Journal of Child Psychology and Psychiatry 47:9 (2006), pp 927–934 doi:10.1111/j.1469-7610.2006.01646.x

Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects

W. Burleson Daviss, Boris Birmaher, Nadine A. Melhem, David A. Axelson, Shana M. Michaels, and David A. Brent

Western Psychiatric Institute & Clinic and the University of Pittsburgh, USA

Background: Previous measures of pediatric depression have shown inconsistent validity in groups with differing demographics, comorbid diagnoses, and clinic or non-clinic origins. The current study re-examines the criterion validity of child- and parent-versions of the Mood and Feelings Questionnaire (MFQ-C, MFQ-P) in a heterogeneous sample of children and adolescents from clinic and non-clinic sources. **Methods:** Among 470 consecutive youth completing semi-structured interviews at a university-based child psychiatry center, total scores from the 33-item MFQ-C and 34-item MFQ-P were examined across subjects with and without mood disorders using analysis of variance, and receiver operating characteristics analysis. **Results:** Mean scores of the MFQ-C and MFQ-P, respectively, differed significantly ($p < .0005$) across youth having major depressive episodes (MDE) (33 and 32, $n = 77$), mood disorders not meeting criteria for current MDE (24 and 28, $n = 75$), and no mood disorders (12 and 10, $n = 318$). In the overall sample, areas under the curve (AUC) for discriminating MDE and any mood disorder, respectively, were .85 and .83 on the MFQ-C, .86 and .90 on the MFQ-P, and .89 and .90 on the MFQ-C and MFQ-P averaged together, suggesting moderate to high criterion validity. Similar findings were noted in subgroups divided by age, sex, race, comorbid psychopathology, and clinic or non-clinic origins. AUCs of these MFQ scores compared favorably with those of the Beck's Depressive Inventory, the Child Behavior Checklist's Anxious/Depressed scale and the Children's Depressive Rating Scale–Revised by the same raters. A score of 29 on the MFQ-C (positive screen rate 21%, sensitivity 68%, specificity 88%) or 27 on the MFQ-P (positive screen rate 23%, sensitivity 61%, specificity 85%) optimally discriminated youth with MDE from the rest of the sample. **Conclusions:** The MFQ-C and MFQ-P, especially used in combination, validly identify MDE or other mood disorders in youth diverse in demographic and clinical characteristics. **Keywords:** Pediatric depression, assessment, validity, reliability. **Abbreviations:** AUC: area under ROC curve; BDI: Beck Depressive Inventory; CDRS-R: Children's Depressive Rating Scale, Revised; CBCL: Child Behavior Checklist; ICC: Intraclass Correlation Coefficient; KSADS-PL: Schedule for Mood Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version; MDE: major depressive episode; MFQ: Mood and Feelings Questionnaire; MFQ-Ave: score averaging child and parent MFQ scores; MFQ-C: child-version of MFQ; MFQ-P: parent-version of MFQ; SE: standard error; ROC: receiver operating characteristics.

Pediatric major depressive episodes (MDE) occur in as many as 2.5% of all children and 8.3% of all adolescents and cause substantial short- and long-term morbidity (Birmaher et al., 1996). Other pediatric depressive disorders not meeting full DSM criteria for a MDE are likewise associated with significant morbidity, under-diagnosed, and may lead eventually to unipolar or bipolar mood disorders of even greater severity (Birmaher et al., 1996). While accurate assessment of depressive disorders is a critical public health concern, such assessment is often complicated by divergent information from parent and child (Birmaher et al., 1996; Cantwell, Lewinsohn, Rohde, & Seeley, 1997), and by high rates of comorbid disorders that share overlapping symptoms with depression, including anxiety, externalizing, and attention deficit hyperactivity disorders (ADHD) (Angold, Costello, & Erkanli, 1999).

In light of these diagnostic challenges, structured or semi-structured interviews of both child and parent are the 'gold standard' for accurately determining the presence of pediatric depressive disorders. These are time-consuming and impractical in most clinic or community settings. Child- and parent-rated questionnaires of depressive symptoms offer a more efficient way to screen for potential cases of depressive disorders, provided that they have sufficient criterion validity, defined as their accuracy discriminating cases of depression as verified by a gold-standard measure such as a semi-structured interview. A recent article reviewed the psychometric properties of many of the most commonly used pediatric depressive measures, and highlighted the continuing need for other questionnaires with

Conflict of interest statement: No conflicts declared.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

broader coverage of depressive symptoms, parallel child and parent versions, and consistent criterion validity across the range of pediatric ages and comorbid disorders (Myers & Winters, 2002).

