Psychological Assessment
2010, Vol. 22, No. 3, 702–710

1040-3590/10/$12.00 DOI: 10.1037/a0019915

Screening for Depressive Disorders Using the Mood and Anxiety Symptoms Questionnaire Anhedonic Depression Scale: A Receiver-Operating Characteristic Analysis

Keith Bredemeier, Jeffery M. Spielberg, Rebecca Levin Silton, Howard Berenbaum, Wendy Heller, and Gregory A. Miller
University of Illinois at Urbana–Champaign

The present study examined the utility of the anhedonic depression scale from the Mood and Anxiety Symptoms Questionnaire (MASQ–AD scale) as a way to screen for depressive disorders. Using receiver-operating characteristic analysis, we examined the sensitivity and specificity of the full 22-item MASQ–AD scale, as well as the 8- and 14-item subscales, in relation to both current and lifetime *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) depressive disorder diagnoses in two nonpatient samples. As a means of comparison, the sensitivity and specificity of a measure of a relevant personality dimension, Neuroticism, was also examined. Results from both samples support the clinical utility of the MASQ–AD scale as a means of screening for depressive disorders. Findings were strongest for the MASQ–AD 8-item subscale and when predicting current depression status. Furthermore, the MASQ–AD 8-item subscale outperformed the Neuroticism measure under certain conditions. The overall usefulness of the MASQ–AD scale as a screening device is discussed, as are possible cutoff scores for use in research.

Keywords: depressive disorders, anhedonic depression, Mood and Anxiety Symptoms Questionnaire, receiver-operating characteristic analysis, screening

There are a variety of strategies that clinical researchers can use to recruit individuals with specific forms of psychopathology. One strategy is to target individuals seeking treatment for the condition of interest. A key limitation of this strategy is that those individuals seeking treatment can be expected to be unrepresentative of individuals who have that condition (du Fort, Newman, & Bland, 1993). An alternative approach is to use specific advertising techniques to target individuals who report having these conditions, although again there is no way to ensure that those who respond are representative. A third approach is to screen, using diagnostic interviews, a very large number of individuals (with the number of individuals to be screened guided by base rates). This strategy can be very inefficient because of the relatively large amount of time that must be devoted to screening each participant.

Keith Bredemeier, Jeffrey M. Spielberg, Rebecca Levin Silton, and Howard Berenbaum, Department of Psychology, University of Illinois at Urbana–Champaign; Wendy Heller and Gregory A. Miller, Department of Psychology and Beckman Institute Biomedical Imaging Center, University of Illinois at Urbana–Champaign.

Rebecca Levin Silton is now at the Department of Psychiatry and Behavioral Medicine, Seattle Children's Hospital, Seattle, Washington.

This work was supported by National Institute of Mental Health Grants R01 MH61358, T32 MH19554, T32 MH067533, and P50 MH079485. We thank Adrienne Abramowitz, Joscelyn Fisher, Brenda Hernandez, and Angela Lawson for their assistance in conducting structured clinical interviews.

Correspondence concerning this article should be addressed to Keith Bredemeier, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: kbredem2@uiuc.edu

A related recruitment approach involves screening a large number of participants with an instrument that can be administered quickly and easily and then conducting follow-up assessments with a subset of these individuals using more extensive diagnostic procedures. This approach has the advantages of being more efficient than conducting full assessments with a large number of participants and having the ability to identify non-treatment-seeking individuals with psychopathology. Of course, the feasibility of adopting this approach is premised on two conditions: (a) that sufficiently predictive instruments have been identified that can accurately distinguish individuals who are likely to meet diagnostic criteria from individuals who are likely to not meet diagnostic criteria and (b) that information is available for determining an appropriate cutoff value that can be used for screening decisions. Both of these conditions can be addressed using receiver-operating characteristic (ROC) analysis (Green & Swets, 1966). In ROC analysis, one obtains a curve in which the sensitivity (i.e., the rate at which the instrument at a given value indicates the presence of a condition when the condition is actually present) is plotted against the specificity (i.e., the rate at which the instrument at a given value indicates the absence of a condition when the condition is not actually present) for the full range of scores on a given measure. The adequacy of a given measure as a screening tool can be determined by calculating the area under the ROC curve (AUC). AUC reflects the probability that a randomly selected case will score higher on the test or measure than a randomly selected control participant (Hanley & McNeil, 1982). Furthermore, sensitivity and specificity for specific scores on the measure can be examined to determine an appropriate clinical cutoff. ROC analysis is growing in popularity as a procedure for

evaluating the utility of specific self-report instruments as screening tools for use in clinical research (e.g., Behar, Alcaine, Zuellig, & Borkovec, 2003; Chen, Faraone, Biederman, & Tsuang, 1994), in part because test results are robust even when the numbers of cases and control participants are unequal in the sample (Rice & Harris, 1995).

