

Getting the Best Return on Your Screening Investment: An Analysis of the Suicidal Ideation Questionnaire and Reynolds Adolescent Depression Scale

Peter M. Gutierrez
Denver VA Medical Center

Augustine Osman
University of Texas at San Antonio

Abstract. Data from 64 adolescent inpatients admitted for serious suicidal ideation, 50 adolescent inpatients admitted following a suicide attempt, and 56 randomly selected high school control participants were used to evaluate the sensitivity, specificity, positive predictive value, and negative predictive value of the Suicidal Ideation Questionnaire (SIQ) and the Reynolds Adolescent Depression Scale—2nd Edition (RADS-2). The purpose of the study was to provide information necessary for selecting assessment tools to use in large-scale screenings of high school students for suicide risk. The hypothesis that a combination of mean scores on the two measures and critical item scores would provide the best clinical utility was partially supported. The SIQ was found to be very useful in discriminating between the study groups, whereas the RADS-2 did not perform as expected. The fewest false negatives would result from relying on RADS-2 scale scores ≥ 63 and SIQ scale scores ≥ 14 . However, the parsimonious and hence most cost-effective approach would be to rely on SIQ scale scores ≥ 14 in combination with SIQ critical Items 2, 3, and 4 for screening purposes.

Systematic screening to identify adolescents at risk is an important component of secondary suicide prevention efforts (e.g., Garland, Shaffer, & Whittle, 1989; Miller & DuPaul, 1996; Shaffer, Garland, Vieland, Underwood, & Busner, 1991; Shaffer et al., 2004); it is a task to which school psychologists are well suited. When conducting such screenings, it is important to assess students for a range of suicide-related risk behaviors (e.g., depressive symptoms and thoughts about suicide (Perez, 2005). A dilemma for schools

is that suicide is a low-base-rate behavior (Kung, Hoyert, Xu, & Murphy, 2008) and large-scale screenings can be quite costly. However, 14.5% of adolescents in a national survey reported seriously considering suicide and 11.3% had made a suicide plan in the prior year (Eaton et al., 2008). Adolescents are clearly reporting risk for suicide, and the costs of adolescent suicide in terms of emotional impact on the school, family, and friends are enormous (Gutierrez & Osman, 2008). As a result, suicide risk screenings seem to be a

Correspondence regarding this article should be addressed to Peter M. Gutierrez, Denver VA Medical Center, MIRECC, 1055 Clermont St., Denver, CO 80220; E-mail: peter.gutierrez@va.gov

Copyright 2009 by the National Association of School Psychologists, ISSN 0279-6015

highly defensible investment of school resources. The challenge is determining how to get the best return on that investment.

Several options exist for assessing potentially high-risk adolescents; two of the most widely disseminated are the Columbia TeenScreen (Columbia University TeenScreen Program, 2007) and the Signs of Suicide (Aseltine & DeMartino, 2004) program. Both have been evaluated by the Substance Abuse and Mental Health Services Administration and are listed in their National Registry of Evidence-based Programs and Practices (Substance Abuse and Mental Health Services Administration, 2006, 2007). A major advantage of these programs is that they are standardized and come with significant amounts of supporting materials. However, there are costs associated with both.

TeenScreen requires supplies (which include laptop computers), postage, and three staff members to administer the program (Substance Abuse and Mental Health Services Administration, 2007). Free consultation, training, and technical assistance are available from the program administrators. Signs of Suicide kits, which include a manual and training video for staff, must be purchased (Substance Abuse and Mental Health Services Administration, 2006). The program also offers in-person training, although it indicates that costs vary depending on the needs of the school. In addition to the financial costs of these programs, which will exist to some extent regardless of the screening method that a school chooses to use, there is currently an absence of published data on the predictive validity (i.e., association between screening results and suicide-related target behaviors) of either program. However, there is evidence that participants in the Signs of Suicide program may be less likely to report a suicide attempt in the 3 months following participation (Aseltine & DeMartino, 2004). This finding, however, is evidence of the effectiveness of the prevention component of the program and not the predictive validity of screening. In other words, this finding does not indicate if adolescents identified as high risk by the screening are more likely to engage in suicide-related behaviors in the future and if students not identified as at

risk will not. As a result, schools may wish to consider other options for screening students that have evidence of predictive validity.

From a statistical standpoint, this requires selecting measures with the most acceptable mix of sensitivity, specificity, positive predictive value, and negative predictive value (Gutierrez & Osman, 2008). In the context of the current study, sensitivity refers to the percentage of inpatient adolescents with a history of suicide attempts who were correctly identified as suicide attempters. It also refers to the percentage of the inpatient adolescents whose suicidal ideation was serious enough to warrant admission being correctly identified in the suicidal ideation group. By contrast, specificity refers to the percentage of youth without a history of suicide attempts (i.e., high school controls) who were correctly identified as controls and the percentage of youth without serious ideation being identified as controls. Positive predictive value refers to the percentage of youth who were predicted to be in the suicide attempter and suicidal ideation groups who were actually suicide attempters or engaging in serious ideation. Negative predictive value refers to the percentage of youth who were predicted as controls (i.e., nonattempters and nonideators) who were actually controls (Loong, 2003). Positive predictive value takes into account false positives and negative predictive value takes into account false negatives. These values vary as a function of the prevalence of a given disorder, in this case suicidal ideation and suicide attempts, in the population of interest.

Using well-chosen self-report measures can be an efficient and effective way to assess adolescents for a range of suicide-related behaviors and risk factors (e.g., Gutierrez & Osman, 2008). This approach is in line with recommendations in the literature (e.g., Reynolds, 1991) to design school-based screening programs such that the initial assessment attempts to evaluate every student with brief measures and additional, more comprehensive assessments are conducted as needed. Self-report measures have the advantage of being easy to administer, they yield data of similar reliability and validity as clinical interviews,

and they may actually provide better information about sensitive subjects such as suicide, because adolescents are able to answer questions without having to verbalize their thoughts (Erdman, Greist, Gustafson, Taves, & Klein, 1987; Kaplan et al., 1994; Kendall, Cantwell, & Kazdin, 1989). A large number of measures exist from which to choose, and a comprehensive review of them is beyond the scope of this article. Goldston's text (2003) is the most comprehensive review to date of available tools to assess risk of suicide in children and adolescents. Gutierrez and Osman (2008) provide detailed psychometric information about a subset of self-report measures frequently used in adolescent research and clinical practice.

Two commonly used self-report measures in school-based screening programs designed to assess suicide risk factors are the Reynolds Adolescent Depression Scale—Second Edition (RADS-2; Reynolds, 2002) and the Suicidal Ideation Questionnaire (SIQ; Reynolds, 1988). The RADS-2 is the renormed version of the original RADS (Reynolds, 1987), which resulted in the creation of several subscales described under Measures later in this article. Data from 9,052 nonclinical and 297 clinical adolescents were used to assess a range of psychometric properties for the RADS-2 (Reynolds, 2002). Internal consistency reliability estimates for the depression total score in the nonclinical sample was .93, with the scale scores ranging from .80 to .87. Estimates were similar in the clinical sample: depression total .94, scale scores from .81 to .87. Item-total score correlations for the RADS-2 were acceptable, with the median correlation being .53. Criterion validity was supported by correlations with an established clinical interview for depression, with the depression total correlation being .82, and scale score correlations ranging from .54 to .79. Correlations between the RADS-2 scores and those on other self-report measures are also acceptable, as reported in the professional manual. Similarly, the SIQ has demonstrated an acceptable internal consistency reliability estimate of .97, with mean interitem correlations $\geq .53$ (Reynolds, 1988). The abil-

ity of the SIQ to differentiate between clinical and nonclinical groups is supported by several studies (e.g., Colle, Belair, DiFeo, Weiss, & LaRoche, 1994; Pinto, Whisman, & McCoy 1997). Evidence of concurrent–convergent validity for the SIQ comes from a range of studies examining correlations between it and other well-established self-report measures (e.g., Mazza, 2000; Reinecke, Du Bois, & Schultz, 2001).

