

The Columbia SuicideScreen: Validity and Reliability of a Screen for Youth Suicide and Depression

DAVID SHAFFER, F.R.C.PSYCH. (LOND), MICHELLE SCOTT, PH.D., HOLLY WILCOX, M.A., CAREY MASLOW, PH.D., ROGER HICKS, B.A., CHRISTOPHER P. LUCAS, M.D., ROBIN GARFINKEL, PH.D., AND STEVEN GREENWALD, M.A.

ABSTRACT

Objective: This study reports on the psychometric properties of a brief, self-administered screening questionnaire, the Columbia SuicideScreen® (CSS), intended to identify high school students at risk for suicide. **Method:** Seventeen hundred twenty-nine 9th- to 12th-grade students completed the CSS and Beck Depression Inventory during school hours in 1991 to 1994. Three hundred fifty-six students who screened positively and 285, group matched on age, gender, and ethnicity, who screened negatively were examined on the Diagnostic Interview Schedule for Children (DISC), version 2.3, to assess validity. The DISC-based suicide risk criterion was suicidal ideation or prior suicide attempt and a *DSM-III-R* diagnosis of major depression or dysthymia or substance use. Test-retest reliability was assessed in a subsample of 85. **Results:** The most balanced algorithm had a sensitivity of 0.75, specificity 0.83, and positive predictive value 16%. Suicidal ideation and prior attempt item reliabilities (κ) were 0.48 and 0.58, respectively. Eight-day test-retest reliability for the most balanced scoring algorithm was 0.32. **Conclusions:** The CSS demonstrated good sensitivity and reasonable specificity identifying students at risk for suicide. A second-stage evaluation would be needed to reduce the burden of low specificity. *J. Am. Acad. Child Adolesc. Psychiatry*, 2004;43(1):71-79. **Key Words:** suicide, depression, adolescents, identification, psychometrics.

Adolescent suicide almost always occurs in the context of an active, often treatable, mental illness (Brent et al., 1999; Groholt et al., 1998; Shaffer et al., 1996b) that has frequently gone unrecognized or untreated. A mood disorder is the most common type of associated psychiatric condition, and there is good evidence for its treatability by either psychopharmacology (Waslick et

al., 2002) or psychotherapy (Mufson and Velting, 2002). There has been a substantial increase in the proportion of depressed adolescents who are now treated with antidepressants (Carlsten et al., 2001; Isaacson, 2000; Olfson et al., 2002, 2003), and it is plausible that the decline in the incidence of suicide among teens over the past 10 years in the United States (NCHS, 2001) is, in part, attributable to this. If the risk factors for suicide are both identifiable and treatable, screening teens for untreated mood disorders should be an important component of any suicide prevention program.

Several instruments have been developed to assess suicide risk (Goldston, 2003). Some, such as the Beck Scale for Suicidal Ideation (Steer et al., 1993), the Suicide Behaviors Questionnaire (Cotton and Range, 1993), the Suicidal Ideation Questionnaire (Reynolds, 1988), and the Depressive Symptom Inventory-Suicidality subscale (Joiner et al., 2002), focus on suicidal ideation or behavior. Others, such as the Beck Hopelessness Scale (Steer et al., 1993) and Reasons for Living Scale (Osman et al., 1996), address constructs such as hopelessness or the relative value of life or death

Accepted June 3, 2003.

Drs. Shaffer, Scott, Lucas, and Garfinkel and Mr. Hicks and Mr. Greenwald are with the Department of Child Psychiatry, New York State Psychiatric Institute, Columbia University, New York. Ms. Wilcox is with the Department of Mental Hygiene, Bloomberg School of Public Health, Johns Hopkins University, Baltimore. Dr. Maslow is with National Development and Research Institutes, Inc., New York.

This work was supported by Centers for Disease Control grant R49/CCR 202598, NIMH grants P30 MH 43878 and ST32MH-16434, and grants from the American Mental Health Foundation and the Joy and William Ruane Center at Columbia University.

Correspondence to Dr. Shaffer, Department of Child Psychiatry, New York State Psychiatric Institute, Columbia University, 1051 Riverside Drive, Unit 78, New York, NY 10032; e-mail: shafferd@childpsych.columbia.edu.

0890-8567/04/4301-0071©2003 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/01.chi.0000096370.43887.68

that do not figure in the *DSM* or *ICD* diagnostic criteria. Only the Suicide Risk Screen (SRS) (Eggert et al., 1994) covers the range of factors—suicide ideation, attempts, depression, and substance use—that have been found to predict suicide in psychological autopsy studies (Brent et al., 1999; Shaffer et al., 1996b). Most of the instruments have been validated against past history of suicidal behavior, and none to our knowledge has been tested against an algorithm of high risk based on a reliable diagnostic interview.

The purpose of this study was to develop a brief, self-report instrument that would assess the most important known risk factors for suicide (Shaffer et al., 1996b) and to validate it against a diagnostic profile of suicide risk obtained through a well-validated structured interview. Additional goals were to measure test-retest reliability and to compare the instrument (the Columbia SuicideScreen® [CSS]) with the Beck Depression Inventory (BDI), a widely used depression screening instrument.

