outcome between teachers and graduate clinical psychology interns in administration of a cognitive-behaviour preventive program, suggesting that the level of cognitive-behavioral professional training may not be an adequate explanation of the lack of long-term benefits.

In relation to length of intervention, the program was of sufficient strength to produce initial improvements in the targeted problem solving skills (Spence et al., 2003). Although it is intuitively appealing to suggest that longer duration of program will be more effective, it should be noted that the Shatté (1997) program involved 12 sessions of 24 hours total duration and was still insufficient to produce lasting effects. Similarly, it appears that a small group-format may not overcome the possible limitations of a whole class approach given that the Shatté (1997) study was conducted in small groups of around 10 children.

Given the findings of the present study, it must be questioned whether universal interventions are the most sensible way to proceed in terms of prevention of depression in

young people. It is tempting to suggest that indicated and selective approaches targeted to individuals who manifest specific risk factors or early symptoms for a disorder may produce superior outcomes in the prevention of depression in youth. Certainly this proposition warrants investigation.

In order to interpret the findings of the present study, it is important to note its strengths and weaknesses. Its strengths included the large sample size, random allocation of matched schools to conditions, psychometric rigor of most outcome measures, and extended duration of follow-up. The sample size, even accounting for drop out at 4-year follow-up ($N = 910$) provided a power of .85 to detect a small effect size (f) of .10 at $\alpha = .05$ for the full sample. For the high-symptom students the final sample size ($N = 224$) provided a power of .85 to detect a small to medium effect size (f) of .20 at $\alpha = .05$. The primary limitation relates to the representativeness of the sample given the potential bias associated with the initial consent and subsequent dropout rates. While the dropout rate was consistent with past research in this area and every effort was made to retain participants, only 60.6% of the original sample remained at 4-year follow-up. Furthermore, there were small but significant correlations between the level of missing data and initial depression scores, problem solving skills and negative life events, suggesting that those who dropped out of the study may have been at higher risk for future depression than those who remained in the study. Although HLM enabled greater retention of individuals within the computations than traditional repeated measures MANOVA techniques, HLM does not overcome all the problems associated with non-random missingness. Theoretically, it is possible that those students who dropped out of the study may have been those who benefited most from the intervention in the long-term. Although this proposition is theoretically possible, in reality it is unlikely. Examination of changes in depression scores from pre- to post-intervention for the PSFL condition revealed no significant differences in outcome associated with different degrees of “missingness” for either the full sample or high-symptom students. Also, the HLM analyses

revealed no significant association between missingness and depression trajectories. However, before firm conclusions can be drawn about the long-term effectiveness of this type of cognitive-behavioral approach to the prevention of depression, evaluations are required that ensure participation and retention of those most at risk.

A further limitation for the study was the lack of interview data for the full sample. As explained earlier, the resources of the project did not permit interview of all 1500 students, with interviews being restricted to those who were initially considered to be at elevated risk for depressive disorders in the future. For evaluation of the full sample, the study relied on student self-report, and future research should aim to include independent informant data. The nature of the monitoring control condition may also represent an experimental weakness. It is feasible that participation in the assessment procedures and the referral of those who already met a diagnosis may have represented a form of intervention in its own right. In practice this possibility is unlikely as the results did not change if the clinically depressed (3% of the total) students were deleted from the analyses.

Despite these methodological limitations, the results of the present study bring into question the long-term benefits of brief, school-based, universal approaches to preventing the development of depression in young people. From the point of view of policy makers, there is little point in investing huge amounts of money in preventive programs if they are not effective. Future research needs to determine what program content, mode of delivery, background and training of practitioners, amount, timing and duration of delivery, and location is most effective for which group of individuals in preventing the development of depression in young people.

Table 1
Means and Standard Deviations on Outcome Measures for Pre-intervention, Post-intervention, 1-yr, 2-yr, 3-yr and 4-yr Follow-up

Risk status and experimental condition	BDI						Total problem solving					
	Pre-	Post-	1-yr	2-yr	3-yr	4-yr	Pre-	Post-	1-yr	2-yr	3-yr	4-yr
High-Symptom Group												
Intervention												
<i>M</i>	19.84	12.07	12.91	13.28	10.41	11.27	51.12	53.69	53.92	53.67	56.60	55.06
<i>SD</i>	8.38	10.80	11.92	11.97	9.10	9.12	13.75	14.93	14.08	14.28	12.88	14.20
<i>n</i>	204	164	133	115	89	102	204	164	130	116	89	102
Control												
<i>M</i>	18.01	14.85	12.73	12.33	11.30	11.51	52.91	53.68	55.10	54.82	56.41	56.44
<i>SD</i>	8.94	11.00	8.36	10.28	8.52	9.19	12.09	13.50	11.82	14.21	13.17	13.80
<i>n</i>	195	168	131	144	132	121	195	165	128	143	132	121
Whole Sample												
Intervention												
<i>M</i>	8.71	6.24	7.80	7.13	7.60	7.23	60.70	62.76	59.22	60.15	60.36	61.10
<i>SD</i>	8.68	7.85	9.80	8.47	9.20	7.71	13.59	15.00	14.18	13.70	13.30	13.57
<i>n</i>	751	627	517	499	389	441	751	624	509	500	390	442
Control												
<i>M</i>	8.02	8.65	7.57	7.43	7.87	7.61	60.34	59.70	59.43	59.90	61.34	61.70
<i>SD</i>	8.10	10.04	8.32	8.65	8.16	8.01	12.62	13.91	13.85	14.05	14.52	14.22
<i>n</i>	749	639	550	575	520	468	749	626	537	574	520	467

