Rapid #: -10158897

MULTIPLE-STAGE SCREENING OF YOUTH DEPRESSION IN SCHOOLS

MELISSA E. MOREY

PRO-ED, Inc.

PRERNA ARORA

University of Maryland

KEVIN D. STARK

The University of Texas at Austin & Texas Child Study Center

Schools present a unique environment in which to conduct universal screenings for youth depression. The present study examines the efficiency of a multiple-stage assessment procedure assessing youth depression in the schools by calculating hit rates and establishing diagnostic accuracy for the measures used. Girls ($N = 3318$) aged 8 to 13, participated in a three-stage screening conducted in schools. At Stage 1, participants completed self-report measures. At Stage 2, participants completed either a second self-report measure or a brief symptom interview. At Stage 3, diagnostic interviews were conducted. Hit rates were calculated at Stages 1 and 2 to compare the efficiency of the different assessments. The diagnostic accuracy of the screening measures was investigated through the calculation of sensitivity and specificity indices, logistic regression, and receiver operating characteristic curve analysis. Bias in identification rates across ethnicities was also examined. Comparison of hit rates indicated that the brief symptom interview had greater efficiency in the identification of depression at Stage 2; moreover, the measures demonstrated accuracy in their prediction of depression. Results also indicated a lack of bias in screening measures, as they did not over-identify depression in different ethnic or racial groups. Recommendations for implementation of multiple-stage screenings in schools are discussed. © 2015 Wiley Periodicals, Inc.

Depression in youth is a serious mental health concern that has been linked to poor school attendance, academic underachievement, and school failure (Garber & Horowitz, 2002; Hammen & Rudolph, 2003), and depressed youth have an increased risk of future school dropout, substance abuse, and suicide (Waslick, Kandel, & Kakouros, 2002). Moreover, if untreated, depression is chronic and recurrent (Keller, 2003; Simons, Rohde, Kennard, & Robins, 2005). Starting in adolescence, girls suffer from depression at a rate of 2:1 compared with boys (Galambos, Leadbeater, & Barker, 2004); this disparity persists, and even increases, during adulthood (Hankin & Abramson, 2001). Further, girls present with different symptoms, often to a more severe degree (Stark, Sander, Yancy, Bronik, & Hoke, 2000). Additionally, evidence has indicated that early-onset depression is predictive of an increasingly severe course (Dekker et al., 2007), which highlights the importance of early detection. Fortunately, a significant research base regarding the use of cognitive–behavioral treatment (CBT) for youth depression exists (Curry, 2001; Weersing & Weisz, 2002), including the use of CBT in schools (Shirk, Kaplinski, & Gudmudsen, 2009).

ASSESSMENT OF DEPRESSION

Assessment of youth depression assists in screening, case formulation, treatment planning, monitoring of progress, and treatment evaluation (Mash & Hunsley, 2005). Further, the role of diagnosis has been underscored as mental health organizations emphasize diagnosis-specific evidence-based treatments (Chorpita & Donkervoet, 2005). There are numerous methods with which the symptoms, severity, and functional impairment of depression can be assessed, the most common of which include self-report measures, parent–teacher questionnaires, diagnostic interviews, observational

Correspondence to: Melissa E. Morey, PRO-ED, Inc., 8700 Shoal Creek Boulevard, Austin, TX 78757. E-mail: mmorey@proedinc.com

methods, and projective techniques. Combining varied methods of assessment from diverse individuals, who have the opportunity to observe the child in different settings, has been proposed as a way to accurately and efficiently identify depressed youth (Chrisman, Egger, Compton, Curry, & Goldston, 2006; Timbremont, Braet, & Dreessen, 2004).

SCHOOL-BASED ASSESSMENT

Schools present a unique environment for identifying and assessing youth experiencing depressive disorders because they are natural entry points for addressing mental health needs due to the significant role these institutions play in students' lives (Mennuti, Christner, & Freeman, 2006). Specifically, mandatory attendance at school provides a captive audience, and families are familiar with the school environment, which can facilitate the assessment process and delivery of additional mental health services. Thus, support for inclusion of mental health care, especially screening practices, in schools has been posited by the Report on the Surgeon General's Conference on Children's Mental Health (U.S. Public Health Service, 2000) and the President's New Freedom Commission on Mental Health (2003).

School personnel have difficulty recognizing when a student is depressed (Reynolds, 1986). Consequently, if a child is not identified, then the traditional model of referring youth for assessment will not be successful. Assessment directed at referred individuals has been deemed insufficient for reasons related to cost, time, and treatment planning (Dowdy, Ritchey, & Kamphaus, 2010). School-based universal screenings, including multiple-stage screenings in particular, for the identification of mental health concerns has been recommended (Dowdy et al., 2010; Levitt, Saka, Romanelli, & Hoagwood, 2007). This holds especially true for internalizing disorders, as these disorders remain more challenging to identify and are often under-reported (Weist, Rubin, Moore, Adelsheim, & Wrobel, 2007). Nonetheless, such universal assessment practices remain uncommon and have not been empirically evaluated. Despite significant support for mental health assessment in schools, Weist and colleagues (2007) note that it violates the family's privacy.

MULTIPLE-STAGE ASSESSMENT

For large-scale depression screenings in schools, a three-step "multiple-stage" procedure has been recommended (Dowdy et al., 2010; Kendall, Cantwell, & Kazdin, 1989; Reynolds, 1986). Children are initially screened (Stage 1) using a self-report measure such as the Children's Depression Inventory (CDI; Kovacs, 1992) to identify students who may be experiencing a clinically significant level of depression. Students scoring above the cutoff criterion (a predetermined threshold) are rescreened using the same measure (Stage 2) a short time later (e.g., 3 weeks; Reynolds, 1986). This second assessment has been recommended due to the tendency of children to endorse greater severity of depressive symptoms on the first administration of a self-report measure relative to the second administration of the same measure (Finch & Saylor, 1985; Seligman et al., 1984). Students who again exceed the cutoff complete a diagnostic interview, such as the Schedule for Affective Disorders and Schizophrenia for School Age Children (K-SADS-P-IVR; Ambrosini & Dixon, 2000) during Stage 3.