Because the majority of individuals with depressive disorders do not seek professional treatment (Flament, Cohen, Choquet, Jeammet, & Ledoux, 2001; Kendler, 1995), recruiting research participants from clinical settings will likely exclude a very large proportion of the population with depressive disorders. Furthermore, motivational deficits, coupled with the stigma associated with mental disorders, may make depressed individuals unlikely to respond to targeted advertisements. Finally, the base rates of depressive disorders, although higher than some other forms of psychopathology, are still low enough that conducting diagnostic procedures with an unselected sample of participants would not be very cost effective. Thus, there is a clear need for self-report instruments that can be administered quickly and easily and can accurately identify individuals likely to have depressive disorders. Research involving ROC analysis has examined the utility of popular self-report measures of depression, such as the Beck Depressive Inventory (BDI; Beck, Steer, & Brown, 1996; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). Findings from these studies have generally been encouraging (e.g., Kumar, Steer, Teitelman, & Villacis, 2002; Lasa, Ayuso-Mateos, Vázquez-Barquero, Díez-Manrique, & Dowrick, 2000). Nevertheless, many of these popular measures have been criticized as having poor discriminant validity, given that they primarily measure general distress or negative affect, which is not unique to depression (see Watson & Clark, 1984). One possible implication of this criticism is that these instruments are likely to have high sensitivity but low specificity (see Sloan et al., 2002). According to the tripartite model of depression and anxiety, low levels of positive affect (anhedonia) are unique to depressive disorders, whereas elevated levels of negative affect are shared by both depressive and anxiety disorders (Clark & Watson, 1991). Self-report instruments have been developed to measure this unique component of depression, perhaps the most popular being the Mood and Anxiety Symptoms Questionnaire (MASQ; Watson, Clark, et al., 1995; Watson, Weber, et al., 1995). The MASQ includes an anhedonic depression scale (MASQ–AD scale), which was designed to measure the low levels of positive affect unique to depression (along with other symptoms that are thought to differentiate depressive disorders from anxiety disorders, e.g., lack of motivation).

Three studies have used ROC procedures to examine the utility of the MASQ–AD scale in clinical settings as a means of identifying individuals with depressive disorders (Boschen & Oei, 2007; Buckby, Yung, Cosgrave, & Cotton, 2007; Buckby, Yung, Cosgrave, & Killackey, 2007). Although all three studies showed that scores on the scale predict depressive disorder diagnoses, some disagreement still exists regarding the ultimate utility of this scale for clinical applications. It is important to note that one of these studies (Buckby, Yung, Cosgrave, & Killackey, 2007) showed that the MASQ–AD scale outperformed a popular measure of depression (the Center for Epidemiologic Studies Depression scale) in predicting depressive disorder diagnoses. To date, no research has examined the utility of the MASQ–AD scale as a means for screening for depressive disorders in nonclinical settings. Such work is critical to exploring the potential utility of the MASQ–AD scale for research applications or for initial screening for depressive disorders in primary health care settings.

Research has shown that items on the MASQ–AD scale load onto two separate factors, one consisting of eight items regarding depressed mood, lack of motivation, and other symptoms of depressive disorders (e.g., "felt really slowed down") and another consisting of 14 reverse-scored items related to experiencing pleasant emotions (e.g., "felt like nothing was very enjoyable"; Nitschke, Heller, Imig, McDonald, & Miller, 2001; Watson, Clark, et al., 1995; Watson, Weber, et al., 1995). Existing ROC research on the MASQ–AD scale has not examined these subscales separately to determine whether one of these subscales outperforms the other and/or the total scale.

In the present project, we examined the utility of the MASQ–AD as a screening tool for depressive disorders using ROC analysis. The utility of the MASQ–AD 22-item scale, as well as that of the 8- and 14-item subscales, was examined in a sample of college students in Study 1 and in a sample of community members in Study 2. We also expanded the present study beyond past research by comparing the MASQ–AD scale and subscales with a measure of Neuroticism, which is a personality trait shown to confer risk for a broad range of psychopathology, including but not limited to depression (Ormel, Rosmalen, & Farmer, 2004).

Study 1

Method

Participants. Participants were 108 university students (60% female) ages 18–22 years (M = 19.0 years, SD = 1.0) who were recruited to participate in a large-scale neuroimaging study. All participants passed exclusion criteria related to a neuroimaging study: having a history of serious brain injury, abnormal hearing or vision, or metal in their body or being left-handed, pregnant, or a nonnative English speaker. For reasons associated with the primary goals of the neuroimaging study, we tried to oversample individuals with symptoms of anxiety and/or depression. To achieve this goal, we initially assessed a large number of individuals (n = 2,637) using self-report measures of anhedonic depression, anxious arousal, and worry. This screening session occurred one to six months prior to the collection of the data reported in this article. Questionnaire scores from this session were used to determine who would participate in the next stage of the research; they were not used in the analyses presented in this article. On the basis of their scores on these questionnaires, five groups of participants were recruited. Specifically, three groups scored above the 80th percentile (percentile levels determined from previous testing; Nitschke et al., 2001) on either the eight-item MASQ–AD subscale (n = 17), the MASQ anxious arousal scale (n = 18), or the Penn State Worry Questionnaire (n = 14; Meyer, Miller, Metzger, & Borkovec, 1990) and below the 50th percentile on the other two scales. A fourth group scored above the 80th percentile on all three measures (n = 29), and the final group scored below the 50th

percentile on all three measures ($n = 29$).[1] All participants received monetary compensation for participating in the study.

Self-report questionnaires.