METHOD

Study Design

A convenience sample of 2,583, 9th- to 12th-grade students from seven diverse high schools was identified from school attendance registers in the greater New York metropolitan area. Those who agreed to participate and were present on the screening day ($n = 1,729$) completed the CSS questionnaire and the BDI, embedded in a larger health survey, at the start of a regular class period during 1991 to 1994. The project and its procedures were presented to parents, school faculty, and administration at appropriate meetings. The project was described as inquiring into the physical and emotional health of teenagers. Students took home a letter to all parents describing the project and offering an opportunity to not participate. Just prior to screening, teachers read a description of the project, described as a "health survey," and distributed the assent form to the students in their classrooms. Students who were interested in participating and whose parents had not denied consent signed the assent form and completed the survey.

Students who endorsed key risk items on the CSS (suicide ideation or attempt, depression, and excess substance use) were then interviewed on the more comprehensive and detailed Diagnostic Interview Schedule for Children version 2.3 (DISC). Halfway through the study, the demographic profile of screen-positive students was described and a group-matched sample of youths who did not endorse any of the risk items was selected for further study as controls and completed the DISC.

Clinicians (second-year child psychiatry residents or licensed clinical psychologists) interviewed the students to verify and elaborate their CSS and DISC responses. Referral for further evaluation or treatment was determined by DISC diagnoses and/or significant social or educational impairment or distress. Names of youths deemed in need of a referral were given to the study case manager, who would interview the parents by telephone or home visit to secure the consent to a referral. The case manager would then set up the appointment. They followed up with the families at 3 and 6 months after referral.

The New York State Psychiatric Institute's institutional review board, the review board of the New York City Board of Education, and the review board of the Archdiocese of New York approved the screening materials and consent procedures.

Sample

Eight schools, with a population of 2,858 students in grades 9 through 12, were chosen to represent the variety of schools found in the New York metropolitan area. These included two parochial single-sex schools (one male and one female), two suburban and six urban schools, and one vocational-technical school. One of the suburban high schools withdrew its participation after a parent representative protested the school's involvement in a noneducational program. Data from these students were not included in the analysis ($n = 275$; 9.6%). Of the remaining 2,583 youths, 340 (13.2%) were absent during the study period, 182 (7%) parents refused to allow their child to participate in the project, and an additional 332 (12.9%) youths declined to participate, leaving a sample of 1,729 (67%) who completed screening forms. The final sample was 57% female, with a mean age of 15.4 ($SD = 1.4$); age range was 11 to 19 years. Of the final sample, 56% was white, 18% African American, 13% Hispanic, and 13% other ethnicity. Each grade represented approximately 25% of the sample.

Students were identified as screen positive if they endorsed any of the following on the CSS: (1) suicidal ideation in the past 3 months, *or* (2) a prior suicide attempt at any time, *or* (3) three of the five emotional items (unhappy, withdrawal, irritability, anxiety, or substance abuse) rated as a bad or very bad problem, *or* (4) indication of a youth's desire to speak to a professional about any of five emotional symptom items. There were 489 (28%; the following percentages use 1,729 as the denominator unless otherwise indicated) students who met one or more of these criteria: 298 (17%) had either attempted suicide in the past ($n = 108$; 6%) or had thought about suicide in the past 3 months ($n = 190$; 11%), 175 (59%) of the suicidal group ($n = 298$) had either rated three of five emotional symptoms as serious (bad or very bad) or requested help on any one of five emotional symptoms. Of the 298 current ideators and prior attempters, 43 (14%) reported prior attempts without three of five associated serious emotional symptoms or any request for help, and 80 (27%) were current ideators without prior attempts, three of five associated serious emotional symptoms, or request for help. Of the 191 nonsuicidal students selected for examination, 127 (66%) signaled a request for help but had no severe emotional symptoms, and 64 (33%) had serious anxiety or depressive symptoms but no suicidality. Of the above, 356 (73%) were interviewed on the DISC between 1 and 14 days after the CSS administration. Students not interviewed included absentees and a small number who refused to be interviewed. Using sampling with replacement, 285 students (23% of all screen negatives) who did not meet the screen-positive criteria and who were available for interview on the designated days were group matched by grade, gender, and ethnicity to the screen-positive sample and were selected, blind to the clinicians or DISC interviewers, to serve as controls.

CSS-positive students who completed the DISC interview compared to CSS-positive students who did not complete the DISC interview were more likely to be female ($\chi^2_1 = 18.25, p < .001$). The CSS-negative controls who completed the DISC were more likely to be female ($\chi^2_1 = 6.54, p = .011$), Hispanic ($\chi^2_1 = 90.97, p < .001$), and older (mean age of 15.6 versus 15.3, $p < .001$) compared to CSS-negative controls who did not complete the DISC interview.

Eighty-five students, from one urban and one suburban school, were selected for test-retest evaluation of the CSS and BDI after an

8-day interval. The reliability sample did not differ significantly from the total screened sample by gender, grade, or mean age, but differed significantly on ethnicity, having significantly more Hispanic (41% versus 11%; $\chi^2_1 = 68.0, p < .001$) and African American youths (34% versus 18%; $\chi^2_1 = 13.8; p < .001$) and significantly fewer white youths (8% versus 61%; $\chi^2_1 = 116.0, p < .001$).