Multiple-stage screening procedures are believed to be time-efficient and cost-effective (Laurent, Hadler, & Stark, 1994; Reynolds, 1986), as they are thought to reduce the number of "false positives" (i.e., participants identified by a screening measure as depressed, but not identified by a diagnostic interview as depressed) prior to engaging in a time-intensive, lengthy diagnostic interview, the gold standard for diagnosis of depression (Reynolds, 1986). Further, multiple-stage assessment procedures are believed to assist in differentiating transient mood symptoms from persistent depressive symptoms requiring intervention (Reynolds, 1986). Conversely, multiple-stage

assessment procedures have been criticized for neither effectively nor efficiently reducing the number of children screened at different stages (Adelman & Taylor, 2006).

Efficiency of Multiple-Stage Assessment

Unfortunately, multiple-stage screening procedures have not been widely reported nor evaluated in the literature. Whereas some have described their use to assess depression (e.g., Kaslow, Stark, Printz, Livingston, & Tsa, 1992; Reynolds, 1986), few have gone beyond description to investigate and compare different assessment methods to utilize in the procedure. Additionally, although the literature recommends the use of self-report measures in both Stages 1 and 2 of the procedure, the efficiency of self-report versus an alternative method, for instance a brief interview, has not been assessed with depressed youth. To our knowledge there are only two published reports documenting use of the procedure to identify depressed youth. Kaslow, Stark, Printz, Livingston, and Tsai (1992) utilized the procedure to identify depressed and/or anxious 8- to 13-year-old children, whereas Reynolds (1986) summarized findings from unpublished research in the identification of depression in adolescents. Neither of these investigations evaluated possible alternative methods of assessment in Stage 2, nor did they evaluate potential racial or cultural bias, which may affect the results of a screening process (Laurent et al., 1994).

To evaluate the efficiency of the multiple-stage procedure, an examination of hit rates (i.e., ratios of participants identified as depressed in the final stage to the number of participants identified in previous stages) of the measures used is needed. Although Reynolds (1986) and Kaslow et al. (1992) used the multiple-stage procedure, hit rates were not directly reported in either study. However, they can be estimated with the numbers provided (i.e., by dividing the number of participants identified with a disorder in Stage 3 by the number of participants who met the cutoff criterion in Stages 1 and 2). Based on information given in the Reynolds study, the first administration of the self-report measure resulted in hit rates between 44% and 60%; hit rates for the second administration of the measure were between 62% and 80%. In the Kaslow et al. (1992) study, the first administration of the self-report measure (the CDI) was determined to have a hit rate of 24%, and the second administration of the CDI was determined to have a hit rate of 41%. The difference in hit rates may be due in part to the age of the participants; Reynolds (1986) summarized studies with adolescents, and the study by Kaslow et al. (1992) assessed children.

In settings like schools, where time and financial resources are limited, improving efficiency in large-scale screenings is a crucial endeavor. Therefore, further examination of the screening measures used in the multiple-stage procedure, especially the investigation of alternative assessments (via the calculation of hit rates) is needed. In regard to the screening measures, the use of diagnostically accurate measures would help make the screening process more efficient. Diagnostic accuracy refers to whether a measure can distinguish individuals with a disorder from those without the disorder. Participants are administered both the measure being studied and the gold standard assessment for that disorder to determine whether they actually have a diagnosis. Once diagnoses have been determined, the diagnostic accuracy of the measure being studied can be investigated. Methods of establishing diagnostic accuracy include computing sensitivity and specificity indices, conducting logistic regression analyses, and performing receiver operating characteristic (ROC)/area under curve analyses (AUC).

The Current Study

In sum, the efficiency of the multiple-stage procedure requires further investigation, and the use of an alternative assessment method at Stage 2 needs to be evaluated. Specifically, with most researchers recommending the repeated use of a self-report measure at both Stages 1 and 2 (Finch

& Saylor, 1985; Reynolds, 1986; Seligman et al., 1984), it is unknown whether the use of a different measure could more efficiently assess youth with depression. To compare different measures, in the present study. we calculated hit rates for the measures at Stages 1 and 2. Using the measure with the higher hit rate should reduce the workload of examiners at Stage 3 and, thus, make the multiple-stage procedure more efficient. Further, multiple-stage procedures have not been investigated to assess whether these procedures are equally efficient and unbiased across diverse populations, which is necessary for use of the procedure (Laurent et al., 1994).

As such, one goal of the present study was to assess efficiency of the multiple-stage assessment procedure by comparing hit rates of screening measures used in a multiple-stage procedure for identification of depressive disorders in 8- to 13-year-old girls to hit rates calculated in previous studies. An additional goal is to compare hit rates of two different measures in Stage 2 of the present study to determine whether one method was more efficient. Moreover, the present study will evaluate the diagnostic accuracy of the two screening measures. First, the multiple-stage assessment procedure used in a school-based CBT treatment for female youth (Stark, Arora, Krumholz, Morey, & Jensen-Doss, 2015 is described. Then, hit rates for the methods used at Stages 1 and 2 are examined. It was hypothesized that the brief symptom interview would have a higher hit rate than the self-report measure, and the potential benefit of using a brief symptom interview at Stage 2, rather than a second self-report measure as proposed by others (Finch & Saylor, 1985; Seligman et al., 1984) is explored.

The present study also sought to evaluate the diagnostic accuracy of the CDI and the Beck Youth Inventory-Depression (BYI-D; Beck, Beck, & Jolly, 2001) through the calculation of sensitivity and specificity indices, logistic regression, and ROC curve analysis. Due to the goals of the larger treatment study, which was to maximize the identification of girls who were depressed, it is hypothesized that the screening measures will possess high sensitivity and the ROC/AUC values for both measures will be in the acceptable range. Finally, possible bias against racial and ethnic groups of the screening measures is investigated. An exploratory hypothesis is that the identification rates will be similar across ethnic and racial groups.