Anhedonic depression. Participants completed the MASQ–AD scale a second time, after being recruited to participate in the neuroimaging study. Scores from this second administration were used in the analyses reported below. On the MASQ–AD scale, individuals indicate how frequently they have experienced a variety of different symptoms during the past week. This scale is composed of 22 items such as "felt like nothing was very enjoyable" and "felt really slowed down." Research has indicated that this scale has good convergent and discriminant validity in undergraduate and community samples (Nitschke, Heller, Palmieri, & Miller, 1999; Nitschke et al., 2001; Watson, Clark, et al., 1995; Watson, Weber, et al., 1995). Because past research has shown that the items of the MASQ–AD scale load onto two separate factors (Nitschke et al., 2001; Watson, Clark, et al., 1995; Watson, Weber, et al., 1995), analyses were conducted with the full 22-item scale as well as the 8- and 14-item subscales. In Study 1, alphas for the 22-, 8-, and 14-item scales were .94, .94, and .86, respectively.

Neuroticism. Participants also completed the 60-item NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) after being recruited to participate in the study. The Neuroticism scale is composed of 12 items related to experiences of negative affect and self-reproach (see Saucier, 1998). Participants rated how characteristic each statement is of them. Research has indicated that this scale has good reliability and convergent validity in a variety of samples (Costa & McCrae, 1992). In the present sample, alpha for the Neuroticism scale was .93.

Diagnostic interview. Within approximately two weeks of completing the questionnaires described above, each participant was interviewed by an advanced doctoral student in clinical psychology using the Structured Clinical Interview for *DSM–IV* Disorders, nonpatient edition (SCID–NP; First, Spitzer, Gibbon, & Williams, 2002) to assess for symptoms of Axis I pathology. All final diagnostic decisions were determined through consensus of the interviewers in consultation with Gregory A. Miller, a licensed clinical psychologist who has supervised over 2,000 diagnostic assessments that used the SCID. Interviewers were unaware of participants' scores on the self-report questionnaires.

For the current study, we used information gathered during the SCID–NP to classify all participants on four variables related to current and lifetime depressive disorder diagnoses. The first variable was based on whether the participant met *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*) diagnostic criteria for a current major depressive episode (MDE) at the time of the interview. The second variable was based on whether the participant met diagnostic criteria for any current *DSM–IV* depressive disorder at the time of the interview. This included individuals who met full diagnostic criteria for a current MDE, as well as individuals who met full diagnostic criteria for dysthymic disorder, substance-induced mood disorder with depressive features, mood disorder due to a general medical condition with depressive features, or depressive disorder not otherwise specified at the time of the interview. The third variable was based on whether the participant met *DSM–IV* diagnostic criteria for lifetime major depressive disorder (MDD). The fourth variable was based on whether the participant had ever met diagnostic criteria for any *DSM–IV* depressive disorder or a bipolar disorder (bipolar I, bipolar II, or cyclothymic disorder) with a history of clinically significant depressive symptoms.[2] These diagnostic variables were not treated as mutually exclusive; thus, participants who qualified for current MDE also qualified for current depressive disorders, those who qualified for current MDE also qualified for lifetime MDD, and those who qualified for current depressive disorders also qualified for lifetime depressive disorders.

To examine interrater reliability, we had a secondary rater (unaware of the original diagnoses) listen to audiotaped SCID–NP interviews of 20 participants (10 randomly selected participants who, according to the original diagnostician, met diagnostic criteria for lifetime MDD and 10 randomly selected participants who did not) and provided ratings for the four diagnostic variables. Kappa was 1.00 for current MDE, 1.00 for current depressive disorders, .70 for lifetime MDD, and .90 for lifetime depressive disorders.

Analyses. The four diagnostic variables (current MDE, current depressive disorders, lifetime MDD, and lifetime depressive disorders) were used as the criteria for evaluating the absolute and relative effectiveness of the self-report questionnaires as a means of screening for depressive disorders, using ROC procedures. In each of these analyses, all participants in the sample who did not qualify for the diagnosis in question were treated as controls for that variable. For example, for analyses conducted with the current MDE variable, all participants in the sample who did not meet full criteria for a current MDE (including remaining participants who qualified for the other diagnostic variables, such as lifetime MDD) were treated as controls. AUCs were calculated for each instrument to quantify the general utility of each scale as a means of screening for current and lifetime depressive disorder. Statistical significance of AUC estimates (i.e., whether these estimates are significantly above chance, which is .50) was determined using nonparametric tests (Hanley & McNeil, 1983). Because these tests are nonparametric, they do not require any statistical assumptions about the distributions of questionnaire scores and/or the base rates on the diagnostic variables. Although no specific guidelines for interpreting the size of AUC estimates are currently available, the following have been used across a wide range of disciplines (e.g., Luna-Herrera et al., 2003; Starr et al., 2004; Thuiller et al., 2003): .50–.60 = fail, .60–.70 = poor, .70–.80 = fair, .80–.90 = good, .90–1.0 = excellent. All analyses were conducted using SPSS, Version 16.0.

To compare the relative effectiveness of each of the scales, we contrasted AUCs for the different self-report scales using the procedures for comparing correlated ROC curves described by DeLong, DeLong, and Clarke-Pearson (1988). These analyses were conducted using locally written Matlab programs (Matlab R2007a, Natick, MA). Also, planned follow-up analyses for the MASQ–AD scale and subscales were conducted by calculating sensitivity, specificity, positive predictive power, and negative predictive power for specific scale values, which, in turn, were

[1] One participant did not meet criteria for any of the five groups. Because group membership was not directly relevant to the present project, data from this individual were included in the analyses.