These two measures have the advantage of being well validated, easy to administer, and standardized on large samples of nonclinical and clinical adolescents. In addition, they tap two essential components of suicide risk: depressive symptoms and current thoughts about suicide. Finally, they are appropriate for group administration, require no special equipment, and can be administered by any school psychologist, social worker, or other mental health professional. According to the professional manuals (Reynolds, 1988, 2002), the SIQ can be administered in 10 min or less and the RADS-2 in 5–10 min.

Specific Aims of the Study

The purpose of the current study was to examine a range of psychometric characteristics of the RADS-2 (Reynolds, 2002) and SIQ (Reynolds, 1988), with a particular emphasis on the ability of scores on the two measures to correctly identify adolescents in different suicide risk categories. Accordingly, the participants included clinical and nonclinical adolescents ages 14–18 years.

Method

Procedure and Participants

The present study used data from large clinical and nonclinical research projects during the 2004 school year to test hypotheses regarding the two measures. Consistent with the standardization samples for the study measures, we recruited only youths ages 14 and older as participants in the study. Following protocol approval from the universities' and hospital institutional review boards for the studies, potential participants provided written

parental informed consent and individual assent before completing the questionnaire packets.

In the high school sample, the questionnaire packets were administered in small groups by trained research assistants, supervised by the lead author. Details of the Adolescent Risk Project, from which data for the current analyses were drawn, have been described in detail by Gutierrez, Watkins, and Collura (2004). Beginning with the 2004 school year, the Adolescent Risk Project was adopted by the school administration as their screening program. All students enrolled in health and government classes (approximately 50% of the overall student body of roughly 1,200) were screened. Their parents were informed of the program and given the opportunity by the school to opt out, but none chose that option. Students were also given the option of not being screened, based on the school's belief that voluntary participation would yield more reliable results. Of students present on screening days, 20% refused to participate.

In the clinical setting, data were collected by a trained school teacher having extensive experience with questionnaire administration for the hospital. Inpatient data were collected within 2–3 days of admission to the unit to control for potential for medication use and the effect of long-term hospitalization. Data collected in the studies were confidential, unless high school participants endorsed critical items (Reynolds, 1988; Reynolds, 2002) indicating a need for additional assessment of risk (Gutierrez et al., 2004). Typically, between 5% and 10% of participating students on a given day required follow-up individual assessment by the clinical psychology graduate student in charge of data collection. Those students meeting preestablished risk criteria (Gutierrez et al., 2004) were then provided appropriate referrals to the school psychologist or school social worker for additional assessment. Their data were not excluded from study analyses.

High school sample. Participants were recruited from an urban high school in the

Midwest. The sample ($N = 430$; $M_{\text{age}} = 15.34$, $SD = 1.16$ years) comprised 189 boys ($M_{\text{age}} = 15.27$, $SD = 1.16$ years) and 241 girls ($M_{\text{age}} = 15.40$, $SD = 1.17$ years). There was no statistically significant difference between boys and girls in age, $t(428) = 1.17$, $p = .24$. Of the 430 participants, 50.0% were Caucasian, 32.3% were African American, 11.2% were Hispanic American, 1.6% were Asian American, and 4.9% were other ethnic/racial groups.

Using the SPSS's Select Cases module, we randomly selected 56 participants (36 boys and 20 girls; $M_{\text{age}} = 15.04$, $SD = 1.06$ years) to serve as the nonclinical control subgroup in some of the validation analyses.

Psychiatric inpatient sample. Participants were consecutive admissions to two adolescent units of a state psychiatric hospital ($N = 202$; $M_{\text{age}} = 15.49$, $SD = 0.99$ years) in a relatively rural area of the Midwest. The study sample included 98 boys ($M_{\text{age}} = 15.53$, $SD = 0.98$ years) and 104 girls ($M_{\text{age}} = 15.45$, $SD = 1.00$ years). In this sample, 74.8% were Caucasian, 12.9% were African American, 5.8% were Hispanic American, 2.5% were Asian American, and 4.0% were other ethnic/racial groups. Boys and girls did not differ significantly in age, $t(200) = 5.6$, $p = .57$. Psychiatric diagnoses were determined by the multidisciplinary assessment and treatment team (i.e., as based on *Diagnostic Statistical Manual*, 4th edition/*Diagnostic Statistical Manual*, 4th edition, text edition, criteria; American Psychiatric Association, 1994, 2000). Of the participants, 41.6% met criteria for major depressive disorder, 17.8% for oppositional defiant disorder, 13.4% for conduct disorder, 9.4% for dysthymia, 5.9% for attention deficit hyperactivity disorder, 6.9% for substance use disorder, and 5.0% for generalized anxiety disorder. Because the units focus on the management of youths with a range of behavior problems, severe psychiatric disorders, such as youth with bipolar disorder and psychosis, were not identified as unique diagnostic groups in these analyses.

Using the team assessment information, we identified two suicidal subgroups for fur-

ther validation analyses. Of the inpatient participants, 64 youths (15 boys and 49 girls) who were admitted because of severe suicidal ideation were assigned to the suicidal ideation subgroup ($M_{\text{age}} = 15.59$, $SD = 0.92$ years) and 50 youths (19 boys and 31 girls) who were admitted because of suicide attempts were assigned to the suicide attempter subgroup ($M_{\text{age}} = 15.28$, $SD = 1.14$ years). For the suicide attempter group, the methods of attempts were documented by the team to include overdose of prescription medications ($n = 25$), cutting/stabbing ($n = 19$), attempts to hang/strangle ($n = 4$), and attempts to suffocate/choke ($n = 2$). There were no statistically significant differences among these groups on demographic variables of age ($p = .11$), gender composition ($p = .10$), and ethnicity ($p = .49$).

Measures

Participants completed a brief demographic information questionnaire (e.g., age and gender) and two self-report measures: the SIQ (Reynolds, 1988) and the RADS-2 (Reynolds, 2002). Critical items were drawn from each completed instrument for some of the key analyses.

SIQ (Reynolds, 1988). The SIQ is a 30-item self-report questionnaire that is used to screen for the frequency of suicidal ideation in adolescents. The item response format of the SIQ ranges from 0 (*I never had this thought*) to 6 (*almost every day*). A total score (range = 0–180) is computed to derive a severity of suicidal ideation score; higher levels indicate greater frequency of suicidal ideation. An example of an item is “I thought about death.” The psychometric properties of the SIQ have been supported in several clinical and nonclinical adolescent samples (e.g., Pinto et al., 1997; Spirito, Stark, Fristad, Hart, & Owens-Stively, 1987). The SIQ total mean score and the 8 critical items (Items 2, 3, 4, 7, 8, 9, 13, and 18) were used in the validation analyses. For the total SIQ score, the recommended raw cutoff score indicative of the potential for suicidal risk is 41 (Reynolds, 1988).