METHOD

Participants

Participants were recruited through multiple-stage screenings for a school-based treatment trial (Stark et al., 2015) for female youth with depressive disorders. Letters, written in both English and Spanish, requesting parental consent for participation in a screening procedure were sent home to primary caregivers of girls attending public schools in two school districts in the southwestern United States. In addition, 45 girls, aged 8 to 13 years, volunteered to participate as the non-depressed control group and were assessed with the K-SADS-P-IVR and determined not to have a depressive disorder. The overall sample comprised 3363 girls (Screening $n = 3318$; Control $n = 45$) between the ages of 8 and 13 ($M = 10.7$, $SD = 1.2$) enrolled in Grades 4 through 7 (22% in fourth grade, 25% in fifth grade, 30% in sixth grade, and 23% in seventh grade). Self-reported ethnicity and race was 35% White (non-Hispanic), 28% White (Hispanic), 11% African American, 3% Asian/Pacific Islander, 1% American Indian, and 6% multi-racial. Fifteen percent did not report this information.

Procedure

Training of Measures Administrators. Doctoral-level graduate research assistants (GRAs) were trained to administer and score the measures and to conduct a brief symptom interview; the Diagnostic and Statistical Manual Brief Symptom Interview for Depression (DSM-BSI; Stark &

Sander, 2002). At least one GRA present at the screenings had prior training on the assessment of suicidal ideation and intent.

Training of the Interviewers.   The K-SADS-P-IVR interviews were conducted by doctoral-level graduate students who completed 6 months of training that included didactic presentations over each diagnostic section, observing interviews, and conducting mock interviews with non-depressed volunteers. Interviewers listened to six audio-taped interviews and accurately determined the absence, presence, and severity of psychological disorders assessed by the interviewer. Before conducting interviews independently, each interviewer in training had to demonstrate competence by providing reliable symptom ratings. Interviewers also had to demonstrate inter-rater reliability with an experienced interviewer (kappa $\geq$ .90) in their first three interviews. All interviewers participated in weekly group supervision to ensure reliability of diagnoses.

Multiple-Stage Screening of Participants.   The multiple-stage screening procedure was conducted in two selected school districts to identify depressed girls for a depression treatment trial. By the end of the 5-year trial, 20 schools had participated. Screenings were conducted at the beginning of the semester in three to six schools yearly. The screenings were rotated so that different schools participated in spring than in the fall; this reduced the workload of the school staff and reduced the likelihood of girls undergoing repeated screenings in close succession. Of the total girls screened, 16% were screened twice, and 1% was screened three times (average time between repeated screenings was 1 year). The procedure was slightly different for the two school districts. During the second year of the study, a new measure of depression in youth, the BYI-D (Beck et al., 2001), was published, and it was included in the present study to evaluate its psychometric properties (Stapleton, Sander, & Stark, 2007). At Stage 1, all of the participants completed the CDI; however, 1958 participants completed the BYI-D in addition to the CDI.

Girls in Grades 4 through 7 ($N = 7737$) were invited to participate in the study. During Stage 1, girls who received parental consent and provided assent ($N = 3318$; 43% of population) completed the CDI and/or the BYI-D. Prior to the screening, school staff were consulted to identify students who could have difficulty understanding the measures due to either cognitive impairments or limited English language proficiency. Trained GRAs were present during the screening to assist the girls as they completed measures and to ensure that participants understood the measures. Bilingual GRAs administered and translated all measures to Spanish-speaking students, either one-on-one or in small groups of two to three students. Students who were unable to complete measures accurately due to comprehension difficulties were excluded from the study.

Participants who scored at least 16 on the CDI ($n = 831$) and/or at least 25 on the BYI-D ($n = 349$) proceeded to Stage 2 ($n = 859$ in total). During Stage 2, girls in Cohort 1 ($n = 162$) completed a second administration of the CDI 1 week later (as recommended by Kendall et al., 1989), whereas girls in Cohorts 2–7 ($n = 697$) completed the DSM-BSI. After the first year of the study, the additional administration of the CDI was replaced with the DSM-BSI at Stage 2 because the second administration of the CDI was not as efficient as previously thought, as it over-identified girls as depressed. The DSM-BSI was administered by a GRA on the same day as the CDI. The data from Cohort 1 were retained in the present study to compare the hit rates of the two methods.

Participants from Cohort 1 who scored 16 or greater on their second CDI ($n = 122$) and participants from Cohorts 2 through 7 who reported at least three depressive symptoms, including a possible mood disturbance, on the DSM-BSI ($n = 334$), were invited, along with their caregivers, to complete the K-SADS-P-IVR in Stage 3. Interviews with the girls were conducted at school and interviews with parents were conducted at school or over the phone. Of the 378 girls who completed

the K-SADS-P-IVR, 225 (60%) received a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev., American Psychological Association, 2000) diagnosis of a depressive disorder.

Instrumentation

Children's Depression Inventory. The CDI (Kovacs, 1992) is the most commonly used self-report measure of depression in children and adolescents and is designed for ages 7 to 17. The measure consists of 27 items assessing the presence and severity of cognitive, affective, and behavioral symptoms of depression over the preceding 2 weeks. For screening purposes, a cutoff score of 16 or above has been shown to have the highest predictive value (Trimbremont, Braet, & Dreessen, 2004). The CDI demonstrates good internal consistency (Kovacs, 1992), whereas test–retest reliability has ranged from .38 to .87 (Kovacs, 1992). The CDI possesses good predictive validity, and it successfully discriminated depressive disorders from anxiety and disruptive disorders (Timbremont et al., 2004). Although the CDI cannot be used by itself for a psychiatric diagnosis (Sitarenios & Kovacs, 1999), there is strong support for the use of the CDI as a screening tool (Timbremont et al., 2004). Internal consistency reliability for the present sample was good ($\alpha = .90$).

Beck Youth Inventory-Depression. The BYI-D (Beck et al., 2001) assesses the presence and severity of depressive symptoms in youth aged 7 to 14 years. The measure includes 20 items assessing negative thoughts, feelings of sadness, and physiological symptoms of depression. According to the authors, a cutoff score of 25 or above indicates moderately elevated symptoms of depression (Beck et al., 2001).

Internal consistency of the BYI-D has been found to be high, with coefficient alphas of .91 for girls aged 7 to 10, .90 for boys aged 7 to 10, .91 for girls aged 11 to 14, and .92 for boys aged 11 to 14. Test–retest reliabilities ranged from .79 to .92 over a retest interval of 7 days (Beck et al., 2001). The BYI-D has also been found to have high convergent validity with the CDI total score ($r = .82$). Internal consistency reliability for the BYI-D in present sample was excellent ($\alpha = .93$).