Note: Pre = pre-intervention; Post- = post-intervention

Table 2
Means and Standard Deviations on Outcome for Attributional Style, Social Functioning, Internalizing and Externalizing Problems at Each Occasion

Risk status and experimental condition	Attributional style					Social Functioning					YSR Internalising					YSR Externalising				
	Pre	1-yr	2-yr	3-yr	4-yr	Pre	1-yr	2-yr	3-yr	4-yr	Pre	1-yr	2-yr	3-yr	4-yr	Pre	1-yr	2-yr	3-yr	4-yr
High-Symptom Group																				
Intervention																				
<i>M</i>	1.68	1.50	1.40	2.07	2.25	66.04	66.93	55.52	65.91	69.33	21.94	17.18	16.56	15.35	14.97	18.06	16.61	15.14	16.65	14.39
<i>SD</i>	3.68	4.03	4.18	3.99	4.50	8.70	10.75	10.23	10.77	9.65	10.21	10.59	10.91	10.10	10.23	9.44	10.40	9.50	9.41	8.71
<i>n</i>	204	126	112	89	102	204	130	114	89	101	204	129	115	89	102	204	129	115	89	102
Control																				
<i>M</i>	1.84	2.18	1.98	2.70	2.62	66.62	67.09	68.26	67.41	70.32	20.91	19.39	17.34	16.18	16.02	17.48	16.14	14.97	14.35	14.64
<i>SD</i>	3.47	3.32	4.29	4.19	4.12	8.54	9.40	9.24	8.80	10.07	9.64	9.90	10.74	9.87	10.44	10.05	9.80	9.18	8.67	9.52
<i>n</i>	195	127	143	132	121	195	130	140	131	121	195	128	140	132	119	195	128	140	132	119
Whole Sample																				
Intervention																				
<i>M</i>	3.80	3.19	3.67	3.73	4.00	71.08	70.43	70.36	70.07	70.83	13.13	12.40	11.67	12.12	12.22	12.41	13.15	12.40	13.66	12.35
<i>SD</i>	3.50	3.93	3.97	3.99	4.13	8.83	10.58	9.63	10.09	9.00	9.42	10.15	9.73	10.66	9.30	7.97	9.82	8.61	10.07	8.66
<i>n</i>	751	499	496	388	441	751	504	496	388	441	751	500	500	389	440	751	500	500	389	440
Control																				
<i>M</i>	3.81	3.44	3.76	4.14	4.18	70.90	70.00	70.73	70.17	72.36	13.18	12.11	12.06	11.63	11.32	13.39	13.19	13.19	12.46	11.75
<i>SD</i>	3.38	3.57	3.90	3.84	3.83	8.57	9.52	9.53	8.80	9.18	8.84	9.33	10.04	9.25	9.09	8.15	9.22	9.45	8.62	8.50
<i>n</i>	749	522	574	540	467	749	544	558	519	467	749	530	563	520	464	749	530	563	520	464

Table 3

Estimated Effects for the Unconditional and Conditional Models for Depression Scores in the Full and High-Symptom Samples

	<i>Coefficient</i>	<i>SE</i>	<i>t ratio</i>	<i>Variance component</i>	χ^2	<i>Approx df</i>	<i>p value</i>
FULL SAMPLE							
Unconditional Model							
Fixed effect							
Intercept							
Mean Initial Status β_{00}	7.970	0.198	40.146			1499	.000
Linear time trend							
Mean growth rate β_{10}	-.003	.005	-.506			1499	.612
Random Effects							
Initial Status τ_0				40.775	4354.261	1429	.000
Growth rate τ_1				.001	2232.915	1429	.000
Level 1 error				38.405			
Conditional Model							
Fixed effect							
Intercept							
Mean Initial Status β_{00}	8.179	.281	29.081			1498	.000
Group β_{01}	-.419	.396	-1.057			1498	.291
Linear time trend							
Intercept β_{10}	-.006	.007	-.860			1498	.390
Group β_{11}	.108	.011	.696			1498	.486
Random Effects							
Initial Status τ_0				40.758	4349.303	1428	.000
Growth rate τ_1				.011	2231.840	1428	.000
Level 1 error				38.405			
HIGH-SYMPATOM SAMPLE							
Unconditional Model							
Fixed effect							
Intercept							
Mean Initial Status β_{00}	16.225	.423	38.345			398	.000
Linear time trend							
Mean growth rate β_{10}	-.114	.023	-9.093			398	.000
Random Effects							
Initial Status τ_0				42.420	845.018	376	.000
Growth rate τ_1				.103	558.605	376	.000
Level 1 error				59.976			
Conditional Model							
Fixed effect							
Intercept							
Mean Initial Status β_{00}	16.122	.604	26.707			397	.000
Group β_{01}	.209	.848	.247			397	.804
Linear time trend							
Intercept β_{10}	-.108	.017	-6.321			397	.000
Group β_{11}	-.012	.025	-.487			397	.626
Random Effects							
Initial Status τ_0				42.603	844.843	375	.000
Growth rate τ_1				.011	558.278	375	.000
Level 1 error				59.964			