Instruments

Columbia SuicideScreen. The CSS is an 11-item, self-report questionnaire that includes items for lifetime suicide attempts and the 3-month prevalence of suicide ideation, negative mood (“unhappy or sad,” “withdrawn,” “irritable,” and “nervous or worried”), and substance abuse (“drugs or alcohol”). To avoid a focus on suicidality, these items are embedded within 32 general health questions and 4 items on relationships and family concerns. Responses are entered on a 5-point, visual analog scale (Aitken, 1969), with anchor points labeled from 1 (“no problem”) to 5 (“very bad problem”). The items assessing suicide ideation and attempt invite only a “yes” or “no” response. Students who endorse a symptom as “bad” or “very bad” (4 or 5) or “yes” to current ideation or prior attempt are then asked contingent questions about the self-perceived need for help and current mental health treatment, if any. The question about self-perceived need for help was asked to identify students who were likely to be more highly motivated. The question about the actual receipt of help was included to avoid postscreening interference with ongoing treatment.

Beck Depression Inventory. The BDI (Beck and Steer, 1993) is a 21-item, self-report inventory that assesses cognitive, behavioral, affective, and somatic aspects of depression. In nonclinical adolescent populations, BDI scores of 16 and above indicate possible dysphoria, scores greater than 21 presumptive clinical depression (Kendall et al., 1987).

Diagnostic Interview Schedule for Children. The DISC, version 2.3 (Shaffer et al., 1996a), is a respondent-based instrument designed to be used by lay interviewers to establish a probable *DSM-III-R* psychiatric diagnosis in children and youth aged 6 to 17. Respondents are asked an initial broad question about a symptom that, if endorsed, is followed by a series of items rating its severity. Computer algorithms calculate diagnoses as well as scores on symptom and criterion scales. In this study, a youth informant was administered only part of the entire interview (i.e., major depression/dysthymia, alcohol and substance use, and anxiety). The anxiety module included sections for individual anxiety disorders.

The test-retest reliability of the DISC 2.3 for lay- versus clinician-administered DISC was 0.37 (κ) for major depression, 0.43 (κ) for dysthymia, and 0.39 (κ) for any anxiety disorder in a community sample (Schwab-Stone et al., 1996). Validity of the DISC 2.3, when comparing a clinician rating of ambiguous responses to the original DISC rating, was 0.79 (κ) for major depression, 0.54 (κ) for dysthymia, and 0.49 (κ) for any anxiety disorder (Schwab-Stone et al., 1996).

Statistical Analyses

Statistical analyses proceeded in several stages. First, sensitivity, specificity, and positive and negative predictive values were calculated against a validity criterion consisting of suicidal ideation or previous suicide attempt and a DISC *DSM-III-R* diagnosis with impairment of major depressive disorder or dysthymia or substance abuse/dependence. To determine the optimal CSS threshold, algorithms using various combinations of CSS questions, determined by clinical judgment and prior research, with varying levels of severity were constructed and assessed. Receiver operating characteristic curve was not used to determine the best-balanced algorithm

because the CSS algorithms are not a simple sum of individual items but a combination of items using “and” or “or” rules based on clinical judgment. The algorithm that showed the highest sensitivity and specificity using the DISC criterion was designated as “best balanced.” Test-retest reliability was computed for individual items and the two best-balanced CSS scoring algorithms. Finally, a comparison was made between the sensitivity and specificity of the CSS scoring algorithms and the BDI using cutoff scores of 16 and 21 (Kendall et al., 1987). Insufficient sample size precluded an analysis to evaluate reliability by ethnicity.

Because predictive values are affected by the prevalence of disease, the positive and negative predictive values were calculated with four different disease-prevalence values (i.e., the weighted prevalence of diagnoses in our total sample and three possible prevalences of diagnosis [1%, 5%, and 10% prevalence]). The test-retest reliability of the scoring algorithm on the CSS, the screen’s individual items, and the screen’s agreement with the BDI and DISC were estimated by the κ statistic and intraclass correlation coefficients when appropriate.

RESULTS

The demographic and symptom characteristics of the sample are shown in Table 1. The item prevalence of CSS questions in various population groups is shown in Table 2. Sensitivity, specificity, and positive and negative predictive values of the CSS and BDI in relationship to the DISC are shown in Tables 3 and 4, respectively.

Risk Item Prevalence

Frequencies of endorsement of the four problem areas (depression, anxiety, substance abuse, and suicidal behavior [attempt and ideation]) used in the CSS scoring algorithms are presented in Table 2. Nominal comparisons are presented. All comparisons use a p value of $<.05$ and are not corrected for multiple comparisons. Girls endorsed the anxiety, unhappy, irritable, and suicide items more frequently than boys. Boys more frequently endorsed substance use ($\chi^2_1 = 4.07, p = .044$). There were significant ethnic differences in endorsements of irritability ($\chi^2_4 = 10.41, p = .034$), suicidal ideation ($\chi^2_4 = 13.70, p = .008$), and suicide attempts ($\chi^2_4 = 13.44, p = .009$). There were significant grade differences in endorsements of unhappy ($\chi^2_3 = 17.23, p < .001$), substance use ($\chi^2_3 = 11.85, p = .008$), anxiety ($\chi^2_3 = 8.01, p = .046$), and irritability ($\chi^2_3 = 9.25, p = .026$).