The Mood and Feelings Questionnaire (MFQ) was developed with many of these needs in mind (Costello & Angold, 1988). The MFQ covers a broad range of cognitive and vegetative symptoms of depression in children and adolescents. Both child-rated (MFQ-C) and parent-rated (MFQ-P) questionnaires are available. The most recent versions of the MFQ-C and MFQ-P have 33 and 34 items, respectively, and ask the child or parent to rate recent depressive symptoms on a Likert scale (0 = not true, 1 = sometimes, 2 = true) with wording simple enough for younger children as well as adolescents. A shorter, 13-item version of the MFQ is also available in child- and parent-versions (Angold, Costello, Messer, & Pickles, 1995).

Two studies have suggested the criterion validity of the MFQ-C or MFQ-P for pediatric depression within specialty child and adolescent psychiatric clinics (Kent, Vostanis, & Feehan, 1997; Wood, Kroll, Moore, & Harrington, 1995). Neither included cases of bipolar disorder or ADHD in their samples, though both reported high rates of comorbid anxiety and externalizing disorders. Neither study reported subgroup analyses of the MFQ's validity in youth having other such psychopathology, or compared the MFQ's validity with other concurrently administered depressive measures.

The purpose of the current paper was to re-examine the criterion validity of the MFQ-C and MFQ-P in a larger, more heterogeneous sample of children and adolescents from both clinic and community settings. The primary objective was to re-examine how well the MFQ-C and MFQ-P discriminated youth with MDE, or any mood disorder, from others without these conditions, both in the overall sample and in various subgroups based on demographic variables, clinic or non-clinic origins, and comorbid diagnoses. Another objective was to compare the criterion validity of the MFQ to other child- and parent-rated measures of pediatric depression administered concurrently.

Methods

Sample

Children and adolescents included in the current sample were consecutive research subjects or clinic patients at least 7 years old assessed at a university-based child and adolescent psychiatry center. To be included, each subject, along with a parent, had completed respective versions of the MFQ as well as a semi-structured interview as part of an intake assessment between June 1, 2001 and October 15, 2004, and had no more than 2 items missing on the MFQ-C or MFQ-P. The sample of 470 subjects included 260 children and adolescents participating in a bipolar offspring study comparing youth with bipolar parents ($n = 150$) to community controls ($n = 110$). An additional 210 subjects were from psychiatric clinics, including adolescents from an ADHD clinic participating in a study examining risk factors of depression ($n = 83$), and other patients assessed at two mood disorder clinics ($n = 127$). Data were included only if the subject and parent had signed an IRB-approved consent form indicating their assent/consent to participate in a study or to have their data included in their clinic's data registry for research purposes.

Diagnostic interview

All subjects and their parents underwent separate, detailed semi-structured interviews to determine mental health diagnoses, using the Schedule for Mood Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997). When youth and parents gave contradictory reports of specific symptoms, interviewers were encouraged to make a best estimate of final diagnoses based on all available information. All interviewers were masters- or doctorate-level research clinicians who had undergone extensive training to assure the reliable and valid administration of this measure. Diagnostic assessments were supervised by board-certified child and adolescent research psychiatrists, and were completed before the MFQ and other rating scales were scored by a computer. Kappas for K-SADS-PL interviews were determined by having all interviewers re-rate other interviewers' recorded assessments in 50 subjects randomly drawn from the four sources of this study. Kappas of relevant diagnostic categories were as follows: major depressive disorder (kappa = .67), any bipolar disorder (kappa = .71), any mood disorder (kappa = .71), ADHD (kappa = .80), any anxiety disorder (kappa = .88) and any externalizing disorder (kappa = .74). All of these kappas were statistically significant ($p < .05$) and suggested acceptable levels of diagnostic accuracy.