RADS-2 (Reynolds, 2002). The RADS-2 is a 30-item, self-report instrument designed to evaluate five domains of the depression severity construct: Dysphoric Mood (8 items), Anhedonia/Negative Affect (7 items), Negative Self-Evaluation (8 items), and Somatic Complaints (7 items). Each item is rated on a 4-point scale ranging from 1 (*almost never*) to 4 (*most of the time*). The professional manual reports good estimates of internal consistency for the total and scale scores. A total RADS-2 score (range = 30–120) is computed to derive severity of depressive symptoms; higher scores suggest greater level of depressive symptoms. An example of an item is “I feel upset.” The factor structure and estimates of concurrent validity of this instrument are also reported in the professional manual. The RADS-2 total mean score and the six critical items (Items 3, 6, 14, 19, 20, and 30) were used in the validation analyses. For the total RADS-2 score, the recommended raw cutoff score that is suggestive of the risk for major depressive symptoms is 76 (Reynolds, 2002).

Data Analytical Plan

Preliminary analyses. First, to identify potential covariates for use in the primary analyses, we examined differences between the clinical and nonclinical groups on several demographic variables using independent-samples t test and χ^2 analytic procedures. Second, we decided to conduct preliminary analyses for the high school and psychiatric inpatient samples separately to examine estimates of internal consistency reliability and relationships between scores on the RADS-2 and SIQ measures. Accordingly, we computed Cronbach’s alpha reliability and Pearson correlation analyses. In each sample, we expected the Cronbach’s alpha estimate to exceed the traditional cutoff score of .80 (Clark & Watson, 1995). Moreover, we expected moderate correlations among scores on the RADS-2 and SIQ within each sample (i.e., zero-order correlations of .40–.60). In addition, we reported descriptive statistics (means and standard deviations) for the RADS-2 and SIQ-2 scale

scores for the high school and psychiatric inpatient samples.

Primary analyses. To examine the ability of the RADS-2 and SIQ total mean scores to differentiate reliably the responses of the high school and the inpatient youths, we used the independent-sample *t* test procedure. We expected the inpatient participants to obtain significantly higher scores on the RADS-2 and the SIQ than the high school participants in each of the analyses. We used Cohen's *d* (1988) estimation procedure to guide interpretations of the results.

To explore potential differences between the high school and the inpatient youths on each of the RADS-2 and SIQ items, we conducted a one-way multivariate analysis of variance (MANOVA). In addition, we reported descriptive statistics (means and standard deviations) for the RADS-2 and SIQ item scores for each sample. We note that screening of the individual item and scale scores (e.g., Shapiro-Wilks's estimates) did not show substantial deviations from normality on the RADS-2 and SIQ item and scale scores. Thus, scores on the items and scales were not transformed for any of the analyses.

To examine the clinical utility of the RADS-2 and the SIQ total mean scores and critical items in differentiating the responses of the suicide attempter and high school control subgroups and suicide ideation and high school control subgroups, we conducted a series of logistic regression and receiver operating characteristic (ROC) curve analyses.

First, separate binary logistic regression analyses were conducted to evaluate the utility of scores on the RADS-2 and SIQ total mean scores to differentiate the responses of the subgroups. In particular, the mean raw score on each scale was entered in the analysis. In addition to the odds ratio (OR), only scale scores with 95% lower and upper confidence intervals (CIs) that reached 1.0 (95% CI \geq 1.0) were identified as useful for further analysis. Next, ROC curve analysis was conducted to evaluate the diagnostic efficacy of scores on each scale and examine a range of cutoff scores with acceptable sensitivity and speci-

ficity estimates. Again, using recommendations in the extant literature (see Lonigan, Anthony, & Shannon, 1998; Matthey & Petrovski, 2002), only instruments with sensitivity $\geq .70$ and specificity $\geq .80$ were considered clinically useful for screening purposes (cf. Pinto et al., 1997). *Sensitivity* is conceptualized as the ability of scores on the instrument to correctly identify youth within the prespecified risk status. *Specificity* is the proportion of youth with the prespecified risk status (e.g., suicide ideation) who are actually positive for the risk status. Positive predictive and negative predictive values were calculated to further assess the clinical utility of the measures. Finally, these analyses were extended to the RADS-2 and SIQ critical items.

Results

Group Differences on Demographic Variables

To identify potential covariates, we conducted independent-sample *t* test and χ^2 analyses. Comparisons of the high school and clinical youth participants on key demographic variables showed that the participants were similar in age, $t(630) = 1.54, p = .12$, and gender composition, $\chi^2(1, N = 632) = 1.15, p = .30$. However, there was a higher proportion of Caucasians in the inpatient sample than in the high school sample, $Z = 6.73, p < .001$. In addition, there were higher proportions of African American, $Z = 5.95, p < .001$, and Hispanic, $Z = 2.36, p < .02$, youth in the high school sample than in the psychiatric inpatient sample. No significant differences were observed with regard to the other ethnic groups (all values of $p > .05$). Thus, these demographic variables were not included as covariates in any of the current analyses.

Estimates of Internal Consistency

The Cronbach's alpha estimate for the RADS-2 for the high school sample (.92, average interitem correlation = .27) was slightly lower than the estimate obtained for the psychiatric inpatient (.95, average interitem correlation = .38) sample. In addition, the Cron-

Table 1
High School Sample: Means, Standard Deviations, and Correlations Among
the RADS-2 and SIQ Scale Scores

Measure	1	2	3	4	5	6
1. RADS-2	—					
2. RADS-DM	.89**	—				
3. RADS-AN	.67**	.42**	—			
4. RADS-NS	.88**	.71**	.50**	—		
5. RADS-SC	.88**	.75**	.45**	.67**	—	
6. SIQ	.55**	.46**	.38**	.57**	.41**	—
<i>M</i>	56.00	16.86	11.75	12.90	14.50	11.55
<i>SD</i>	13.97	4.91	3.17	4.30	4.24	16.74

Note. RADS-2 = Reynolds Adolescent Depression Scale—2nd edition; RADS-DM = RADS—Dysphoric Mood; RADS-AN = RADS—Anhedonia/Negative Affect; RADS-NS = RADS—Negative Self-Evaluation; RADS-SC = RADS—Somatic Complaints; SIQ = Suicidal Ideation Questionnaire.

** $p < .01$.

bach's alpha estimate for the SIQ for the high school sample (.95, average interitem correlation = .40) was slightly lower than the estimate obtained for the psychiatric inpatient (.98, average interitem correlation = .63) sample. However, estimates for both measures are acceptable indicators of internal consistency.

Correlational Analysis

To examine links between scores on the RADS-2 and the SIQ in the separate high school and psychiatric inpatient groups, we computed the Pearson correlation analyses. Table 1 shows the means, standard deviations, and intercorrelations among the study measures for the high school nonreferred sample. Table 2 shows the means, standard deviations, and intercorrelations among the study measures for the inpatient psychiatric sample. The results of the independent correlation analysis showed that the link between the RADS-2 total mean score and the SIQ total mean score were not substantially different for the high school and the psychiatric inpatient samples, $Z = 0.34$, $p = .73$.