Validity

The CSS algorithm that gave the best sensitivity (0.88) but the worst specificity (0.72) was any suicidal ideation or previous attempt (algorithm I) (Table 3). Eight subjects denied suicidal ideation or prior attempt on the CSS but admitted to suicidal behavior on the DISC. These eight cases were eligible for the DISC

TABLE 1
Characteristics of SuicideScreen Samples

	Screen Neg (<i>n</i> = 1240)	Screen Pos (<i>n</i> = 489)	Any Ideation/Attempt (<i>n</i> = 298)	Any Ideation/Attempt ≥ 3 Any Depressive Item and ≥ 3 on Anxiety (<i>n</i> = 207)
Sex				
Male	587 (78)	165 (22)	89 (12)	47 (6)
Female	653 (67)	324 (33)	209 (21)	160 (16)
Total	1240 (72)	489 (28)	298 (17)	207 (12)
Grade				
9	341 (77)	103 (23)	72 (16)	46 (10)
10	291 (71)	120 (29)	68 (17)	44 (11)
11	290 (68)	136 (32)	89 (21)	73 (17)
12	285 (71)	117 (29)	62 (15)	40 (10)
Missing	33 (72)	13 (28)	7 (15)	4 (9)
Total	1240 (72)	489 (28)	298 (17)	207 (12)
Ethnicity				
White	724 (75)	248 (26)	161 (17)	116 (12)
Black	217 (69)	99 (31)	43 (14)	29 (9)
Hispanic	141 (69)	76 (35)	51 (24)	32 (15)
Asian	72 (67)	36 (33)	25 (23)	22 (20)
Other	40 (65)	22 (36)	15 (24)	7 (11)
Missing	46 (85)	8 (15)	3 (6)	1 (2)
Total	1240 (72)	489 (28)	298 (17)	207 (12)
Age				
Mean (SD)	15.4 (1.4)	15.5 (1.3)	15.4 (1.3)	15.5 (1.3)

Note: Values represent *n* (%).

evaluation because they either reported three or more emotional problems at a high severity level or requested help. The CSS algorithm that gave both the highest specificity and the best balance between sensitivity and specificity (0.75 and 0.83, respectively) was algorithm VI (suicidal ideation or previous attempt and a score of ≥ 3 for unhappy or withdrawal or irritability and anxiety). Using the prevalence of disease to be 4% (see above), algorithm I and algorithm VI had positive predictive values of 12% and 16%, respectively, while both had negative predictive values of 99%. The other four a priori algorithms had higher specificities, ranging from 0.85 to 0.96, but very poor sensitivities, ranging from 0.04 to 0.09, and—using the prevalence of DISC positives of 4%—positive predictive values under 7% and negative predictive values of 96%.

Test–Retest Reliability

Test–retest reliabilities for “suicidal ideation” and “suicide attempt” were 0.48 and 0.58 (κ), respectively. Test–retest reliability for the two best-balanced CSS algorithms were 0.55 (algorithm I) and 0.32 (algorithm VI). κ was calculated for the BDI at two cutoff points, a score of 16 or more ($\kappa = 0.48$) and a score of 21 or

more ($\kappa = 0.52$). The continuous BDI scores yielded an intraclass correlation coefficient of 0.79.

Comparison of Columbia SuicideScreen and Beck Depression Inventory

Using the same validity criterion and a cutpoint of 16, the BDI had a sensitivity of 0.68 and a specificity of 0.78. When the Beck threshold was raised to 21, sensitivity dropped to 0.45 and specificity increased to 0.88 (Table 4). The sensitivities for the BDI, at either cutpoint, were thus lower than those of CSS scoring algorithms I and VI.

DISCUSSION

This study of 1,729 high school students tested the reliability and validity of a brief screening questionnaire, the CSS, designed to allow adolescents to self-report the presence of the major risk factors for completed suicide. Different combinations and threshold values for individual CSS items were examined. The algorithm that afforded the best balance against the validity criterion (i.e., mood or substance use disorder on the DISC, coupled with suicidal ideation or prior attempt) was suicidal ideation or previous at-

TABLE 2
Prevalence of Risk Items by Demographic Characteristics

Item	Percent Positive	
	Boys (<i>n</i>)	Girls (<i>n</i>)
	(<i>n</i> = 752)	(<i>n</i> = 977)
Anxiety**	12.5 (94)	24.9 (243)
Unhappy**	10 (75)	21.8 (213)
Irritable**	23.5 (177)	36.1 (353)
Withdrawn	6.3 (47)	6.8 (66)
Substance Use	2.5 (19)	1.2 (12)
Attempted Suicide**	3.5 (25)	8.7 (83)
Suicidal Ideation**	10.7 (77)	18.8 (179)