Mood and Feelings Questionnaire

All subjects and their parents completed their respective 33-item and 34-item versions of the MFQ at the time of the diagnostic interview, while the other was completing the K-SADS-PL. Cronbach's alphas in the overall sample were high for all items on the MFQ-C (alpha = .95) and MFQ-P (alpha = .96), suggesting high internal consistency. In youths from the ADHD clinic study whose depressive symptoms were re-rated one month after their diagnostic assessment, retest reliability was high ($p < .0001$) for both the MFQ-C (ICC = .80, $n = 63$) and MFQ-P (ICC = .80, $n = 66$).

Other depressive measures

Additional measures completed during the assessment depended on the clinic or study. Ninety-six adolescents at least 13 years old from the mood disorder clinics completed the Beck Depressive Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), a 21-item self-report measure of depression with demonstrated

validity in adolescent samples (Myers & Winters, 2002). A total of 241 parents of children and adolescents in the bipolar offspring study completed the Child Behavior Checklist (CBCL) (Achenbach, 1991), which includes 113 symptoms of psychopathology, rated on a 3-item Likert scale (0 = not true to 2 = often true). The Anxious/Depressed scale for the CBCL was used for this analysis based on its reported validity identifying internalizing disorders and major depression in non-clinic and clinic populations (Achenbach, 1991; Biederman, Monuteaux, Kendrick, Klein, & Faraone, 2005). Raw scores on the CBCL were converted to T-scores using the Assessment Data Manager Program, Version 4.0 (Achenbach System of Empirically Based Assessment; Burlington, VT). A total of 142 adolescents and their parents in the ADHD clinic study and a mood disorder clinic completed the Children's Depressive Rating Scale–Revised (CDRS-R) (Poznanski, Freman, & Mokros, 1985), a 17-item measure with well-documented validity in which the clinician rates depressive symptomatology on a 7-point Likert scale based on symptom-reports by the child and parent as well as the clinician observations (Myers & Winters, 2002).

Statistical analysis

ROC analyses were completed using Intercooled Stata 8.0 for Windows (Stata Corporation; College Station, TX, USA), while other statistical analyses were completed using SPSS-PC Version 11.0 (SPSS, Inc.; Chicago, IL, USA). Scores for missing items on the MFQ were imputed using the average score of the entire sample for that item. Groups with and without mood disorders were compared, using Chi-squared tests for categorical variables, and Mann–Whitey U-tests for individual item responses on questionnaires, and Student *t*-tests or one-way analysis of variance with post hoc Tukey B-tests for continuous variables. Receiver operating characteristics analysis (ROC) curves were used to assess the validity of total scores from the MFQ and other measures in discriminating subjects with and without MDE, or with and without any mood disorder, both in the overall sample and in various subgroups. In ROC, a graph is constructed plotting rates of false positives (1 – specificity) versus true positives (sensitivity) at varying cutpoints of the measure. The area under the curve (AUC) for this ROC graph ranges from .5 to 1.0. An AUC < .7 suggests 'low' diagnostic accuracy, from .7–.9 'moderate' diagnostic accuracy, and ≥.9 'high' diagnostic accuracy (Swets & Pickett, 1982). Optimal cutpoints in the current analysis were selected to maximize Cohen's Kappas of agreement between diagnoses and groups dichotomously categorized by MFQ total scores (Kraemer, 1992). *Z*-scores were calculated to quantify AUC differences between measures obtained within the same subjects (Hanley & McNeil, 1983). All statistical tests were judged significant at *p*-values < .05.

Results

Comparison of groups by mood disorder severity

Among the overall sample of 470 youth, 77 (16.4%) were diagnosed with a current major depressive episode (MDE), including 68 with unipolar MDE and 9 with bipolar MDE. An additional 75 subjects had mood disorders not meeting full DSM-IV criteria for MDE. These included 32 with depressive disorders not otherwise specified (NOS), 14 with dysthymia alone, 12 with major depressive disorder in partial remission, 11 with bipolar disorder NOS, 4 with bipolar disorder I current episode manic, and 2 with cyclothymia. Other diagnoses in the overall sample included 85 with externalizing disorders, either oppositional defiant disorder or conduct disorder, and 135 with ADHD. Sixty-eight youth had at least one anxiety disorder, including separation anxiety, generalized anxiety, social phobia, post-traumatic stress disorder, obsessive compulsive disorder, panic disorder, and acute stress disorder. As expected, there were significantly higher rates in the 210 clinic subjects than in the 260 non-clinic subjects of mood disorders (58.6% versus 11.2%, $\chi^2 = 119.4$, $p < .0005$) and non-mood psychiatric disorders (57.6% versus 31.2%, $\chi^2 = 33.2$, $p < .0005$).