[2] All analyses for the lifetime depressive disorders variable were rerun excluding individuals who qualified for bipolar disorders, and the results were virtually identical in both samples.

used to explore optimal clinical cutoffs.[3] To determine optimal cutoff scores, the Youden (1950) index was computed, which has been shown to yield lower misclassification rates than other commonly used methods (Perkins & Schisterman, 2006).

Results and Discussion

Table 1 provides means, standard deviations, and ranges for the self-report measures. The descriptive statistics for this sample closely resemble those reported in past studies involving unselected student samples (e.g., Watson, Clark, et al., 1995). On the basis of the SCID–NP, three participants (2.7%) met criteria for a current MDE and six participants (5.6%) met criteria for current depressive disorders. Seventeen participants (15.7%) met criteria for lifetime MDD, and 28 (25.9%) participants met criteria for lifetime depressive disorders. The rates in this sample of participants, whose mean age was 19.0 years, are comparable to rates reported in epidemiological studies of depressive disorders in older adolescents (e.g., Lewinsohn, Rohde, & Seeley, 1998).

Table 2 contains AUCs for the four criterion variables: current MDE, current depressive disorders, lifetime MDD, and lifetime depressive disorders. As can be seen in Table 2, the self-report scales effectively predicted depressive disorder diagnoses, particularly for current MDE and depressive disorders. Specifically, the MASQ–AD eight-item subscale and the Neuroticism scale reliably predicted whether participants met criteria for a current MDE. The AUCs for both scales were in the good range, with the Neuroticism scale bordering on the excellent range. The MASQ–AD eight-item subscale and the Neuroticism scale also predicted lifetime MDD, with AUCs in the fair range. In addition, the full 22-item MASQ–AD scale predicted lifetime MDD, although the AUC for this scale bordered on the poor range. All of the scales predicted current depressive disorders, with all of the AUCs for the MASQ–AD scale and subscales in the good range and the AUC for the Neuroticism scale bordering on the excellent range. In addition, all of the scales predicted lifetime depressive disorders, with the AUCs for the full 22-item MASQ–AD scale, the eight-item subscale, and the Neuroticism scale in the fair range and the AUC for the MASQ–AD 14-item subscale in the poor range.

Given that the results of the ROC analyses largely supported the effectiveness of all of the self-report measures as a means of screening for depressive disorders, we proceeded to examine whether the four scales differed from one another by conducting pairwise comparisons of the AUCs. The results revealed one statistically significant difference: the Neuroticism scale outperformed the MASQ–AD 14-item subscale as a predictor of lifetime depressive disorders, $\chi^2 = 4.29$, $p = .04$. No other pairs of scales differed significantly from one another.

Given that one of the primary goals of the project was to explore possible cutoff scores on the MASQ–AD scale and subscales that could be used to screen for depressive disorders, sensitivity, specificity, positive predictive power, and negative predictive power were computed for specific scale values. This was done using the eight-item subscale to predict current MDE status because that subscale slightly outperformed the full 22-item scale and the 14-item subscale in the ROC analyses, and the results were stronger for current than for lifetime depressive disorders.[4] The results are presented in Table 3. A cutoff score of 21 on the MASQ–AD eight-item subscale maximized the Youden index, thus achieving a balance of sensitivity and specificity. At this cutoff, negative predictive power was excellent (1.0), although positive predictive power was fairly low (.13).

Overall, the results from Study 1 support the utility of the MASQ–AD scale as a means of screening for depressive disorders, although the MASQ–AD scale did not significantly outperform the Neuroticism scale in predicting current or lifetime diagnostic status. The results were stronger for all of the self-report measures when predicting current rather than lifetime diagnostic status.

Nevertheless, these findings are qualified by some important limitations. The sample for this study was one of convenience, consisting of undergraduate participants preselected on the basis of their scores on several self-report scales, including one of the MASQ–AD subscales. Consequently, it is possible that certain portions of the population distributions for the self-report measures (in particular, the MASQ–AD scale and subscales) were unrepresented or underrepresented in this sample. In turn, this may have inflated their discriminative power. Furthermore, the fact that the sample consisted solely of college students raises questions about the generalizability of the results. To address these limitations, in Study 2, we sought to replicate the findings from Study 1 in a sample of unselected community participants.

Study 2

Method

Participants. Participants were 167 community members (65% women) ages 19–51 years ($M = 34.7$ years, $SD = 9.2$). Participants were recruited though advertisements targeting adults interested in participating in a neuroimaging study and were screened for the same exclusion criteria as used in Study 1. All participants received monetary compensation for participating.

Self-report questionnaires.

Anhedonic depression. Participants completed the MASQ–AD scale. As in Study 1, analyses examined the 22-, 8-, and 14-item scales. In the present sample, alphas were 92, .94, and .80, respectively. One participant had missing data on the 22- and 14-item versions of the MASQ–AD scale and was excluded from analyses involving these scales.

Neuroticism. Participants also completed the NEO-FFI. In the present sample, the alpha for the Neuroticism scale was .88. Twelve participants had missing data on the NEO-FFI and were excluded from analyses involving this scale.