Diagnostic and Statistical Manual Brief Symptom Interview for Depression (DSM-BSI; Stark & Sander, 2002). This semi-structured interview was created for the purpose of screening and monitoring depressive symptoms. It contains 18 items that are rated as possibly present or absent. If the participant indicates that a symptom might be present (e.g., it is a problem for most days within the past 2 weeks, and is distressing or clinically impairing), the interviewer rates it as present and goes to the next item. The administration time is 10 to 20 minutes, and parents do not complete this interview. The DSM-BSI was significantly correlated with the CDI ($r = .46$, $p < .001$) and the BYI-D ($r = .45$, $p < .001$), supporting the convergent validity of the measure. Internal consistency reliability for the DSM-BSI in present sample was good ($\alpha = .86$).

The Schedule for Affective Disorders and Schizophrenia for School Age Children. The K-SADS-P-IVR (Ambrosini & Dixon, 2000) is a semi-structured diagnostic interview designed for use with both a primary caregiver and child, and it is the most widely used semi-structured diagnostic interview for children and adolescents (Klein, Dougherty, & Olino, 2005). The K-SADS-P-IVR contains five diagnostic supplements, including affective disorders, anxiety disorders, behavioral disorders, psychotic disorders, and substance abuse, eating, and tic disorders. Most items are rated on a 6- or 7-point range of severity, with specific criteria indicated for scoring level. In the present study, the child and parent interviews took approximately 45 and 60 minutes, respectively, to administer.

Earlier versions of the K-SADS-P-IVR have been found to have high test–retest reliability for the diagnosis of depressive disorders (kappa = .90s; Apter, Orvaschel, Laseg, Moses, & Tyano, 1989; Kaufman et al., 1997) and for the symptom scales (intra-class correlation coefficients =

Table 1
Number of Participants Assessed at Each Stage of Screening and Hit Rates

Stage	Cohort	Measure	Total Assessed	Met Cutoff Criterion for Next Stage/Diagnosed with Depressive Disorder in Stage 3	% of Total Assessed	Hit Rate, %
1	1	CDI	514	162	32	24
	2–7	CDI (BYI-D)	2804 (1958)	669 (349)	24 (18)	
	2–7	BYI-D and/or CDI	2804	697	25	28
	1–7	Total	3,318	859	26	26
2	1	2nd CDI	162	122	75	47
	2–7	DMS-BSI	697	334	48	63
	1–7	Total	859	456	53	
3	1	KSADS	83	39	47	
	2–7	KSADS	295	186	63	
	1–7	Total	378	225	60	

Note. In Stage 1 for Cohorts 2–7, all participants ($n = 2804$) were administered the CDI, and a subset of participants ($n = 1958$) were administered both the CDI and the BYI-D. Values for the BYI-D are in parentheses. In total, 697 participants in Cohorts 2–7 met the cutoff criterion on the CDI and/or the BYI-D at Stage 1.

.72–.83; Apter et al., 1989). Inter-rater reliability was found to be high for the diagnoses of major depression, and dysthymic disorder (Ambrosini, 2000). Overall, evidence of high diagnostic, scale, and symptom reliability support the K-SADS as a reliable diagnostic tool for use with children (Ambrosini, Metz, Prabucki, & Lee, 1989). Internal reliability for the total depression score used in this study was excellent ($\alpha = .91$).

RESULTS

Results are presented by stage, followed by results of diagnostic accuracy analyses and examination of possible racial or ethnic bias of the screening measures. Numbers of participants in each stage are presented in Table 1.

Stage 1: Screening

The CDI scores for the 3318 girls who participated at Stage 1 ranged from 0 to 47 ($M = 9.8$, $SD = 8.3$). A subset of girls ($n = 1958$) also completed the BYI-D at Stage 1, and the scores ranged from 0 to 57 ($M = 14.9$, $SD = 9.8$). Twenty-six percent of the sample met the cutoff criterion at Stage 1. A significant difference was found in the CDI scores between girls identified as depressed ($M = 24.5$, $SD = 7.2$) and not depressed ($M = 19.8$, $SD = 7.2$), $t(376) = 6.33$, $p < .001$. A significant difference was also found in the BYI-D scores between girls identified as depressed ($M = 30.1$, $SD = 9.4$) and not depressed ($M = 19.7$, $SD = 10.5$), $t(318) = 9.39$, $p < .001$.

Stage 2: Identification

In Cohort 1, 75% of participants identified in Stage 1 continued to meet the cutoff criterion for the CDI at the second stage of screening. In Cohorts 2 through 7, 48% of those identified in Stage 1 reported possible depressive symptoms on the DSM-BSI in Stage 2.

Stage 3: Diagnosis

Girls who were identified as possibly depressed in Stage 2 were interviewed after parental consent was obtained. Of the 456 girls identified, 378 completed interviews at Stage 3. Parental permission was not obtained (either parents did not respond to attempts to contact them or they would not grant permission) for 78 girls. In Cohort 1, 47% of girls received a diagnosis of a depressive disorder. In Cohorts 2 through 7, when the DSM-BSI was used, 63% of girls received a diagnosis of a depressive disorder. Based on results of the K-SADS-P-IVR, 9% of the fourth graders, 7% of the fifth graders, 5% of the sixth graders, and 7% of the seventh graders screened received diagnoses of a depressive disorder.

Hit Rates

Hit rates of measures at Stages 1 and 2 (displayed separately for Cohort 1 and Cohorts 2 through 7) are presented in Table 1. Hit rates were calculated by dividing the number of girls who met cutoff criterion on a screening measure for the next stage (and received parental permission to participate) by the number diagnosed as depressed by the K-SADS-P-IVR. To determine whether hit rates were significantly different between the CDI and DSM-BSI at Stage 2, a chi-square test of goodness-of-fit was performed. Hit rates were significantly higher for the DSM-BSI compared with the CDI, $\chi^2(1, N = 378) = 6.9, p = .01$.