Group Differences on the RADS-2 Total Mean Score

To assess for substantive differences between the high school and the psychiatric in-

patient groups on the RADS-2 total mean score, we conducted an independent-samples t test analysis. The mean scores (and standard deviations) on the RADS-2 total and scales are presented in Tables 1 and 2 for the high school and psychiatric inpatient samples, respectively. It was expected that the psychiatric inpatient group would obtain substantially higher scores on the RADS-2 total mean score than the high school sample. Results showed that this hypothesis was supported, $t(630) = 5.99$, $p < .001$; Cohen's $d = 0.51$, 95% CI = 0.34, 0.68. The psychiatric inpatient youths scored significantly higher than the high school youths on the RADS-2 total mean score. The obtained estimate of 0.51 indicates clinically useful differences between these groups in the RADS-2 total mean score.

Group Differences on the SIQ Total Mean Score

To examine differences between the high school and the psychiatric inpatient groups on the SIQ total mean score, we conducted an independent-samples t test analysis. Tables 1 and 2 show the mean scores (and standard deviations) on the SIQ total scale for the high school and psychiatric inpatient samples, respectively. As expected, results showed that the psychiatric inpatient youth

Table 2
Psychiatric Inpatient Sample: Means, Standard Deviations, and Correlations
Among the RADS-2 and SIQ Scale Scores

Measure	1	2	3	4	5	6
1. RADS-2	—					
2. RADS-DM	.93**	—				
3. RADS-AN	.72**	.52**	—			
4. RADS-NS	.93**	.83**	.61**	—		
5. RADS-SC	.88**	.81**	.48**	.74**	—	
6. SIQ	.57**	.51**	.39**	.58**	.47**	—
<i>M</i>	63.93	18.63	12.91	15.95	16.44	42.40
<i>SD</i>	18.35	5.69	4.20	6.13	4.96	45.98

Note. RADS-2 = Reynolds Adolescent Depression Scale—2nd edition; RADS-DM = RADS—Dysphoric Mood; RADS-AN = RADS—Anhedonia/Negative Affect; RADS-NS = RADS—Negative Self-Evaluation; RADS-SC = RADS—Somatic Complaints; SIQ = Suicidal Ideation Questionnaire.

** $p < .01$.

scored significantly higher than the high school youths on the SIQ total mean score, $t(630) = 12.29, p < .001$; Cohen's $d = 1.05$, 95% CI = 0.87, 1.22. The obtained effect size estimate suggests clinically useful differences between the groups in the SIQ total mean score.

Descriptive Statistics for the RADS-2 Items

The mean scores (and standard deviations) of the RADS-2 items for the high school sample are presented in Table 3 (Columns 2 and 3). Table 3 (Columns 4 and 5) shows the mean scores (and standard deviations) of the RADS-2 items for the clinical inpatient sample. To assess potential differences between the high school and the psychiatric inpatient youth on the RADS-2 items, we conducted a one-way MANOVA. We found that the overall MANOVA was statistically significant, Pillai's trace = .45, $p < .001$, partial $\eta^2 = .45$. Using Cohen's d estimates of values ≤ 0.49 as small effect, 0.50–0.79 as medium effect, and 0.80 or higher as large effect, we found that 19 items had small effect sizes in differentiating between the groups, and 11 items had medium effect sizes. In particular, Items 13 (*discouragement*), 18 (*fatigue*),

and 22 (*anger*) were most trivial in differentiating between these groups.

Descriptive Statistics for the SIQ Items

The mean scores (and standard deviations) of the SIQ items for the high school sample are shown in Table 4 (Columns 2 and 3). Table 4 (Columns 4 and 5) shows the mean scores (and standard deviations) of the SIQ items for the psychiatric inpatient sample. To assess potential differences between the high school and psychiatric inpatient youth on the SIQ items, we conducted a one-way MANOVA. The overall MANOVA was found to be statistically significant, Pillai's trace = .42, $p < .001$, partial $\eta^2 = .30$. We found that only two of the items (Items 5 and 6) were trivial in differentiating between the groups. Eight of the items resulted in medium effect sizes in differentiating between the groups, and 20 resulted in large effect sizes in differentiating between the groups.

Clinical Utility of Responses on the RADS-2 Scale

Attempter versus control subgroups. Results of the analyses are presented in Table 5. When we used the total mean in the logistic

Table 3
Raw Score Means, Standard Deviations, and Cohen's *d* Values for the RADS-2

Abbreviated Item	High School Sample		Clinical Sample		Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
1. Reduced affect	1.40	0.58	1.70	0.75	0.47
2. School anxiety	3.08	0.76	2.44	0.96	0.77
3. Loneliness	2.19	0.95	2.50	0.99	0.32
4. Rejection	1.69	0.92	2.04	1.02	0.37
5. Self-worth	1.71	0.79	1.91	0.86	0.25
6. Withdrawal	1.83	0.94	2.22	1.01	0.41
7. Sadness	2.14	0.89	2.51	0.93	0.41
8. Crying	1.99	0.91	2.36	0.99	0.40
9. Worthlessness	1.71	0.90	2.08	0.98	0.40
10. Peers anhedonia	1.44	0.69	1.88	0.87	0.59
11. Feeling sick	1.81	0.84	2.00	0.96	0.22
12. Low self-worth	1.39	0.69	1.81	0.90	0.55
13. Discouragement	1.63	0.88	1.79	0.97	0.18
14. Self-injurious	1.30	0.66	1.79	1.02	0.62
15. Self-esteem	1.92	0.94	2.18	1.05	0.27
16. Irritability	1.99	0.91	2.34	0.95	0.38
17. Pessimism	2.07	0.99	2.39	1.03	0.32
18. Fatigue	2.54	0.95	2.48	0.97	0.06
19. Self-reproach	1.67	0.83	2.06	0.94	0.45
20. Self-deprecation	1.47	0.80	1.92	0.96	0.53
21. Self-pity	1.50	0.76	1.92	0.99	0.50
22. Anger	2.28	0.94	2.46	0.97	0.19
23. Reduced speech	1.90	0.91	2.04	0.92	0.15
24. Sleep disturbance	2.17	1.06	2.45	1.08	0.26
25. Others anhedonia	1.33	0.64	1.69	0.83	0.51
26. Worry	2.13	0.93	2.33	0.99	0.21
27. Stomachaches	1.60	0.83	2.07	0.95	0.54
28. Loss of interest	2.02	0.95	2.61	1.03	0.60
29. Appetite	2.58	0.93	1.88	0.88	0.77
30. Helplessness	1.51	0.69	2.09	1.01	0.72

Note. RADS-2 = Reynolds Adolescent Depression Scale—2nd edition. Critical items are indicated in bold.

regression analysis, we found that the OR was 1.06 and had the following desired CI: 95% CI = 1.03, 1.09. Next, in the ROC curve analysis, we found that the area under the curve (AUC) indicated moderate accuracy of the RADS-2 in differentiating between the suicide attempter and the high school nonreferred control youths, $AUC = .75$, $SE = .048$, 95% CI = 0.65, 0.83. Table 5 (upper portion)

shows several cutoff scores for this instrument. Using the predefined sensitivity and specificity criteria, we found that none of the cutoff scores attained the expected criteria.

Ideation versus control subgroups.