	Percent Positive				
	White	Black	Hispanic	Asian	Other
Anxiety	21.2 (206)	15.5 (49)	18.0 (39)	25.0 (27)	14.5 (9)
Unhappy	17.9 (165)	14.9 (47)	20.7 (45)	17.6 (19)	17.7 (11)
Irritable*	29.6 (288)	36.7 (116)	34.6 (75)	23.1 (25)	27.4 (17)
Withdrawn	6.5 (63)	6.3 (20)	8.3 (18)	7.4 (8)	4.8 (3)
Substance Use*	2.4 (23)	0.6 (2)	1.4 (3)	2.8 (3)	0 (0)
Attempted Suicide**	6.0 (57)	4.0 (12)	11.5 (24)	9.3 (10)	8.8 (5)
Suicidal Ideation**	14.4 (137)	12.2 (36)	20.8 (43)	21.7 (23)	23.3 (14)

	Percent Positive			
	Grade 9	Grade 10	Grade 11	Grade 12
Anxiety*	16.7 (74)	19.0 (78)	24.2 (103)	19.9 (80)
Unhappy**	11.5 (51)	19.5 (80)	21.4 (91)	16.2 (65)
Irritable*	26.8 (119)	31.6 (130)	36.2 (154)	29.9 (120)
Withdrawn	5.9 (26)	5.6 (23)	7.0 (30)	7.5 (30)
Substance Use**	0.7 (3)	2.2 (9)	3.5 (15)	1.0 (4)
Attempted Suicide	6.9 (29)	5.6 (22)	7.9 (33)	6.1 (24)
Suicidal Ideation	14.6 (62)	15.7 (62)	18.8 (78)	12.0 (47)

* $p < .05$; ** $p < .01$.

tempt and an intermediate or higher score on the scales for unhappy or withdrawal or irritability and anxiety. As with other suicide risk instruments, the CSS has the potential for having high (0.88) sensitivity at the expense of specificity. Sensitivity remained high but was somewhat reduced (0.75) when the algorithm having the best balance with specificity (0.83) was used. Positive predictive values for this algorithm ranged from 16% to 33%, based on the hypothesized prevalence of diagnosis. We believe that the validity criterion matches empirical evidence for the conditions that are most likely to lead to suicide. The use of a highly structured criterion has the separate advantage of allowing reliable duplication. Although the specificity is generally better than for other instruments, it is our impression that the relatively low specificity requires that it be used in the context of a two-stage procedure. Test-retest reliabilities of individual CSS items were in

the good to excellent range. The best-balanced CSS scoring algorithm had a test-retest reliability score of 0.32, in the poor range (Landis and Koch, 1977).

The CSS compares favorably to the BDI, which had slightly higher test-retest reliability but was both less sensitive and less specific against the suicide validity criterion. We suspect that the CSS is more sensitive than the BDI in predicting the suicide validity criterion either because of its longer time frame (3 months versus 1 week) or because the CSS asks about both suicidal ideation and about suicide attempt behaviors.

These measures of accuracy cannot readily be compared with the sensitivities of 0.87 and 0.91 and specificity of 0.60 of the 20-item Suicide Risk Scale (SRS) (Thompson and Eggert, 1999), which has been tested in a sample of high-risk students and validated against two clinician-completed, comprehensive, unstructured assessments. The SRS has also been referenced against

TABLE 3
Validity of Columbia SuicideScreen Against DISC Criterion^a

Algorithm	CSS Items						Sens	Spec	Prev	PPV	NPV
	Att	Id	Dep	Anx	Subs	Help					
I.	YES	YES					0.88	0.72	0.04	0.12	0.99
									0.01	0.03	1.0
									0.05	0.14	0.98
									0.10	0.26	0.98
II.	YES						0.07	0.96	0.04	0.07	0.96
									0.01	0.02	0.99
									0.05	0.08	0.95
									0.10	0.16	0.90
III.		YES					0.09	0.92	0.04	0.04	0.96
									0.01	0.01	0.99
									0.05	0.06	0.95
									0.10	0.11	0.90
IV.			YES	or	YES	or	YES		0.04	0.93	0.04
									0.01	0.00	0.99
									0.05	0.03	0.95
									0.10	0.06	0.90
V.						YES	0.07	0.85	0.04	0.02	0.96
									0.01	0.00	0.99
									0.05	0.02	0.95
									0.10	0.05	0.89
VI.	YES	or	YES	and	YES	and	YES		0.75	0.83	0.04
									0.01	0.05	1.0
									0.05	0.19	0.98
									0.10	0.33	0.97

Note: DISC = Diagnostic Interview Schedule for Children version 2.3; CSS = Columbia SuicideScreen; Att = ever attempt on CSS; ID = current suicidal ideation on CSS; Dep = depression on CSS; Anx = anxiety on CSS; Subs = substance problem on CSS; Help = request for help on CSS; Sens = sensitivity; Spec = specificity; Prev = prevalence; PPV = positive predictive value; NPV = negative predictive value.

^a Suicidal ideation or previous suicide attempt and a DISC DSM-III-R diagnosis of major depressive disorder or dysthymia or substance abuse/dependence. Positive scores on each of these items required impairment.

the Suicide Ideation Questionnaire-Junior High (SIQ-JR), yielding a sensitivity of 100% at 31 and 23 cutoff scores and a specificity of 54% (using a cutoff of 31) and 57% (using a cutoff of 23) (Thompson and Eggert, 1999).