Diagnostic Accuracy

The sensitivity indices and specificity indices of the predictor measures (i.e., CDI and BYI-D) were calculated and examined. Positive predictive values can also be derived when examining diagnostic accuracy; however, they can be greatly influenced by the prevalence rate of the disorder in the particular sample being studied. In a sample with a high prevalence rate (such as the one used in the present study), the positive predictive value of a measure will be higher than if the same measure was used in a sample with a lower prevalence rate. Thus, although they may provide information about the clinical importance of a test, they are not indices of diagnostic accuracy (Pepe, 2003; Swets, 1996) and will not be calculated for the present study. The sensitivity index reflects the ability of a measure to identify depressed individuals correctly, which is the most important attribute of a screening test, particularly when assessing serious conditions such as depression. The specificity index reflects the ability of a measure to identify non-depressed individuals correctly. The ROC is the standard method used to evaluate accuracy, and the AUC is the index of accuracy (Pepe, 2003; Swets, 1996). A ROC graph represents the proportion of times that a student was correctly identified as depressed by the screening measure relative to the proportion of times that a student was incorrectly identified as depressed across the range of all possible screener scores. ROC/AUC values range from 0 (representing no predictive ability) to 1 (representing perfect predictive ability). Diagnostic accuracy of the screening measures for depression (e.g., the CDI and BYI-D) is crucial and requires further investigation.

To examine the predictor measures' (i.e., CDI and BYI-D) diagnostic accuracy, a binary classification analysis was conducted using a group of 423 participants (i.e., 378 participants assessed at Stage 3 and 45 participants from the control group). Girls were dichotomized into four groups based on their depression status as determined by the KSADS-P-IVR and their scores on the predictor measures, and multiple 2×2 frequency matrices were created for the BYI-D and CDI. In the matrices, the number of participants correctly identified by the measure is represented by cells *a* and *d*. Cell *a* represents true positives (i.e., identified by the predictor measure and K-SADS-P-IVR as depressed), and cell *d* represents true negatives (i.e., identified by the predictor measure and the K-SADS-P-IVR as not depressed). The number of participants who were not correctly identified

Table 2
Diagnostic Accuracy and ROC/AUC Curve Analyses for the BYI-D and CDI

Criterion	Cut Score	Sensitivity Index	Specificity Index	ROC/ AUC	Classification Accuracy	True Positives	False Positives	True Negatives	False Negatives
BYI-D	25	.72	.69	.77	.70	118	49	107	46
	26	.66	.73		.69	108	42	114	56
	27	.62	.76		.69	102	37	119	62
	28	.55	.78		.67	91	34	122	73
	29	.53	.82		.67	87	28	128	77
	30	.49	.85		.67	81	24	132	83
CDI	16	.95	.22	.70	.66	214	119	34	11
	17	.90	.27		.65	202	111	42	23
	18	.84	.37		.65	190	96	57	35
	19	.81	.45		.66	182	84	69	43
	20	.72	.57		.66	161	66	87	64
	21	.66	.61		.65	149	57	96	76
	22	.62	.70		.65	139	46	107	86

is represented by cells *b* and *c*. Cell *b* represents false positives (i.e., identified by the predictor measure as depressed, but not identified by the K-SADS-P-IVR as depressed), and cell *c* represents false negatives (i.e., not identified on the predictor measure as depressed, but identified by the K-SADS-P-IVR as depressed).

Table 2 reports results of diagnostic accuracy analyses. Authorities vary regarding how large a measure's sensitivity and specificity indices should be to justify its use for screening purposes. Jansky (1978) recommends that the sensitivity and specificity indices reach .70; however, others are willing to tolerate high false-positive rates for screening purposes (Wood, Flowers, Meyer, & Hill, 2002, November). The ROC/AUC is calculated using the full range of continuous scores; thus, the value does not vary by score. In general, ROC/AUC values between .70 and .90 represent accuracies that are acceptable, and higher values represent better accuracy (Swets, 1996). The ROC/AUC values, as well as the sensitivity of the BYI-D, are acceptable; however, the specificity was .69, which just missed being in the acceptable range. Table 2 demonstrates that a cutoff score of 25 on the BYI-D provides the best balance between sensitivity and specificity (sensitivity = .72, specificity = .69, ROC/AUC = .77). In regard to the CDI, results of diagnostic accuracy analyses indicated that the cutoff score of 16 resulted in a high level of sensitivity (.95), but low specificity (.22). The ROC/AUC value (.70) was in the acceptable range. It is important to note, however, that the CDI was not administered to the control group, and as a result, its analyses contained fewer true negatives than the analyses conducted for the BYI-D. The lower number of true negatives may have limited the investigation of specificity of the CDI; this index might have been higher had the CDI been administered to the control group.

A simultaneous logistic regression was also conducted with the BYI-D and CDI to further assess the overall classification accuracy of the measures and predict which students would become depressed (see Table 3). A test of the full model against a constant only model was statistically significant, indicating that the measures reliably distinguished between depressed and non-depressed girls, $\chi^2 (2, N = 275) = 35.68$, $p < .001$. The Wald criterion demonstrated that the BYI-D and CDI made significant contributions to prediction ($p = .017$ and $p = .008$, respectively).

Table 3
Results from Logistic Regression Predicting Depression

Predictors	Estimate	*SE*	Ratio	95% CI— Lower Bound	95% CI— Upper Bound
Intercept	–2.32	.52			
BYI-D	.052	.02	1.05	1.00	1.09
CDI	.063	.03	1.07	1.02	1.12

Note. CI = confidence interval.

Evaluation of Bias

To investigate differences in identification rates across ethnic and racial groups for each stage of the procedure, chi-square tests were performed for each measure. Participants' scores were categorized as low in depressive symptomatology (below the cutoff score) or high in depressive symptomatology (at or above the cutoff score). Identification rates (low vs. high) for each ethnicity were not significantly different for the BYI-D, $\chi^2(5, N = 1880) = 7.0$, $p = .224$; CDI, $\chi^2 (5, N = 2799) = 7.1$, $p = .217$; DSM-BSI, $\chi^2 (5, N = 653) = 1.7$, $p = .893$; and K-SADS-P-IVR, $\chi^2 (5, N = 382) = 9.7$, $p = .08$.