When the RADS-2 total mean score was used in the logistic regression analysis, the OR was 1.08; the desired CI was attained: 95%

Table 4
Raw Score Means, Standard Deviations and Cohen's *d* Values for the SIQ
Items

Abbreviated Item	High School Sample		Clinical Sample		Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
1. Thought not alive	0.63	1.06	1.69	2.02	0.74
2. Thought killing	0.35	0.75	1.61	1.79	1.06
3. Thought how	0.23	0.69	1.34	1.78	0.96
4. Thought when	0.15	0.53	1.31	1.77	1.06
5. People dying	1.38	1.69	1.90	2.01	0.29
6. Thought death	1.66	1.65	2.22	2.11	0.31
7. Suicide note	0.20	0.64	1.06	1.74	0.77
8. Writing a will	0.23	0.56	0.81	1.56	0.58
9. Plan to kill self	0.07	0.39	0.92	1.51	0.93
10. People happier	0.55	1.09	1.69	2.03	0.78
11. People would feel	0.64	1.11	1.76	2.15	0.74
12. Wish dead	0.36	0.84	1.43	1.93	0.83
13. How easy to end	0.27	0.74	1.51	2.00	0.97
14. Solves problems	0.21	0.58	1.51	2.03	1.05
15. Better off dead	0.32	0.88	1.51	2.00	0.89
16. Nerve to kill self	0.16	0.55	1.06	1.82	0.80
17. Never been born	0.42	0.90	1.56	2.09	0.82
18. Would kill self	0.15	0.54	1.19	1.75	0.96
19. Ways to kill	0.52	0.95	1.52	2.00	0.73
20. Would not do it	0.39	0.86	1.32	1.77	0.76
21. Having accident	0.48	0.92	1.33	1.93	0.64
22. Not worth living	0.35	0.86	1.47	1.92	0.86
23. Life is too rotten	0.23	0.76	1.58	2.05	1.03
24. Way to be noticed	0.12	0.54	1.11	1.85	0.87
25. Worth caring about	0.20	0.64	1.34	1.95	0.93
26. No one cared	0.35	0.93	1.41	1.99	0.78
27. Hurting myself	0.40	0.89	1.57	1.95	0.88
28. Nerve to kill	0.22	0.63	1.27	1.90	0.88
29. Things not better	0.19	0.69	1.37	1.94	0.96
30. Right to kill	0.11	0.61	1.02	1.77	0.81

Note. SIQ = Suicidal Ideation Questionnaire. Critical items are indicated in bold.

CI = 1.05, 1.11. In the subsequent ROC curve analysis, the AUC also indicated medium accuracy of the RADS-2 scores for differentiating between these subgroups, AUC = .82, SE = .038, 95% CI = 0.74, 0.88. Table 5 (lower portion) shows a range of cutoff values for the RADS-2. Overall, cutoff scores of 63 and 67 met the criteria for clinically useful screening. However, a score of 67 achieves superior specificity

with only a slight decrease in sensitivity and therefore was considered most useful.

Clinical Utility of Responses on the SIQ Scale

Attempter versus control subgroups.

Results of the analyses are presented in Table 6. When we used the total mean score in the

Table 5
Cutoff Scores on the RADS-2 by
Study Subgroups

Cutoff Estimate	Sensitivity (≥.70)	Specificity (≥.80)	PPV	NPV
Attempter vs. Control Subgroups				
50	.86	.45	.58	.78
55	.78	.68	.68	.78
61	.58	.75	.67	.67
63	.56	.80	.72	.67
67	.50	.86	.76	.66
76	.26	.95	.81	.59
Ideation vs. Control Subgroups				
50	.86	.45	.64	.74
55	.83	.68	.75	.78
61	.73	.75	.77	.71
(63)	.73	.80	.81	.73
(67)	.70	.86	.85	.72
76	.50	.95	.91	.62

Note. RADS-2 = Reynolds Adolescent Depression Scale—2nd edition; PPV = positive predictive value; NPV = negative predictive value. Values in parentheses meet the targets for an acceptably high level of sensitivity and specificity.

Table 6
Cutoff Scores on the SIQ by Study
Subgroup

Cutoff Estimate	Sensitivity (≥.70)	Specificity (≥.80)	PPV	NPV
Attempter vs. Control Subgroups				
(12)	.86	.80	.80	.87
(14)	.82	.89	.87	.85
(19)	.72	.91	.88	.79
27	.46	.96	.92	.67
42	.34	.98	.94	.63
49	.28	.98	.93	.60
58	.22	.98	.92	.59
Ideation vs. Control Subgroup				
(12)	100	.80	.85	100
(14)	100	.89	.91	100
(21)	.98	.91	.93	.98
(27)	.95	.96	.97	.95
(42)	.78	.98	.98	.80
(49)	.73	.98	.98	.76
(52)	.70	.98	.98	.74
58	.69	.98	.98	.73

Note. SIQ = Suicidal Ideation Questionnaire; PPV = positive predictive value; NPV = negative predictive value. Values in parentheses meet the targets for an acceptably high level of sensitivity and specificity.

logistic regression analysis, we found an OR of 1.21. As expected, the desired CI was attained: 95% CI = 1.07, 1.18. Next, in the ROC curve analysis, the AUC showed high accuracy, AUC = .90, SE = .03, 95% CI = 0.82, 0.95. Table 6 (upper portion) shows a range of cutoff values for the SIQ for differentiating the attempter and control groups. Cutoff scores of 12, 14, and 19 all met the preestablished criteria for clinically useful screening, but a score of 19 maximizes specificity while retaining acceptable sensitivity.

Ideation versus control subgroups.

Results of the analyses are presented in Table 6. In the logistic regression analysis involving scores on the mean SIQ for the groups, we found an OR of 1.17. The pre-defined CI was attained: 95% CI = 1.10, 1.26. In the subsequent ROC curve analysis, the related AUC attained high accuracy, AUC = 0.99,

SE = 0.01, 95% CI = 0.95, 0.99. Table 6 (lower portion) shows a range of cutoff values for the SIQ that are useful for differentiating between the groups. The maximum cutoff score for meeting the preestablished criteria is 52, which is the point at which acceptable sensitivity is achieved with the greatest specificity.

Overview of the Clinical Utility of Responses on the RADS-2 and SIQ Critical Items

Because of the unreliability of scores on single items for screening purposes, all the critical items of each instrument were entered simultaneously in the analyses to identify a set of items that may be useful for differentiating the responses of the suicidal and control participants (e.g., Pinto et al., 1997). The set of items retained was included in the final anal-

Table 7
Critical Items on the RADS-2 by Study Subgroups

Critical Item	OR	95% CI ≥ 1	Sensitivity ($\geq .70$)	Specificity ($\geq .80$)	PPV	NPV	AUC
Attempter vs. Control Subgroups							
3	1.61	1.06, 2.43	.54	.66	.59	.62	.62
6	2.12	1.36, 3.31	.38	.86	.70	.61	.70
14	2.76	1.55, 4.90	.48	.88	.77	.65	.68
19	2.55	1.49, 4.36	.72	.61	.62	.71	.69
20	2.34	1.40, 3.90	.62	.79	.72	.70	.69
30	2.17	1.28, 3.66	.70	.54	.58	.68	.66
Ideation vs. Control Subgroups							
3	2.66	1.72, 4.11	.70	.66	.70	.66	.74
6	3.15	2.02, 4.91	.83	.61	.71	.76	.78
14	3.72	2.14, 6.47	.61	.88	.85	.66	.76
19	3.52	2.14, 5.80	.80	.61	.70	.72	.77
20	4.02	2.39, 6.79	.81	.79	.81	.79	.81
30	3.37	2.08, 5.48	.81	.54	.67	.71	.77

Note. RADS-2 = Reynolds Adolescent Depression Scale—2nd edition; CI = confidence interval; OR = odds ratio; PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve.

ysis to help identify the set of critical items that is most useful for differentiating the responses of the subgroups. As in the previous analyses, logistic regression analyses were implemented separately for the RADS-2 and SIQ critical items. Unlike the previous analysis, cutoff scores were not computed for the single items. In each logistic regression analysis, the scored mean was the predictor variable and the suicidal (coded = 1) and high school (coded = 0) subgroups served as the dependent variable.