The SIQ-JR (Reynolds, 1991) is a 15-item self-completion scale that assesses frequency of suicidal cognitions on a 7-point Likert scale. It is frequently used with the Suicidal Behavior Inventory, a semistructured interview that rates aspects of suicidal behavior. Suicidal ideation and behavior have been referenced against each other, but, as each measures a different aspect of suicidality, it is not clear whether differences reflect content or their adequacy as measuring devices. The SIQ-JR does not assess predisposing conditions such as depression or substance abuse. In a small but unique study, the SIQ-JR was used to predict suicidality in the context of a very intense epidemic of suicidal behavior over the short term. It attained sensitivity of 0.80, specificity of 0.86, and a positive

predictive value of 27% when using a future attempt as the criterion (Keane et al., 1996), but it is not clear how this would generalize to a less-acute community sample.

The Suicide Probability Scale (SPS) (Cull and Gill, 1988) is a 36-item self-completion questionnaire measuring suicidal ideation, hopelessness, low self-esteem, and hostility toward others. Cutoffs were derived empirically (Larzelere et al., 1996) and were tested in a high-risk population. Its sensitivity is well below that of other similar scales.

In sum, the CSS, which has a novel structure, is more specific and has comparable sensitivity when compared to other screens.

Limitations

A screening program to prevent suicide has to strike a balance between (1) high sensitivity, so that individuals who are at risk for suicide or suicidal behaviors are not missed; and (2) adequate specificity, so that the

TABLE 4
Validity of BDI Against DISC Criterion^a

BDI	DISC Pos	DISC Neg	Sens	Spec	Prev of Disease	PPV	NPV
≥16							
Pos	44	116					
Neg	21	404	0.68	0.78	0.04	0.11	0.98
					0.01	0.03	1.0
					0.05	0.14	0.98
					0.10	0.26	0.96
≥21							
Pos	29	61					
Neg	36	459	0.45	0.88	0.04	0.14	0.98
					0.01	0.04	0.99
					0.05	0.16	0.97
					0.10	0.29	0.94

Note: BDI = Beck Depression Inventory; DISC = Diagnostic Interview Schedule for Children; Sens = sensitivity; Spec = specificity; Prev = prevalence; PPV = positive predictive value; NPV = negative predictive value.

^a Suicidal ideation or previous suicide attempt and a DISC DSM-III-R diagnosis of major depressive disorder or dysthymia or substance abuse/dependence. Positive scores on each of these items required impairment.

operation of the program is not overburdened with false positives. Although the sensitivity of the CSS is excellent, in practice a specificity of 0.83 would deliver many who were not at risk for suicide, and that could reduce the acceptability of a school-based prevention program. The CSS's positive predictive value of 16% (determined by a weighted prevalence of DISC positive in the sample) would result in 84 nonsuicidal teens being referred for further evaluation for every 16 youths correctly identified. This problem could be addressed by restricting the screen to pathognomonic items such as prior attempt, but, as these are not universally present in those at risk, this would greatly reduce the sensitivity of the screen. Requiring a second-level evaluation would seem preferable. However, it is important not to lose sight of the fact that many of these so-called false-positive cases may be experiencing painful depressive symptoms with social and academic impairment and are likely to benefit from treatment.

The test-retest reliability of the CSS was lower than the SIQ, SIQ-JR, and SPS, which had better test-retest reliabilities at 0.72 (Reynolds, 1988), 0.89 (Reynolds and Mazza, 1999), and 0.92 (Cull and Gill, 1988), respectively. In our study, the test-retest sample was not a random probability sample but was a small convenience sample from two schools, and inferences from such a sample should be limited. There are consider-

able difficulties in knowing how to interpret low test-retest reliabilities in adolescent psychometrics. Low reliability implies poor validity coefficients; thus, improving reliability results in improved validity coefficients (Lord and Novick, 1968). This theory is valid in the presence of random error. Unfortunately, in test-retest studies of psychiatric measures, systematic and random errors are present. For example, discrepancies between responses at time 1 and time 2 frequently follow a systematic pattern, with fewer endorsements at time 2 than at time 1 (Lucas et al., 1999). The reasons for the phenomenon of symptom attenuation that was found in this sample (80% of inconsistent replies) are not known, nor, more importantly, do we know whether inconsistent responses are less valid or predictive than consistent endorsements. Since we do not know which is more valid (test or retest), this further supports the idea of imposing a second-stage evaluation for those youths deemed positive by the screen.

Poor test-retest reliability could be related to the ephemeral nature of suicidal ideation and depressive feelings among teens (Larsson et al., 1991). We had limited data on reliability and ethnicity, but Hispanics, who constituted 39% ($n = 33$) of the sample, had poor test-retest reliabilities on the best-balanced CSS scoring algorithm at 0.44. Other ethnic group test-retest reliabilities could not be calculated due to lack of variation and small cell sizes. There appeared to be no variation in reliability between the ethnic-stratified and full samples. Nonetheless, we are reluctant to draw inferences from the limited data presented in this paper because the test-retest sample was not representative.