DISCUSSION

Early detection of depressed youth is crucial due to the pernicious nature of the disorder, and considerable evidence supporting the use of CBT for depression for youth in schools exists (Curry, 2001; Weersing & Weisz, 2002). A multiple-stage procedure has been recommended for use in schools to identify depressed youth (e.g. Reynolds, 1986). This study sought to evaluate the efficiency of a multiple-stage procedure used in the context of recruitment for a larger treatment study. Hit rates for screening measures were calculated for Stages 1 and 2, and diagnostic accuracy of screening measures was also examined. Results indicate that the present study utilized a multiple-stage procedure that successfully identified girls between the ages of 8 and 13 who were experiencing a depressive disorder. Specifically, of the original sample (3318 girls), 859 girls (26%) met the cutoff criterion at the initial screening (Stage 1). However, only 378 girls (11% of initial sample) had to be assessed with the time-intensive diagnostic interview. At completion of the multiple-stage procedure, 225 girls (7% of initial sample) were identified as experiencing a depressive disorder. The overall hit rate (26%) for the first administration of a self-report measures in the present study was similar to the hit rate (24%) calculated by Kaslow et al. (1992). As for Stage 2, the hit rate (47%) for the second administration of the CDI in Cohort 1 of the present study was comparable to the hit rate (41%) calculated for the second administration of the CDI by Kaslow et al. (1992). However, the hit rate (63%) for the brief symptom interview used in Stage 2 in Cohorts 2 through 7 of the present study was higher. The similarity between the hit rates for Cohort 1 of the present study and the previous study suggests generalizability of the results. Interestingly, the brief symptom interview had a higher hit rate than did the previous study, which indicates it may be an improvement to the procedure.

Concerning the efficiency of the procedure, 26% of girls reported elevated symptoms of depression at Stage 1. As indicated by the diagnostic interview, a large number of these girls were not experiencing a depressive disorder. This finding would point to the inefficiency of the multiple-gate procedure, as critiqued by others (e.g., Adelman & Taylor, 2006). However, Stage 2, which was included to reduce the number of false positives prior to engagement in lengthy clinical interviews

with both the child and the caregiver, further reduced the number of girls. With regard to Cohort 1, inclusion of Stage 2 reduced the number of girls who met the cutoff on a second administration of the CDI from 162 girls to 122 girls, a 25% decrease.

Regarding Cohorts 2 through 7, inclusion of a brief symptom interview during Stage 2 reduced the number of girls who met the cutoff criterion for Stage 3 from 697 girls to 334 girls, a 52% decrease. Moreover, the DSM-BSI had a significantly higher hit rate on the KSADS-P-IVR relative to the CDI. These findings indicate that use of a brief symptom interview is more efficient than a self-report measure at Stage 2, as it greatly reduced the number of diagnostic interviews that needed to be conducted at Stage 3. The advantage of the DSM-BSI over the CDI may be related to the nature of the brief interview, namely, that questions can be clarified and participant responses can be queried, which could elicit more comprehensive and diagnostically useful responses.

Another goal of the present study was to evaluate the diagnostic accuracy of the CDI and the BYI-D by examining their sensitivity, specificity, and ROC/AUC indices. Analyses indicated that the cutoff scores used in the present study for both measures resulted in acceptable levels of sensitivity. With regard to specificity, the indices for the CDI and BYI-D were lower than .70, which was not unexpected, given that the aim of the multiple-stage screenings in the present study was to maximize the identification of depressed girls for a school-based treatment trial (Stark et al., 2015). As such, it was still acceptable, given the goals of the study. The ROC/AUC indices for the BYI-D and CDI were in the acceptable range; further, results of the logistic regression indicated that the measures reliably distinguished between depressed and non-depressed girls. Thus, the diagnostic accuracy of both measures was supported.

Lastly, schools must also ensure that measures they employ during universal screenings are not biased against ethnic or racial groups. A concern about large-scale screening of mental health disorders is the over-identification of depression in certain racial or ethnic groups. Results indicated that rates of identification of depression for different racial and ethnic groups for each measure did not differ; as such, there was no relationship between race or ethnicity and scores on any of the measures. This finding provides support for use of this procedure and these measures in racial and ethnic minority populations.

Implications

Universal mental health screenings that are guided by research can help efficiently and accurately identify students in need of intervention and support (Weist et al., 2007). Although these screenings would potentially require increased resources from school personnel, there are ways to increase the efficiency and cost-effectiveness of the procedure. First, the Stage 3 assessments required in the multiple-stage procedure can be time-intensive and conducting a large number of them would likely not be feasible in schools where the workload of the staff is already high and resources are limited. Therefore, a school might focus on increasing the ability of a measure to categorize non-depressed students more accurately rather than depressed students. In other words, the cutoff score on screening measures could be raised to lower the number of false positives (thus increasing the specificity of the measure). For example, to decrease the number of false positives on the measures used in the present study, the cutoff score could be raised to from 16 to 22 on the CDI and from 25 to 26 on the BYI-D. If resources are very limited, the cutoff scores might be raised even more, perhaps to a standard deviation and a half above the mean. As can be seen in Table 4, if the cutoff score for the BYI-D was raised by a standard deviation and a half (a score of 30), number of false positives would have been reduced by more than half. However, it is important to note that the number of false negatives would also increase, and some depressed students would not be identified.

Perhaps students whose scores fell between the original cutoff score and the higher cutoff score could be considered at risk and be monitored more closely to see if they actually develop depression.

In addition to increasing specificity of self-report measures in Stage 1, efficiency of the procedure was further enhanced by using a brief interview at Stage 2. When examining the multiple-stage procedure, it was posited that some of the girls who were false positives in Stage 1 were experiencing a transient mood disturbance (but not a depressive disorder) or were over-reporting their symptoms. An informal review of the clinician's notes on the DSM-BSI indicated that many of the girls who reported high depressive symptoms on the CDI or BYI-D were actually experiencing a transient mood disturbance (e.g., having a bad day), had misinterpreted items, or did not understand scaling of the responses. However, further research that investigates the decline in scores is needed. In addition, it should be noted that the threshold for advancing to Stage 3 was set very low for the DSM-BSI, as the interview was new and the investigators did not want to exclude a participant who was actually depressed. It would have been possible to set a more stringent standard DSM-BSI that would have further reduced the number of false positives and thus, reduced the time needed for follow-up assessments.