Clinical Utility of Responses on the RADS-2 Critical Items

Attempter versus control subgroups.

The RADS-2 critical items included in the initial logistic regression analyses were Items 3, 6, 14, 19, 20, and 30 (see Table 7). We examined the lower and upper CIs of each item, as we did with the total mean scores, to identify items useful for differentiating between the subgroups.

Although all the items met the CI crite-

tion, none of scores on the RADS-2 critical items met the preestablished sensitivity and specificity criteria for differentiating between the responses of the subgroups (see Table 7, upper portion).

Ideation versus control subgroups.

Results of the initial logistic regression analysis are presented in Table 7 (lower portion). Although each item met the CI criteria, again none of the items reached the expected sensitivity and specificity criteria for differentiating the responses of the subgroups.

Clinical Utility of Responses on the SIQ Critical Items

Attempter versus control subgroup.

The SIQ critical items included in the initial logistic regression analyses were as follows: Items 2, 3, 4, 7, 8, 9, 13, and 18. Using the preestablished CI criteria, all the items were identified as potentially useful for differentiating between the study groups (see Table 8, upper portion). However, when we examined

Table 8
Critical Items on the SIQ by Study Subgroups

Critical Item	OR	95% CI ≥ 1	Sensitivity ($\geq .70$)	Specificity ($\geq .80$)	PPV	NPV	AUC
Attempter vs. Control Subgroups							
(2)	31.24	9.72, 100.44	.90	.84	.83	.90	.90
3	8.37	3.30, 21.24	.66	.86	.81	.74	.77
4	13.70	4.30, 43.68	.62	.93	.89	.73	.78
7	3.22	1.52, 6.85	.48	.84	.73	.64	.67
8	2.52	1.15, 5.53	.30	.89	.71	.59	.60
9	29.01	3.70, 227.54	.46	.98	.96	.67	.72
(13)	3.39	1.74, 6.63	.70	.84	.80	.76	.79
18	7.51	2.47, 22.83	.52	.93	.87	.68	.73
Ideation vs. Control Subgroups							
(2)	37.78	7.80, 183.00	.97	.84	.87	.96	.97
(3)	17.29	5.46, 54.70	.92	.86	.88	.91	.94
(4)	38.35	9.51, 154.77	.92	.93	.94	.91	.95
(7)	5.42	2.53, 11.63	.77	.84	.85	.76	.84
8	3.56	1.80, 7.01	.56	.89	.86	.64	.75
(9)	43.92	5.57, 346.37	.72	.98	.98	.75	.86
(13)	3.61	2.02, 6.46	.83	.84	.86	.81	.87
(18)	27.43	7.64, 98.44	.88	.93	.93	.87	.92

Note. SIQ = Suicidal Ideation Questionnaire; CI = confidence interval; OR = odds ratio; PPV = positive predictive value; NPV = negative predictive value; AUC = area under the curve. Values in parentheses meet the targets for an acceptably high level of sensitivity and specificity.

the results for the specificity and sensitivity criteria, we found that only 2 of the 8 items (i.e., Items 2 and 13) met the preestablished criteria.

Ideation versus control subgroups.

Results of the logistic regression analyses are presented in Table 8 (lower portion). As can be seen, all the SIQ critical items met CI preestablished criterion. Examination of the results showed further that all the items except 1 (i.e., Item 8) met the sensitivity and specificity criteria.

Discussion

Preliminary analyses indicated that correlations among the RADS-2 subscale scores and between the mean RADS-2 and SIQ scores were in the expected direction and of adequate magnitude to support the link be-

tween depressive symptoms and thoughts of suicide. However, correlations were not so large as to suggest that the two scales are assessing the same construct—hence, there was support for the basic reliability and validity of the measures. In addition, because the correlations were similar in the high school and clinical samples, comparisons across subgroups of the two samples are justified.

We next compared the mean scores on the two measures across the high school and clinical sample to test the hypothesis that the inpatient adolescents would score significantly higher than the high school sample. For the RADS-2, both groups' mean score was below the clinical cutoff of 76, but the inpatient adolescents scored significantly higher. It is understandable that even though the inpatient adolescents were reporting higher levels of depressive symptoms than the high school stu-

dents, they were still below the clinical cutoff. Recall that these adolescents were being treated for a range of psychiatric disorders, and not just affective disorders. Therefore, the higher mean scores likely represent a combination of symptoms being endorsed by those adolescents who were depressed and had higher levels of distress than those exhibiting other disorders. It is also possible that the inpatient adolescents' scores were lower than they would have been at admission (i.e., point of likely greatest symptom severity). However, by administering the protocols to them within 2–3 days of admission, we attempted to minimize the effect of treatment on their self-report. Results of the SIQ comparison were more striking. The clinical sample not only scored significantly higher than the high school sample, but their mean score was above the clinical cutoff, whereas the high school sample's was well below it. This finding suggests that even though some high school students may be having thoughts about suicide, overall they could be characterized as not suicidal. Adolescents in the clinical sample who had been hospitalized following a suicide attempt likely contributed to the much higher levels of ideation in that group. Together, these results support the hypothesis of group mean differences.

Another way to make comparisons between the two groups was at the item level. We were interested in determining which RADS-2 and SIQ items best distinguished between the high school and clinical adolescents. Clinically, comparisons such as this can be useful in guiding interventions by providing information on particular areas of concern. As was true with the comparisons of group means, a larger number of individual items on the SIQ produced large effect sizes compared with RADS-2 items. In fact, no single item on the RADS-2 resulted in a large effect size, suggesting that the symptoms endorsed by the two groups were more similar in the depression domain than in the suicidal ideation domain. Examining the effect sizes of the critical items on the RADS-2 (i.e., 3, 6, 14, 19, 20, and 30) indicated half of them resulted in small effect sizes and half resulted in medium effect

sizes. Items 14, 20, and 30 showed larger differences between the inpatient adolescents and high school controls, but not as large as would have been expected. Two of the SIQ critical items produced medium effect sizes and the rest produced large effect sizes, indicating the two groups differed to a greater extent on the SIQ than the RADS-2. We are not advocating for the use of single items to identify at-risk youth. Rather, we are attempting to highlight the more dramatic differences in psychiatric inpatients' thoughts about suicide, as opposed to depressive symptoms, compared with high school adolescents.