Diagnostic interview. Within approximately two weeks of completing the questionnaires described above, each participant was interviewed by a clinical psychology graduate student using the SCID–NP, which was used classify participants on the same four variables described in Study 1 (current MDE, current depressive disorders, lifetime MDD, and lifetime depressive disorders).

[3] Positive and negative predictive power were included because some have argued that these indices are more clinically meaningful than sensitivity and specificity (Kessel & Zimmerman, 1993; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984).

[4] Statistics for specific values on the other self-report scales that were examined, as well as for the other diagnostic variables, are available from the authors on request.

Table 1
Descriptive Statistics for the Self-Report Measures From Studies 1 and 2

	Study 1			Study 2		
Self-report scale	*M*	*SD*	Range	*M*	*SD*	Range
MASQ–AD 22-item scale	55.4	13.9	26–101	54.4	13.0	30–92
MASQ–AD 8-item subscale	15.7	5.1	8–35	15.1	4.6	8–32
MASQ–AD 14-item subscale	39.7	10.8	14–66	39.2	10.1	14–67
NEO-FFI Neuroticism scale	35.3	9.3	12–60	33.2	8.9	14–58

Note. In Study 2, $n = 166$ for the MASQ–AD 22-item scale and 14-item subscale, $n = 167$ for the MASQ–AD 8-item subscale, and $n = 154$ for the NEO-FFI Neuroticism scale. MASQ–AD = Mood and Anxiety Symptoms Questionnaire anhedonic depression scale; NEO-FFI = NEO Five-Factor Inventory.

As in Study 1, these diagnostic variables were not treated as mutually exclusive, interviewers were unaware of participants' scores on the self-report questionnaires, and all final diagnostic decisions were determined through consensus.

Again, a secondary rater listened to audiotaped SCID–NP interviews of 20 participants (10 randomly selected participants who, according to the original diagnostician, met criteria for lifetime MDD and 10 randomly selected participants that did not) and provided ratings for the four diagnostic variables. Kappa was 1.00 for current MDE, .77 for current depressive disorders, 1.00 for lifetime MDD, and .76 for lifetime depressive disorders.

Analyses. The same analyses were conducted as in Study 1. Again, all analyses were conducted using SPSS, Version 16.0, and locally written Matlab programs.

Results and Discussion

Table 1 contains means, standard deviations, and ranges for self-report measures. The descriptive statistics for this sample closely resemble those reported in past studies involving unselected adult samples (e.g., Watson, Clark, et al., 1995). On the basis of the SCID–NP, five participants (3.0%) met criteria for a current MDE and 11 (6.6%) participants met criteria for current depressive disorders. Fifty-five participants (32.9%) met criteria for lifetime MDD and 81 (48.5%) participants met criteria for lifetime depressive disorders. The rates in this sample for current diagnoses were comparable to rates reported in epidemiological research, although the rates for the two lifetime variables were higher than estimates from past research (e.g., American Psychiatric Association, 2000; Kessler et al., 2003).

As can be seen in Table 2, for the most part, the self-report scales effectively predicted current MDE and depressive disorders, with the eight-item MASQ–AD subscale performing particularly well. Specifically, all four self-report scales predicted whether participants met criteria for a current MDE, with the AUCs for the full 22-item MASQ–AD scale and the eight-item subscale in the good range and the AUCs for the Neuroticism scale and the 14-item MASQ–AD subscale in the fair range. Likewise, all four self-report scales predicted whether participants met criteria for a current depressive disorder, with the AUCs for each scale falling into these same ranges. Unlike Study 1, the self-report scales did not predict lifetime MDD and depressive disorders very well. Specifically, only the full 22-item MASQ–AD scale and the 14-item subscale significantly predicted lifetime MDD diagnosis, although the AUCs were both in the poor range. The full 22-item MASQ–AD scale, the eight-item subscale, and the Neuroticism scale predicted lifetime depressive disorders, although again the AUCs for all of these scales were in the poor range.

Again, the results of comparisons of the AUCs for the self-report measures revealed very few significant differences. In this sample, the MASQ–AD eight-item subscale outperformed the Neuroticism scale, $\chi^2 = 3.71$, $p = .05$, and the MASQ–AD 22-item scale outperformed the 14-item subscale, $\chi^2 = 3.77$, $p = .05$, as a means of predicting current depressive disorders. Also, the Neuroticism scale outperformed the MASQ–AD 14-item sub-

Table 2
Areas Under the ROC Curves for the Four Self-Report Questionnaires Predicting Current and Lifetime Depressive Disorders

	Current				Lifetime			
	MDE		Depressive disorders		MDD		Depressive disorders	
Self-report scale	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
MASQ–AD 22	.83	.83**	.88**	.84**	.70*	.62*	.70**	.61*
MASQ–AD 8	.87*	.85**	.88**	.88**	.72**	.59	.72**	.61*
MASQ–AD 14	.77	.79**	.86**	.79**	.64	.61*	.66*	.58
Neuroticism	.90*	.78*	.90**	.76**	.79**	.59	.79**	.68**

Note. ROC = receiver-operating characteristic; MDE = major depressive episode; MDD = major depressive disorder; MASQ–AD = Mood and Anxiety Symptoms Questionnaire anhedonic depression scale; Neuroticism = Neuroticism scale of the NEO Five-Factor Inventory. Asterisks represent asymptotic significance.
* $p < .05$. ** $p < .01$.