Efficiency and cost-effectiveness of the multiple-stage procedure can also be improved by involving teachers as well as other mental health professionals in the process. An application of the multiple-stage assessment in schools could be executed as follows: Teachers could monitor distribution and return of consent forms and administer Stage 1 self-report measures, whereas school psychologists, school counselors, and/or school nurses could score and interpret measures to protect confidentiality. However, although allowing teachers to conduct Stages 1 and 2 of the screenings would make the procedure more cost-effective, issues of confidentiality and training cannot be ignored, especially if teachers score the measures. In addition, scoring and interpreting the measures used in the present study require qualifications that teachers usually do not have (e.g., graduate-level courses in assessment). Moreover, the measures like the CDI and BYI-D contain items concerning self-harm or suicidality and if students endorse those items, they should be assessed by a qualified professional and provided with appropriate services. Thus, school psychologists or other qualified individuals should conduct the DSM-BSI at Stage 2 and the clinical interviews at Stage 3. Additionally, school districts could partner with community mental health centers when conducting screenings to reduce the workload. Students experiencing moderate to severe distress could be referred for treatment in the community or receive counseling at school. School psychologists may also consider forming groups to help identified students learn skills to cope with their feelings, negative thoughts, and maladaptive behaviors.

Limitations

There are several limitations to the current investigation. The diagnostic accuracy analysis of the CDI was limited since it was not administered to the non-depressed control group. Moreover, due to time and budget constraints of the larger treatment study, the KSADS-P-IVR interviews were only administered to girls who met the cutoff at Stage 2, which also limits the examination of the hit rates and accuracy of the measures. Future studies of the multiple-stage procedure should assess more of the participants who did not exceed the cutoffs in Stages 1 and 2 with clinical interviews to better examine diagnostic accuracy of measures. Also, information regarding previous diagnoses and intervention history of the participants could be collected in future studies, as it could facilitate better understanding of the nature and course of depression.

There are also limitations concerning other aspects of the measures. The present study used the first editions of the BYI-D and the CDI; second editions have since been published. The second editions of the measures are very similar to the first editions, although they have updated normative

information and reportedly improved psychometrics. The diagnostic accuracy of the second editions could be investigated in future studies. In addition, the measures were written in English, and limited English language proficiency can affect students' comprehension of the questions. In the present study, participants self-identified as having language limitations and school staff were informally consulted to help identify students with language difficulties; future studies could investigate the effect of language proficiency more systematically.

Another concern is the time needed to administer the measures and clinical interviews. Although the self-report measures used in the present study are group administered, brief, and easy to administer and score, the assessment time needed for the students is not insignificant, especially due to the need for individual clinical interviews. Future studies could include teacher ratings of students in addition to other measures, as teachers could quickly complete ratings of students. Teacher ratings would not require additional assessment time for students, although it would increase the workload of the teachers. The time needed for teachers or other staff to implement the multiple-stage procedure is an important issue, and it needs to be evaluated in future investigations of efficiency of the multiple-stage procedure.

Future Directions

Depression is a serious disorder that adversely impacts youth. To assist depressed youth, they must first be identified, and, despite the limitations of the multiple-stage procedure, it facilitates large-scale screenings that identify depressed youth. Although there are concerns about the appropriateness of schools for mental health screenings (e.g., Laurent et al., 1994), it may be in the school's best interests to identify and treat depressed youth, given the effect of depression on school outcomes. Even if resources available to schools limit the number of children who can be screened, at least some depressed children will be identified and potentially receive help. The multiple-stage procedure should continue to be refined, and future studies could examine how to reduce false positives at Stages 1 and 2. Furthermore, studies of how schools independently apply the procedure would provide valuable information about the feasibility of the procedure, as it would be conducted solely by school staff.

REFERENCES

Adelman, H. S., & Taylor, L. (2006). Mental health in schools and public health. Public Health Reports, 121, 294–298.

Ambrosini, P. J. (2000). Historical development and present status of the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS). Journal of the American Academy of Child and Adolescent Psychiatry, 39, 49–58.

Ambrosini, P. J., & Dixon, J. F. (2000). Schedule for Affective Disorders & Schizophrenia for School Age Children (6–18 years)–Kiddie-SADS (K-SADS-P-IVR) (present state and lifetime version) K-SADS-IVR (Revision of K-SADS-IIIR). Unpublished manuscript, Eastern Pennsylvania Psychiatric Institute, Philadelphia, PA.

Ambrosini, P. J., Metz, C., Prabucki, K., & Lee, J. (1989). Videotape reliability of the third revised edition of the K-SADS. Journal of the American Academy of Child and Adolescent Psychiatry, 28, 723–728.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.

Apter, A., Orvaschel, H., Laseg, M., Moses, T., & Tyano, S. (1989). Psychometric properties of the K SADS-P in an Israeli adolescent inpatient population. Journal of the American Academy of Child and Adolescent Psychiatry, 28, 61–65.

Beck, J. S., Beck, A. T., & Jolly, J. B. (2001). Beck Youth Inventories of Emotional and Social Impairment manual. San Antonio, TX: The Psychological Corporation.

Chorpita, B. F., & Donkervoet, C. (2005). Implementation of the Felix consent decree in Hawaii: The impact of policy and practice development efforts on service delivery. In R. G. Steele & M. C. Roberts (Eds.). *Handbook of Mental Health Services for Children, Adolescents, and Families* (pp. 317–332). New York: Kluwer.

Chrisman, A., Egger, H., Compton, S. N., Curry, J., & Goldston, D. B. (2006). Assessment of childhood depression. Child and Adolescent Mental Health, 11, 111–116.

Curry, J. F. (2001). Specific psychotherapies for childhood and adolescent depression. Biological Psychiatry, 49, 1091–1100.

Dekker, M. C., Ferdinand, R. F., vanLang, N. D. J., Bongers, I. L., vander Ende, J., & Verhulst, F. C. (2007). Developmental trajectories of depressive symptoms from early childhood to late adolescence: Gender differences and adult outcome. Journal of Child Psychology and Psychiatry, 48, 657–666.

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. School Mental Health, 2, 166–176.