There are some limitations specific to the screening procedure. This study was conducted on seven unselected convenience samples instead of a random sample. The samples were chosen to represent educationally, ethnically, and socially diverse schools. This study was further limited by an appreciable rate of student absentees. Between-school participation rates varied from 47% to 70%, with an outlier of 91% at the all-girls parochial school. Screen-positive rates varied from 22% to 33%, with an outlier of 44% at an inner-city public school. Demographic and school features influence participation rates for a study of this kind. Our participation rates fell below the standard of 80%, leading one to question the generalizability of our results. However, in general, the rates of CSS-positive-and-DISC-positive youths ranged from 14% to 25.4%. Thus, despite different participation rates, the CSS in

combination with the DISC appeared to perform the same across the majority of the school cohorts. Nonetheless, it will be important for further research to find the optimal incentive and to determine the generalizability of our findings.

Clinical Implications

The second-stage evaluation used in the present study was the interviewer-administered, computer-assisted interview technique using the DISC format and script. Space limitations can be an important problem in school-based screening, although the introduction of a voice-administered version of the DISC (Shaffer et al., 2000; Wasserman et al., 2002) should ease that problem. The Voice-DISC can be given simultaneously to a number of students sitting together in a single classroom. Confidentiality is preserved through the students listening to the questions through headphones and responding on laptop computers placed in front of them. This approach also obviates the need for individual interviewers. Further, because self-completion interviews yield higher endorsement rates for suicide, substance use, and sexual items (Turner et al., 1998), this approach might yield more valid results than the interviewer-administered format.

The value of diagnosing mood disorders and/or suicidality is considerable. These disorders are regularly underdiagnosed. While 22% of 7- to 11-year-olds visiting their pediatrician had one or more *DSM-III* diagnoses, a review of studies on pediatric practices showed that only 4% to 7% of youths are identified (Costello, 1986; Costello et al., 1988). In the present sample (Scott, unpublished), professionals in the school setting, including the principal and guidance counselors, were not aware of almost half of the youth who were identified by the CSS and validated on the DISC as at risk.

A screening technique of this type might be the most effective strategy for suicide prevention. Research on telephone crisis lines and clinics has failed to show an impact on community suicide rate (Mishara and Daigle, 2000; Shaffer et al., 1990). A variety of psychosocial interventions, including access to resources and skills training (Garland et al., 1989; Hawton et al., 2000) have, similarly, been found to be ineffective. On the other hand, there is growing evidence that the declining suicide rate in adolescents may be due to greater use of antidepressant medication (Carlsten et al., 2001; Isaacson et al., 2000; Olfson et al., 2002; Olfson et al., 2003). For antidepressant medication to be used effec-

tively, cases of depression need to be identified, adding to the importance of screens of the type described here.

REFERENCES

- Aitken RCB (1969), Measurement of feelings using visual analog scales. *Proc R Soc Med* 62:989-993
- Beck AT, Steer RA (1993), *Beck Depression Inventory Manual*. San Antonio, TX: Psychological Corporation
- Brent DA, Baugher M, Bridge J, Chen T, Chiappetta L (1999), Age- and sex-related risk factors for adolescent suicide. *J Am Acad Child Adolesc Psychiatry* 38:1497-1505
- Carlsten A, Waern M, Ekedahl A, Ranstam J (2001), Antidepressant medication and suicide in Sweden. *Pharmacoepidemiol Drug Saf* 10:525-530
- Costello EJ (1986), Primary care pediatrics and child psychopathology: a review of diagnostic, treatment, and referral practices. *Pediatrics* 78:1044-1051
- Costello EJ, Costello AJ, Edelbrock C et al. (1988), Psychiatric disorders in pediatric primary care. *Arch Gen Psychiatry* 45:1107-1116
- Cotton CR, Range LM (1993), Suicidality, hopelessness, and attitudes toward life and death in children. *Death Stud* 17:185-191
- Cull JG, Gill WS (1988), *Suicide Probability Scale*. Los Angeles: Western Psychological Services
- Eggert L, Thompson E, Hering J (1994), A Measure of Adolescent Potential for Suicide (MAPS): development and preliminary findings. *Suicide Life Threat Behav* 24:359-381
- Garland A, Shaffer D, Whittle B (1989), National survey of school-based adolescent suicide prevention programs. *J Am Acad Child Adolesc Psychiatry* 28:931-934
- Goldston D (2003), *Measuring Suicidal Behavior and Risk in Children and Adolescents*. Washington, DC: American Psychological Association
- Groholt B, Ekeberg O, Wichstrom L, Haldorsen T (1998), Suicide among children and younger and older adolescents in Norway: a comparative study. *J Am Acad Child Adolesc Psychiatry* 37:473-481
- Hawton K, Townsend E, Arensman E et al (2000), Psychosocial versus pharmacological treatments for deliberate self harm (Cochrane Review). In: The Cochrane Library, 4, 2000. Update Software, Oxford. *Cochrane Database Syst Rev* CD001764
- Isaacsson G (2000), Suicide prevention: a medical breakthrough? *Acta Psychiatr Scand* 102:113-117
- Joiner RE, Pfaff JJ, Acres JG (2002), A brief screening tool for suicidal symptoms in adolescents and young adults in general health settings: reliability and validity data from the Australian National General Practice Youth Suicide Prevention Project. *Behav Res Ther* 40:471-481
- Keane EM, Wiegman-Dick R, Bechtold DW, Manson SM (1996), Predictive and concurrent validity of the suicidal ideation questionnaire among American Indian adolescents. *J Abnorm Child Psychol* 24:735-747
- Kendall PC, Hollon SD, Beck AT, Hammen CL, Ingram RE (1987), Issues and recommendations regarding use of the Beck Depression Inventory. *Cognit Ther Res* 11:289-299
- Landis JR, Koch GG (1977), The measurement of observer agreement for categorical data. *Biometrics* 33:159-174
- Larsson B, Melin L, Breitholtz E, Andersson G (1991), Short-term stability of depressive symptoms and suicide attempts in Swedish adolescents. *Acta Psychiatr Scand* 83:385-390
- Larzelere RE, Smith GL, Batenhorst LM, Kelly DB (1996), Predictive validity of the Suicide Probability Scale among adolescents in group home treatment. *J Am Acad Child Adolesc Psychiatry* 35:166-172
- Lord FM, Novick MR (1968), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley
- Lucas CP, Fisher P, Piacentini J et al (1999), Features of interview questions associated with attenuation of symptom reports. *J Abnorm Child Psychol* 27:429-437
- Mishara BL, Daigle MS (2000), *Helplines and Crisis Intervention Services: Challenges for the Future in Suicide Prevention: Resources for the Millennium*. Philadelphia: Brunner/Mazel
- Mufson L, Veltling D (2002), Psychotherapy for depression and suicidal behavior in children and adolescents. In: *The Many Faces of Depression in Children and Adolescents*, Shaffer D, Waslick BD, eds. Washington, DC: American Psychiatric Publishing, pp 37-72