Table 3
Sensitivity, Specificity, Positive Predictive Power (PPP), and Negative Predictive Power (NPP) for Specific Values From the MASQ–AD Eight-Item Subscale Predicting Current Major Depressive Episode Status

	Sensitivity		Specificity		PPP		NPP	
Cutoff score	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
11	1.00	1.00	.11	.12	.03	.03	1.00	1.00
12	1.00	1.00	.17	.21	.03	.03	1.00	1.00
13	1.00	1.00	.23	.32	.04	.04	1.00	1.00
14	1.00	.80	.31	.42	.04	.04	1.00	.98
15	1.00	.80	.41	.54	.05	.04	1.00	.98
16	1.00	.80	.53	.62	.06	.05	1.00	.99
17	1.00	.80	.56	.73	.06	.06	1.00	.99
18	1.00	.80	.63	.77	.07	.08	1.00	.99
19	1.00	.80	.68	.85	.08	.10	1.00	.99
20	1.00	.80	.73	.87	.10	.16	1.00	.99
21	**1.00**	.80	**.80**	.90	**.13**	.19	**1.00**	.99
22	.67	.80	.83	.90	.10	.20	.99	.99
23	.33	**.80**	.85	**.93**	.06	**.25**	.99	**.99**
24	.33	.40	.87	.95	.07	.20	.98	.98
25	.33	.40	.89	.96	.08	.22	.98	.98
26	.33	.40	.89	.97	.08	.29	.98	.98
27	.33	.40	.90	.99	.09	.50	.98	.98
28	.33	.20	.93	.99	.13	.50	.98	.98
29	.00	.20	.93	.99	.00	.50	.98	.96

Note. Statistics for those values that maximized the Youden index in each sample are shown in bold. MASQ–AD = Mood and Anxiety Symptoms Questionnaire anhedonic depression scale.

scale, $\chi^2 = 4.83$, $p = .03$, as a means of predicting lifetime depressive disorders.

Table 3 presents sensitivity, specificity, positive predictive power, and negative predictive power for the eight-item scale when predicting current MDE status (see footnote 4). A cutoff score of 23 on the MASQ–AD eight-item subscale maximized the Youden index, thus achieving a balance of sensitivity and specificity. At this cutoff, negative predictive power was once again excellent (.99). Positive predictive power, although low, was better than in Study 1 (.25).

The results of Study 2 provide additional support for the utility of the MASQ–AD scale and subscales as a means of screening for depressive disorders, thus replicating the main finding from Study 1 without its limitations. Furthermore, unlike in Study 1, the MASQ–AD eight-item subscale outperformed the Neuroticism measure when predicting current depressive disorders. This is consistent with the notion that anhedonia is specific to depressive disorders (relative to anxiety disorders; Clark & Watson, 1991), and thus measures that were specifically developed to tap this dimension of depression are likely to have higher specificity.

The results for all of the self-report measures were again stronger when predicting current rather than lifetime clinical diagnostic status. Unlike Study 1, only the full 22-item MASQ–AD scale and the 14-item subscale predicted lifetime MDD at a level above chance, and the AUCs were not very strong in either case. Furthermore, although the full 22-item MASQ–AD scale, the eight-item subscale, and the Neuroticism scale predicted lifetime depressive disorders at a level above chance, again, none of the AUCs were very strong. This suggests that screening for lifetime history of depressive disorders is difficult in an unselected sample of participants with a broad age range.

General Discussion

In both evaluations of the MASQ–AD scale as a means of screening for depressive disorders, all of the self-report scales that were examined predicted current depressive disorders. In Study 1, all four self-report scales also predicted lifetime depressive disorders, and the MASQ–AD eight-item subscale and the Neuroticism scale predicted current MDE and lifetime MDD. In Study 2, all four self-report scales also predicted current MDE. Furthermore, the eight-item scale significantly outperformed the Neuroticism scale when predicting current depressive disorders.

Overall, these findings support the usefulness of the MASQ–AD scale as way to screen for depressive disorders in nonclinical settings and suggest that the eight-item subscale may be the best means of screening for depressive disorders (among the scales that were examined). Results were stronger for this subscale in several analyses compared with the 14-item subscale and the full 22-item scale. More important, this subscale requires less time to administer. The MASQ–AD scale and subscales appear to be more effective as a means of screening for current than for lifetime depressive disorders. Although the MASQ–AD scale and subscales predicted lifetime MDD and lifetime depressive disorders in Study 1, the results were much less impressive for these two variables in Study 2. This discrepancy could be due to systematic differences between the two samples (e.g., the larger age range of participants in Study 2 relative to Study 1). It is important to note that our results are applicable to categorically defined diagnostic entities (e.g., those in the *DSM–IV*) and not to dimensional definitions of psychopathology.