Although it was necessary to establish overall group differences between the two samples, the primary analyses of interest were those comparing the suicidal ideation and the suicide attempter subgroups to the nonclinical controls. Despite attaining moderate accuracy at differentiating between suicide attempters and controls, no single RADS-2 cutoff score achieved the preestablished criteria for optimum sensitivity and specificity. Relying on a single cutoff score to differentiate between adolescents with a history of a suicide attempt and those who had not resulted in a less than acceptable mix of false positives and false negatives. In terms of a school's resource allocation, more youth would be referred for follow-up assessment than would be necessary and a larger than desirable number of adolescents who were indeed at risk would not be evaluated. The RADS-2 did a much better job of distinguishing between the suicidal ideation and control groups. Relying on a cutoff score that is actually lower than recommended in the test manual maximizes the potential for correctly classifying adolescents as at risk (i.e., significant suicidal ideation) or not at risk (i.e., high school controls). Utilizing this cutoff score, 85% of adolescents who are having serious suicidal ideation and 72% of those who are not would be correctly identified. Taken together, cutoff scores on the RADS-2 are more clinically useful for differentiating the responses of youth with suicidal ideation and related controls than those with past suicide attempts and controls. This finding is likely a result of the greater similarity between

the types of thoughts adolescents have about suicide and depressive symptoms than between the myriad factors contributing to a suicide attempt and those same symptoms. However, in school-based screenings where it is assumed that students are functioning well enough to not require hospitalization, youth scoring 67 or higher on the RADS-2 can reasonably be expected to require follow-up assessment without risk of missing large numbers of students in need of further assessment or referring students not at risk.

Examination of a range of SIQ cutoff scores, based on high accuracy discrimination between the two clinical groups and controls, identified very different scores as useful. A score of 19, which is far below the cutoff of 41 suggested in the professional manual, correctly identified 88% of the suicide attempter group and 79% of the controls. A score of 52, however, which is well above the published cutoff score, correctly identified 98% of the suicide ideation group and 74% of the controls.

In evaluating the utility of test scores, a balance is struck between sensitivity and specificity, based on the relative risk of false negatives to false positives (Loong, 2003). Because the goal of the present study is to help schools make informed decisions about the best measures to use in screening students for suicide risk, we need to primarily consider the risk of false negatives (i.e., sensitivity and positive predictive value). False positives are certainly a concern. A false positive on a risk screening means that a student is identified as needing follow-up assessment who is not really at elevated risk. The cost associated with this error is more staff time devoted to the student and potentially unnecessarily worried parents. This is not a trivial effect, as school resources are stretched increasingly thin by demands of providing more and more services for students. It also speaks to the issue of needing an adequate infrastructure in place to handle referrals generated by any screening procedure implemented (Gutierrez et al., 2004). One possible solution is to partner with community agencies to spread the burden more widely.

A false negative would identify a student as not needing follow-up assessment when he or she is actually at elevated risk of suicide. The cost of this error is unnecessary continued suffering on the part of the student and the potential for the youth to attempt suicide and die. Fortunately, this is a case where the low base rate of suicide works in favor of screening efforts. The negative predictive value of any test rises as the prevalence of the disorder to be identified decreases (Loong, 2003). Taking all of the possible acceptable cutoff scores on the RADS-2 and SIQ into consideration for maximum negative predictive value, it appears that scores of 63 on the RADS-2 and 14 on the SIQ will result in the fewest false negatives, which is the most important psychometric consideration when evaluating a screening instrument. Until the results of this study have been replicated in a broader range of school settings, it is premature to recommend schools use only the SIQ as their suicide risk screen, but it appears warranted to use it as a central component of the process.

To begin testing the hypothesis that the combination of mean scores and critical items will have the greatest utility in correctly identifying at-risk youth, we next turned our attention to the sensitivity and specificity of critical items on the RADS-2. None of the RADS-2 critical items was identified as potentially useful for differentiating the responses of youths with suicidal ideation or attempts and the high school nonreferred subgroup. For screening purposes, the mean total score had greater utility for correctly identifying at-risk youth than the critical items. Examination of adolescents' responses on these items may provide useful clinical information regarding factors affecting their depressive symptoms, but they should not be relied on for screening purposes.

As with the analyses examining the utility of the mean scores, the SIQ critical items performed better than those of the RADS-2. Two items attained acceptable sensitivity and specificity for discriminating between those in the suicidal ideation group and the high school controls. All but one item was found to be acceptable in discriminating between the suicide

attemper and high school control groups. For screening purposes, Items 2, 3, and 4 would result in between 90% and 96% of adolescents not at risk correctly being identified as such. These findings lead to the conclusion that an adolescent with a total mean SIQ score of 14 or higher who also scores high on those 3 items should be assessed more thoroughly for suicide risk, as this profile is unlikely in a youth who is not suicidal. The inability of RADS-2 critical items to discriminate between either of the clinical groups and the controls means this hypothesis was only partially supported.

The purpose of this study was not to provide a cost-benefit analysis comparing self-report measures versus one of the established programs available to schools for screening adolescents. TeenScreen and Signs of Suicide are more than just screening programs; they are also designed to target factors contributing to adolescent suicide risk, with the goal of affecting change in those factors, and by extension reducing risk. Each school must decide the scope of program they can reasonably provide for their students, taking into consideration the costs of the programs and evidence for their effectiveness. The most basic level involves school-wide screenings to identify at-risk youth, followed by more comprehensive assessments for identified youth, and finally providing services to the youth most in need (Reynolds, 1991). Our findings suggest that a school with limited resources desiring to provide suicide screenings for their students in a cost-effective manner could rely on administration of just the SIQ. Although the RADS-2 provides useful information regarding suicide risk, and there is an association between scores on the two measures, it does not perform nearly as well as the SIQ for identifying at-risk youth. Some useful information will be lost by only screening for one component of suicide risk, so schools could consider using the RADS-2 as part of the more comprehensive assessment conducted of the students screening positive in the first phase.

We offer these recommendations with several caveats. All of the groups for this study were drawn from settings in the Mid-

west. In addition, the high school sample came from an urban setting whereas the inpatient samples came from a more rural area. As a result, there were differences in the ethnic diversity of the groups. The high school sample was fairly diverse, whereas the inpatient sample was not. There are differences in suicide risk factors (e.g., Canino & Roberts, 2001; Harris & Molock, 2000; Hovey & King, 1997; Shiang, 1998) and rates of suicidal ideation and attempts across racial and ethnic groups (e.g., Canino & Roberts, 2001; Garrouette et al., 2003; Roberts, 2000; Shiang, 1998). Our data do not allow for more fine-grained analyses examining the sensitivity, specificity, and associated positive predictive and negative predictive value for the RADS-2 and SIQ across ethnic groups. It is possible that the scales would perform differently for one or more groups. Replication of these findings in samples in which such comparisons are possible would therefore increase the general applicability of our findings. Also, the inpatient sample did not represent all types of serious psychopathology (e.g., bipolar disorder) and comparisons were not made across diagnostic categories. However, the intention of this study was not to determine whether the RADS-2 and SIQ can differentiate between adolescents with various disorders, but rather between those at different levels of suicide risk. The ultimate test of the utility of an assessment tool is how well it predicts future outcomes, which requires prospective research designs. At present, such data are lacking for existing screening measures (Goldston, 2003).