References

- Achenbach, T. M. (1991). *Manual for the Youth Self-Report and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Arnold, C. L. (1992). An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development*, 25(2), 58-90.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Wiley.
- Bennett, D. S., Ambrosini, P. J., Bianchi, M., Barnett, D., Metz, C., & Rabinovich, H. (1997). Relationship of Beck Depression Inventory factors to depression among adolescents. *Journal of Affective Disorders*, 45, 127-134.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, Ca.: Sage Publications.
- Cardemil, E. V. (2002). Preventing paralysis in culture-based research: Negotiating obstacles. *Prevention and Treatment*, 5.
- Carter, C. L., & Dacey, C. M. (1996). Validity of the Beck Depression Inventory, MMPI, and Rorschach in assessing adolescent depression. *Journal of Adolescence*, 19, 223-231.
- 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. B., Lewinsohn, P. M., & Seeley, J. M. (1995). Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: A randomized trial of group cognitive intervention. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34(3), 312-321.
- D'Zurilla, T. J., & Nezu, A. (1980). A study of the generation-of-alternatives process in social problem solving. *Cognitive Therapy and Research*, 4(1), 67-72.
- D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (1997). *Manual for the Social Problem-Solving Inventory-Revised (SPSI-R)*. North Tonawanda, N. Y.: Multi-Health Systems, Inc.
- Gillham, J. E., Reivich, K. J., Jaycox, L. H., & Seligman, M. E. P. (1995). Prevention of depressive symptoms in schoolchildren: Two-year follow-up. *Psychological Science*, 6(6), 343-351.
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2(1), 64-78.

- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, 35(2), 125-141.
- Merry, S., McDowell, H., Hetrick, S., Bir, J., & Muller, N. (2003). *Psychological and/or educational interventions for the prevention of depression in children and adolescents*, 2003
- 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.
- Price, C., Spence, S. H., Sheffield, J., & Donovan, C. L. (2002). The development and psychometric properties of a measure of social and adaptive functioning for children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 31(1), 111-122.
- 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.
- Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International, Inc.
- Reynolds, W. M. (1994). Assessment of depression in children and adolescents by self-report questionnaire. In W. M. Reynolds & H. F. Johnston (Eds.), *Handbook of depression in children and adolescents* (pp. 209-234). New York: Plenum Press.
- Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30(1), 58-66.
- Seligman, M. E. P., Kaslow, N. J., Allow, L. B., Peterson, C., Tannenbaum, R. L., & Abramson, L. Y. (1984). Attributional style and depressive symptoms in children. *Journal of Abnormal Psychology*, 93, 235-238.
- Shapiro, R., & Keller, M. (1979). *Longitudinal Interval Follow-up Evaluation (LIFE)*. Boston: Massachusetts General Hospital.
- Shatté, A. (1997). *Prevention of depressive symptoms in adolescents: Issues of dissemination and mechanisms of change*. University of Pennsylvania, USA.
- Shochet, I. M., Dadds, M. R., Holland, D., Whitefield, K., Harnett, P. H., & Osgarby, S. (2001). The efficacy of a universal school-based program to prevent adolescent depression. *Journal of Clinical Child Psychology*, 30(3), 303-315.

- Silverman, W. K., & Albano, A. M. (1996). *Anxiety Disorders Interview schedule for DSM-IV - Child Version: Parent Interview Schedule*. San Antonio: The Psychological Corporation - Harcourt, Brace & Company.
- Silverman, W. K., & Eisen, A. R. (1992). Age differences in the reliability of parent and child reports of child anxious symptomatology using a structured interview. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31(1), 117-124.
- Silverman, W. K., & Nelles, W. B. (1988). The Anxiety Disorders Interview Schedule for Children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27(6), 772-778.
- Silverman, W. K., Saavedra, L. M., & Pina, A. A. (2001). Test-retest reliability of anxiety symptoms and diagnoses with anxiety disorders interview schedule for DSM-IV : Child and parent versions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(8), 937-944.
- Spence, S. H., Sheffield, J., & Donovan, C. L. (2003). Preventing adolescent depression: An evaluation of the Problem Solving for Life program. *Journal of Consulting and Clinical Psychology*, 71(1), 3-13.
- Thompson, M., Kaslow, N. J., Weiss, B., & Nolen-Hoeksema, S. (1998). Children's Attributional Style Questionnaire--Revised: Psychometric examination. *Psychological Assessment*, 10(2), 166-170.