The AUCs obtained for the MASQ–AD scale and subscales predicting current MDE and depressive disorders were strong in

both samples and were comparable to those reported for some common biomedical tests (Swets et al., 1979) as well as other self-report measures used to screen for psychopathology (e.g., the Penn State Worry Questionnaire as a means of screening for generalized anxiety disorder; Fresco, Mennin, Heimberg, & Turk, 2003; the Beck Depression Inventory as a means of screening for MDD; Kumar et al., 2002). Thus, the MASQ–AD scale appears to be a potentially useful tool for researchers who wish to screen for current depressive disorders. In recent years, increased attention has been devoted to identifying untreated depressed individuals in primary care settings (e.g., Zich, Attkisson, & Greenfield, 1990). The results of the present study indicate that the MASQ–AD scale and, in particular, the eight-item subscale may also be quite useful for depression screening in primary care settings.

To explore what might be an appropriate clinical cutoff score on the MASQ–AD scale when screening for depressive disorders, we examined sensitivity, specificity, positive predictive power, and negative predictive power for specific values in both samples. Results were consistently strongest for the MASQ–AD eight-item subscale and when predicting current diagnostic status, so analyses focused on the eight-item subscale predicting current MDE. Results from Study 1 showed that the optimal cutoff score (based on the Youden index) was 21; in Study 2, the optimal cutoff score was 23. Of course, the most appropriate cutoff score to use for any specific application will depend on the nature of the sample, as well as the relative importance of sensitivity and specificity.

To concretize the implications of present findings for research on depressive disorders, we constructed a table of results for a hypothetical sample of 500 unselected community participants, using data from Study 2 to estimate how participants would be classified as meeting criteria for a current MDE on the basis of a cutoff score of 23 on the MASQ–AD eight-item subscale. Table 4 shows that, of 48 individuals at or above that cutoff, 12 (71%) of the 17 individuals who would meet diagnostic criteria for a current MDE would be correctly identified. The remaining 36 would be judged false positives. Perhaps more important, 447 individuals who would not qualify for a current MDE would be correctly screened out. Clearly, this approach would be much more efficient than conducting full diagnostic assessments with all 500 participants.

The participants in Study 1 were college students selected on the basis of their scores on several questionnaires, including the MASQ–AD scale. Although the means and standard deviations for the MASQ–AD from this sample were comparable to those from past research involving unselected student samples (e.g., Watson, Clark, et al., 1995), concerns could be raised about whether the results from Study 1 would generalize to an unselected sample. To address this concern, in Study 2, we replicated the findings using an unselected sample of community participants.

Table 4
Hypothetical Breakdown of Classification Accuracy Using the MASQ–AD Eight-Item Subscale in an Unselected Sample of 500 Participants to Screen for Current Major Depressive Episodes

Score	Accurate	Inaccurate	Total
Below cutoff (<23)	447	5	452
At or above cutoff (≥23)	12	36	48

Note. Numbers were rounded up for inaccurate classifications and down for accurate classifications. MASQ–AD = Mood and Anxiety Symptons Questionnaire anhedonic depression scale.

As expected, the rates of current depressive disorder diagnoses in both samples were comparable to estimates for the general population. Although the absolute rates of current depressive disorders were low (6.5% in Study 1 and 6.6% in Study 2), this sort of sample is appropriate for examining the utility of screening for depressive disorders in research and primary care settings. The rates of lifetime depressive disorder diagnoses in Study 1 were comparable to the rates reported in previous research examining older adolescents, although the rates of lifetime depressive disorder diagnoses in Study 2 were higher than usually found in the general population. As previously noted, one of the major strengths of ROC methodology is that test results are robust even when the numbers of cases and control participants are unequal in the sample (Rice & Harris, 1995). The fact that our results were very comparable to results from past studies on the MASQ–AD scale using ROC methods (Boschen & Oei, 2007; Buckby, Yung, Cosgrave, & Cotton, 2007; Buckby, Yung, Cosgrave, & Killackey, 2007) serves to bolster confidence in the applicability of our findings. Nonetheless, to be confident that the MASQ–AD scale can be used to successfully select individuals with depressive disorders, whether for research purposes or in primary care settings, additional replication is needed.

The present comparison of the MASQ–AD scale and subscales to a self-report measure of Neuroticism indicated that the MASQ–AD eight-item subscale outperformed the Neuroticism scale under certain circumstances (i.e., when predicting current depressive disorders in an unselected sample). Thus, the MASQ–AD scale may be more appropriate as a means of screening for current depressive disorders than a measure of Neuroticism in unselected participants from a wide range of age groups. Coupled with results from research showing that the MASQ–AD scale can outperform other popular measures of depression (Buckby, Yung, Cosgrave, & Killackey, 2007), these findings also suggest that the MASQ–AD scale may be more appropriate to use for such purposes than scales that primarily gauge high levels of general distress or negative affect. Future research should directly compare the effectiveness of the MASQ–AD scale and subscales and other popular measures of depression (e.g., the Beck Depression Inventory) when screening for depressive disorders in nonclinical settings.

In summary, the findings from the present studies support the utility of the MASQ–AD scale and subscales as a means of screening for depressive disorders. Results were stronger for current than for lifetime depressive disorders and suggest that the eight-item subscale offers the best discriminative power. Investigators may be able to maximize efficiency by using this measure as an initial screening tool in clinical research.

References

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

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961).

An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Behar, E., Alcaine, O., Zuellig, A. R., & Borkovec, T. D. (2003). Screening for generalized anxiety disorder using the Penn State Worry Questionnaire: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 34,* 25–43.