- National Center for Health Statistics, Centers for Disease Control (2001), Deaths: final data for 1999. *Natl Vital Stat Rep*, vol. 49, no. 8. Hyattsville, MD
- Olfson M, Marcus SC, Weissman MM, Jensen PS (2002), National trends in the use of psychotropic medications by children. *J Am Acad Child Adolesc Psychiatry* 41:514–521
- Olfson M, Shaffer D, Marcus SC, Greenberg T (2003), Relationship between antidepressant medication treatment and suicide in adolescents. *Arch Gen Psychiatry* 60:978–982
- Osman A, Kopper B, Barrios F, Osman J, Besett T, Linehan M (1996), The Brief Reasons for Living Inventory for adolescents. *J Abnorm Child Psychol* 24:433–443
- Reynolds WM (1988), *Suicidal Ideation Questionnaire (SIQ): Professional Manual*. Odessa, FL: Psychological Assessment Resources
- Reynolds WM (1991), A school-based procedure for the identification of adolescents at risk for suicidal behaviors. *Fam Community Health* 14:64–75
- Reynolds WM, Mazza J (1999), Assessment of suicidal ideation in inner-city children and young adolescents: reliability and validity of the Suicidal Ideation Questionnaire–JR. *Sch Psychol Rev* 28:17–30
- Schwab-Stone ME, Shaffer D, Dulcan MK et al. (1996), Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3). *J Am Acad Child Adolesc Psychiatry* 35:878–888
- Shaffer D, Fisher P, Dulcan MK et al. (1996a), The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study (Methods for the Epidemiology of Child and Adolescent Mental Disorders Study). *J Am Acad Child Adolesc Psychiatry* 35:865–877
- Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME (2000), NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* 39:28–38
- Shaffer D, Garland A, Fisher P, Bacon K, Vieland V (1990), Suicide crisis centers: a critical reappraisal with special reference to the prevention of youth suicide. In: *Preventing Mental Health Disturbances in Childhood*, Goldston FE, Heinecke CM, Pynoos RS, Yager J, eds. Washington, DC: American Psychiatric Press, pp 135–165
- Shaffer D, Gould MS, Fisher P et al. (1996b), Psychiatric diagnosis in child and adolescent suicide. *Arch Gen Psychiatry* 53:339–348
- Steer R, Kumar G, Beck A (1993), Hopelessness Scale: scale for suicidal ideation, suicidal intent scale with adolescents. *Adv Adolesc Ment Health* 3:219–231
- Thompson EA, Eggert LL (1999), Using the Suicide Risk Screen to identify suicidal adolescents among potential high school dropouts. *J Am Acad Child Adolesc Psychiatry* 38:1506–1514
- Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JM, Sonenstein FL (1998), Adolescent sexual behavior, drug use, and violence: increased reporting with computer-survey technology. *Science* 280:867–873
- Waslick BD, Kandel R, Kakouros A (2002), Depression in children and adolescents: an overview. In: *The Many Faces of Depression in Children and Adolescents*, Shaffer D, Waslick BD, eds. Washington, DC: American Psychiatric Publishing, pp 1–36
- Wasserman GA, McReynolds LS, Lucas CP, Fisher P, Santos L (2002), The voice DISC-IV with incarcerated male youths: prevalence of disorder. *J Am Acad Child Adolesc Psychiatry* 41:314–321