Finch, A. J., & Saylor, C. F. (1985, February). Test-retest reliability on the Children's Depression Inventory with normal school children. Paper presented at the annual meeting of the Society for Personality Assessment, Berkeley, CA.

Galambos, N. L., Leadbeater, B. J., & Barker, E. T. (2004). Gender differences in and risk factors for depression in adolescence: A 4-year longitudinal study. International Journal of Behavioral Development, 28, 16–25.

Garber, J., & Horowitz, J. L. (2002). Depression in children. In I. H. Gotlib & C. L. Hammen (Eds.), Handbook of depression (pp. 510–540). New York, NY: Guilford.

Hammen, C., & Rudolph, K. D. (2003). Childhood mood disorders. In E. J. Mash & R. A. Barkley (Eds.), Child psychopathology (pp. 233–278). New York, NY: Guilford Press.

Hankin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. Psychological Bulletin, 127, 773–796.

Jansky, J. J. (1978). A critical review of some developmental and predictor precursors of reading disabilities. In A. L. Benton & D. Pearl (Eds.), Dyslexia (pp. 412–516). New York, NY: Oxford University Press.

Kaslow, N. J., Stark, K. D., Printz, B., Livingston, R., & Tsai, S. (1992). Cognitive Triad Inventory for Children: Development and relation to depression and anxiety. Journal of Clinical Child Psychology, 21, 339–347.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). X. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry, 36, 980–988.

Keller, M. (2003). Past, present, and future directions for defining optimal treatment outcome in depression: Remission and beyond. Journal of the American Medical Association, 289, 3152–3160.

Kendall, P. C., Cantwell, D. P., & Kazdin, A. E. (1989). Depression in children and adolescents: Assessment issues and recommendations. Cognitive Therapy & Research, 13, 109–146.

Klein, D. N., Dougherty, L. R., & Olino, T. M. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34, 412–432.

Kovacs, M. (1992). Children's Depression Inventory (CDI) manual. North Tonawanda, NY: Multi-Health Systems, Inc.

Laurent, J., Hadler, J. R., & Stark, K. D. (1994). A multiple-stage screening procedure for the identification of childhood anxiety disorders. School Psychology Quarterly, 9, 239–255.

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of implementation. Journal of School Psychology, 45, 163–191.

Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. Journal of Clinical Child and Adolescent Psychology, 34, 362–379.

Mennuti, R. B., Christner, R. W., & Freeman, A. (2006). An introduction to a school-based cognitive-behavioral framework. In R. B. Mennuti, A. Freeman, & R. W. Christner (Eds.), *Cognitive behavioral interventions in educational settings: A handbook for practice* (pp. 3–19). New York: Routledge.

Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. New York, NY: Oxford.

President's New Freedom Commission on Mental Health. (2003). Achieving the promise: Transforming mental health care in America. Retrieved from http://govinfo.library.unt.edu/mentalhealthcommission/reports/Final Report/downloads/FinalReport.pdf

Reynolds, W. M. (1986). A model for the screening and identification of depressed children and adolescents in school settings. Professional School Psychology, 1, 117–129.

Seligman, M. E. P., Peterson, C., Kaslow, N. J., Tanenbaum, R. L., Alloy, L. B., & Abramson, L. Y. (1984). Attributional style and depressive symptoms among children. Journal of Abnormal Psychology, 93, 235–238.

Shirk, S., Kaplinski, H., & Gudmudsen, G. (2009). School-based cognitive-behavioral therapy for adolescent depression: A benchmarking study. Journal of Emotional and Behavioral Disorders, 17, 106–117.

Simons, A. D., Rohde, P., Kennard, B. D., & Robins, M. (2005). Relapse and recurrence prevention in the treatment for adolescents with depression study. Cognitive and Behavioral Practice, 12, 240–251.

Sitarenios, G., & Kovacs, M. (1999). Use of the Children's Depression Inventory. In M. E. Maruish (Ed.), The use of psychological testing for treatment planning and outcomes assessment (2nd ed., pp. 267–298). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Stapleton, L. M., Sander, J. B., & Stark, K. D. (2007). Psychometric properties of the Beck Depression Inventory for Youth in a sample of girls. Psychological Assessment, 19, 230–235.

Stark, K. D., Arora, P., Krumholz, L., Morey, M. E., & Jensen-Doss, A. (2015). Evaluation of cognitive behavioral therapy with and without parent training for the treatment of depressed girls. Manuscript in preparation.

Stark, K. D., & Sander, J. B. (2002). Diagnostic and statistical manual brief symptom interview for depression. Unpublished manuscript, University of Texas at Austin.

Stark, K. D., Sander, J. B., Yancy, M. G., Bronik, M. D., & Hoke, J. A. (2000). Treatment of depression in childhood and adolescence: Cognitive-behavioral procedures for the individual and the family. In P. C. Kendall (Ed.), Child and adolescent therapy: Cognitive-behavioral procedures (2nd ed., pp. 173–234). New York, NY: Guilford Press.

Swets, J. A. (1996). Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trimbremont, B., Braet, C., & Dreessen, L. (2004). Assessing depression in youth: Relation between the children's depression inventory and a structured interview. Journal of Clinical Child and Adolescent Psychology, 33, 149–157.

Public Health Service, U. S. (2000). Report on the Surgeon General's Conference on Children's Mental Health: A national action agenda. Washington, DC: U.S. Government Printing Office.

Waslick, B. D., Kandel, R., & Kakouros, A. (2002). Depression in children and adolescents: An overview. In D. Shaffer & B. D. Waslick (Eds.), The many faces of depression in children and adolescents (pp 1–36). Washington, DC: American Psychiatric Publishing.

Weersing, V. R., & Weisz, J. R. (2002). Mechanisms of action in youth psychotherapy. Journal of Child Psychology and Psychiatry, 43, 3–29.

Weist, M. D., Rubin, M., Moore, E., Adelsheim, S., & Wrobel, G. (2007). Mental health screening in schools. Journal of School Health, 77, 53–58.

Wood, F., Flowers, L., Meyer, M., & Hill, D. (2002, November). How to evaluate and compare screening tests: Principles of science and good sense. Paper presented at the annual meeting of the International Dyslexia Association, Atlanta, GA.