Boschen, M. J., & Oei, T. P. S. (2007). Discriminant validity of the MASQ in a clinical sample. *Psychiatry Research, 150,* 163–171.

Buckby, J. A., Yung, A. R., Cosgrave, E. M., & Cotton, A. M. (2007). Distinguishing between anxiety and depression using the Mood and Anxiety Symptoms Questionnaire (MASQ). *British Journal of Psychology, 46,* 235–239.

Buckby, J. A., Yung, A. R., Cosgrave, E. M., & Killackey, E. J. (2007). Clinical utility of the Mood and Anxiety Symptoms Questionnaire (MASQ) in a sample of young help-seekers. *BMC Psychiatry, 7,* 50.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1017–1025.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach. *Biometrics, 44,* 837–845.

du Fort, G. G., Newman, S. C., & Bland, R. C. (1993). Psychiatric comorbidity and treatment seeking: Sources of selection bias in the study of clinical populations. *Journal of Nervous and Mental Disease, 181,* 467–474.

First, M. B., Spitzer. R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM–IV Axis I Disorders, research version, non-patient edition.* New York, NY: New York State Psychiatric Institute, Biometrics Research.

Flament, M. F., Cohen, D., Choquet, M., Jeammet, P., & Ledoux, S. (2001). Phenomenology, psychosocial correlates, and treatment seeking in major depression and dysthymia of adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 1070–1078.

Fresco, D. M., Mennin, D. S., Heimberg, R. G., & Turk, C. L. (2003). Using the Penn State Worry Questionnaire to identify individuals with generalized anxiety disorder: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 34,* 283–291.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29–36.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

Kendler, K. S. (1995). Is seeking treatment for depression predicted by a history of depression in relatives? Implications for family studies of affective disorder. *Psychological Medicine, 25,* 807–814.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 54,* 395–399.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS–R). *JAMA, 289,* 3095–4105.

Kumar, G., Steer, R. A., Teitelman, K. B., & Villacis, L. (2002). Effectiveness of Beck Depression Inventory—II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment, 9,* 164–170.

Lasa, L., Ayuso-Mateos, J. L., Vázquez-Barquero, J. L., Díez-Manrique, F. J., & Dowrick, C. F. (2000). The use of the Beck Depression Inventory to screen for depression in the general population: A preliminary analysis. *Journal of Affective Disorders, 57,* 261–265.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review, 18,* 765–794.

Luna-Herrera, J., Marínez-Cabrera, G., Parra-Maldonado, R., Enciso-Moreno, J. A., Torres-López, J., Quesada-Pasqual, F., . . . Franzblau, S. G. (2003). Use of receiver operating characteristic curves to assess the performance of a microdilution assay for determination of drug susceptibility of clinical isolates of *Mycobacterium tuberculosis*. *European Journal of Clinical Microbiology & Infectious Diseases, 22,* 21–27.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28,* 487–495.

Nitschke, J. B., Heller, W., Imig, J. C., McDonald, R. P., & Miller, G. A. (2001). Distinguishing dimensions of anxiety and depression. *Cognitive Therapy and Research, 25,* 1–22.

Nitschke, J. B., Heller, W., Palmieri, P. A., & Miller, G. A. (1999). Contrasting patterns of brain activity in anxious apprehension and anxious arousal. *Psychophysiology, 36,* 628–637.

Ormel, J., Rosmalen, J., & Farmer, A. (2004). Neuroticism: A noninformative marker of vulnerability to psychopathology. *Social Psychiatry and Psychiatric Epidemiology, 39,* 906–912.

Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology, 163,* 670–675.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737–748.

Saucier, G. (1998). Replicable item-cluster subcomponents in the NEO Five-Factor Inventory. *Journal of Personality Assessment, 70,* 263–276.

Sloan, D. M., Marx, B. P., Bradley, M. M., Strauss, C. C., Lang, P. J., & Cuthbert, B. C. (2002). Examining the high-end specificity of the Beck Depression Inventory using an anxiety sample. *Cognitive Therapy and Research, 26,* 719–727.

Starr, A. J., Smith, W. R., Frawley, W. H., Borer, D. S., Morgan, S. J., Reinert, C. M., & Mendoza-Welch, M. (2004). Symptoms of posttraumatic stress disorder after orthopaedic trauma. *Journal of Bone and Joint Surgery, 86,* 1115–1121.

Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., & Freeman, B. A. (1979, August 24). Assessment of diagnostic technologies. *Science, 205,* 753–759.

Thuiller, W., Vayreda, J., Pino, J., Sabate, S., Lavorel, S., & Gracia, C. (2003). Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology & Biogeography, 12,* 313–325.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465–490.

Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *Journal of Abnormal Psychology, 104,* 15–25.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology, 104,* 3–14.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and *DSM–III*. *Archives of General Psychiatry, 41,* 1005–1012.

Youden, W. J. (1950). An index for rating diagnostic tests. *Cancer, 3,* 32–35.

Zich, J. M., Attkisson, C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: The CES–D and the BDI. *International Journal of Psychiatry in Medicine, 20,* 259–277.

Received June 1, 2009
Revision received December 15, 2009
Accepted March 15, 2010 ■