Conclusion

A good return on a school's screening investment can be derived from using the SIQ, and to a lesser extent the RADS-2, as the first phase of a tiered screening approach (Reynolds, 1991). For schools with limited financial and personnel resources, this option might be more cost-effective than using one of the more comprehensive screening programs available (e.g., Columbia TeenScreen; Signs of Suicide). The evidence for the predictive validity of these two scales adds significantly to their

value, and should assist schools in making informed decisions about resource allocation.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Aseltine, R. H., & DeMartino, R. (2004). An outcome evaluation of the SOS suicide prevention program. *American Journal of Public Health, 94*, 446–451.
- Canino, G., & Roberts, R. E. (2001). Suicidal behavior among Latino youth. *Suicide and Life-Threatening Behavior, 31*, 122–131.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colle, L., Belair, J., DiFeo, M., Weiss, J., & LaRoche, C. (1994). Extended open-label fluoxetine treatment of adolescents with major depression. *Journal of Child and Adolescent Psychopharmacology, 4*, 225–232.
- Columbia University TeenScreen Program. (2007). *Training manual for school-based screenings*. New York: Author.
- Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., et al. (2008). Youth risk behavior surveillance—United States, 2007. *Morbidity and Mortality Weekly Reports, 57*(SS-4), 1–136.
- Erdman, H. P., Greist, J. H., Gustafson, D. H., Taves, J. E., & Klein, M. H. (1987). Suicide risk prediction by computer interview: A prospective study. *Journal of Clinical Psychiatry, 48*, 464–467.
- Garland, A., Shaffer, D., & Whittle, B. (1989). A national survey of school-based, adolescent suicide prevention programs. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 931–934.
- Garroute, E. M., Goldberg, J., Beals, J., Herrell, R., Manson, S. M., & the AI-SUPERPFP team. (2003). Spirituality and attempted suicide among American Indians. *Social Science & Medicine, 56*, 1571–1579.
- Goldston, D. B. (2003). *Measuring suicidal behavior and risk in children and adolescents*. Washington, DC: American Psychological Association.
- Gutierrez, P. M., & Osman, A. (2008). *Adolescent suicide: An integrated approach to assessment of risk and protective factors*. DeKalb, IL: Northern Illinois University Press.
- Gutierrez, P. M., Watkins, R., & Collura, D. (2004). Suicide risk screening in an urban high school. *Suicide and Life-Threatening Behavior, 34*, 421–428.
- Harris, T. L., & Molock, S. D. (2000). Cultural orientation, family cohesion and family support in suicide ideation and depression among African American college students. *Suicide and Life-Threatening Behavior, 30*, 341–353.
- Hovey, J. D., & King, C. A. (1997). Suicidality among acculturating Mexican Americans: Current knowledge and directions for research. *Suicide and Life-Threatening Behavior, 27*, 92–103.
- Kaplan, M. L., Asnis, G. M., Sanderson, W. C., Keswani, L., DeLecuona, J. M., & Joseph, S. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychology, 50*, 294–298.
- Kendall, P. C., Cantwell, D. P., & Kazdin, A. E. (1989). Depression in children and adolescents: Assessment issues and recommendations. *Cognitive Therapy and Research, 13*, 109–146.
- Kung, H.-S., Hoyert, D. L., Xu, J., & Murphy, S. L. (2008). Deaths: Final data for 2005. *National Vital Statistics Reports, 56*(10). Retrieved January 16, 2008, from http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56_10.pdf
- Lonigan, C. J., Anthony, J. L., & Shannon, M. P. (1998). Diagnostic efficiency of posttraumatic symptoms in children exposed to disaster. *Journal of Clinical Child Psychology, 27*, 255–267.
- Loong, T.-W. (2003). Understanding sensitivity and specificity with the right side of the brain. *British Medical Journal, 327*, 716–719.
- Matthey, S., & Petrovski, P. (2002). The Children's Depression Inventory: Error in cutoff scores for screening purposes. *Psychological Assessment, 14*, 146–149.
- Mazza, J. J. (2000). The relationship between posttraumatic stress symptomatology and suicidal behavior in school-based adolescents. *Suicide and Life-Threatening Behavior, 30*, 91–103.
- Miller, D. N., & DuPaul, G. J. (1996). School-based prevention of adolescent suicide: Issues, obstacles, and recommendations for practice. *Journal of Emotional and Behavioral Disorders, 4*, 221–230.
- Perez, V. W. (2005). The relationship between seriously considering, planning, and attempting suicide in the Youth Risk Behavior Survey. *Suicide and Life-Threatening Behavior, 35*, 35–49.
- Pinto, A., Whisman, M. A., & McCoy, K. J. M. (1997). Suicidal ideation in adolescents: Psychometric properties of the Suicidal Ideation Questionnaire in a clinical sample. *Psychological Assessment, 9*, 63–66.
- Reinecke, M. A., Du Bois, D. L., & Schultz, T. M. (2001). Social problem solving, mood, and suicidality among inpatient adolescents. *Cognitive Therapy and Research, 25*, 743–756.
- Reynolds, W. M. (1987). *Reynolds Adolescent Depression Scale*. Odessa, FL: Psychological Assessment Resources, Inc.
- Reynolds, W. M. (1988). *Suicidal Ideation Questionnaire: Professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Reynolds, W. M. (1991). A school-based procedure for the identification of adolescents at risk for suicidal behaviors. *Family Community Health, 14*, 64–75.
- Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale—2nd Edition: Professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Roberts, R. E. (2000). Depression and suicidal behaviors among adolescents: The role of ethnicity. In I. Cuéllar & F. A. Paniagua (Eds.), *Handbook of multicultural mental health: Assessment and treatment of diverse populations* (pp. 359–388). San Diego: Academic Press.
- Shaffer, D., Garland, A., Vieland, V., Underwood, M., & Busner, C. (1991). The impact of curriculum-based suicide prevention programs for teenagers. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 588–596.

- Shaffer, D., Scott, M., Wilcox, C., Maslow, C., Hicks, R., Lucas, C. P., et al. (2004). The Columbia Suicide-screen: Validity and reliability of a screen for youth suicide and depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*, 71–79.
- Shiang, J. (1998). Does culture make a difference? Racial/ethnic patterns of completed suicide in San Francisco, CA 1987–1996 and clinical applications. *Suicide and Life-Threatening Behavior*, *28*, 338–354.
- Spirito, A., Stark, L., Fristad, M., Hart, K., & Owens-Stively, J. (1987). Adolescent suicide attempters hospitalized on a pediatric unit. *Journal of Pediatric Psychology*, *12*, 171–190.
- Substance Abuse and Mental Health Services Administration. (2006). *Intervention summary: SOS Signs of Suicide*. Retrieved June 16, 2008, from http://www.nrepp.samhsa.gov/programfulldetails.asp?PROGRAM_ID=66
- Substance Abuse and Mental Health Services Administration. (2007). *Intervention summary: TeenScreen*. Retrieved June 16, 2008, from http://www.nrepp.samhsa.gov/programfulldetails.asp?PROGRAM_ID=108
- Date Received: June 18, 2008
Date Accepted: January 27, 2009
Action Editor: Tanya Eckert ■

Peter M. Gutierrez, PhD, is Clinical/Research Psychologist with the VA VISN 19 Mental Illness Research, Education and Clinical Center. He studied adolescent and young adult suicide for over 10 years before joining the VA. He is a Past President of the American Association of Suicidology.

Augustine Osman, PhD, is Professor of Clinical Psychology and Associate Dean of Research and Graduate Studies at The University of Texas at San Antonio. He is an expert in scale development and psychometrics, having developed or codeveloped numerous measures of adolescent suicide-related constructs.