Table 1 compares three groups – youths with no mood disorder, a non-MDE mood disorder, or MDE – regarding demographic variables and rates of other psychiatric diagnoses. Mean ages were significantly older in each successive group. Girls were more frequent in the MDE group, while boys were more frequent in the other two groups. Mean scores on both MFQ-C and MFQ-P were similar comparing boys and girls within the separate groups having non-MDE mood disorders or MDE. There were no significant racial differences across the three groups. Rates of comorbid disorders including ADHD, externalizing disorders, and anxiety disorders differed significantly across groups, with highest rates observed in the non-MDE mood disorder group.

As shown in Table 1, there were significant differences ($p < .0005$) across groups in mean scores on both the MFQ-C and MFQ-P. Post hoc testing showed the MDE group to have the highest mean scores for MFQ-C and MFQ-P, followed by the non-MDE mood disorder group, followed by the group without a mood disorder. Post hoc testing showed that all three groups' mean scores differed significantly ($p < .05$) from each other. Scores from the MFQ-C and MFQ-P in the overall sample correlated strongly with each other ($r = .61$, $p < .0005$).

All items of the MFQ-C and MFQ-P were rated significantly higher ($p < .05$) in the group having MDE than in others not having MDE ($p < .05$), except for item #4 (ate more) on both the MFQ-C and MFQ-P. All items on the MFQ-C and MFQ-P were rated significantly higher ($p < .05$) in the group having any mood disorder than in others without any mood disorder.

In the MDE group, mean MFQ-C or MFQ-P scores were similar comparing 9 youth with and 68 youth without bipolar disorders. Likewise, within subjects having non-MDE mood disorders, mean scores on the MFQ-C or MFQ-P were similar comparing 17

Table 1 Demographic and clinical characteristics across mood disorder subgroups

	No mood disorder ($n = 318$)	Non-MDE mood disorder ($n = 75$)	MDE ($n = 77$)	Test
Demographics				
Age in yrs: Mean (SD)	12.7 (3.1)[a]	13.5 (2.7)[b]	14.7 (2.2)[c]	$F_{2,467} = 15.6$**
Males: *n* (%)	159 (50.4)	41 (54.7)	23 (29.9)	χ^2 (2) = 11.9*
Females: *n* (%)	159 (50.0)	34 (45.3)	54 (70.1)	
White: *n* (%)	260 (81.8)	63 (84.0)	67 (87.0)	χ^2 (2) = 2.8
African-American: *n* (%)	43 (13.5)	7 (9.3)	8 (10.4)	
Biracial: *n* (%)	9 (2.8)	3 (4.0)	1 (1.3)	
Other: *n* (%)	6 (1.9)	2 (2.7)	1 (1.3)	
Other diagnoses: *n* (%)				
ADHD	80 (25.2)	37 (49.3)	18 (23.4)	χ^2 (2) = 18.6**
Externalizing disorder	37 (11.6)	29 (38.7)	19 (24.7)	χ^2 (2) = 32.6**
Anxiety disorder	32 (10.1)	22 (29.3)	14 (18.2)	χ^2 (2) = 19.2**
Total scores: Mean (SD)				
MFQ-C	11.6 (9.9)[a]	24.0 (14.0)[b]	32.8 (13.5)[c]	$F_{2,467} = 126.9$**
MFQ-P	9.5 (10.1)[a]	27.6 (13.3)[b]	32.0 (12.2)[c]	$F_{2,467} = 177.4$**

Externalizing disorder = oppositional defiant or conduct disorders; Anxiety disorder = generalized anxiety, social phobia, panic, separation anxiety, post traumatic stress, or obsessive compulsive disorders. Significance of group comparisons: *$p < .005$, **$p < .0005$. Superscripts indicate group means differing significantly ($p < .05$) in post hoc analysis.

youth with and 58 youth without bipolar mood disorders.

ROC analysis discriminating MDE

In Table 2, results of ROC analysis are summarized for the overall sample and subgroups based on demographic factors, clinic or non-clinic origins, and comorbid diagnoses. As shown, prevalences of MDE varied widely across subgroups. AUCs and standard errors (SE) for discriminating MDE were significant in all analyses (all *p*-values < .01). ROC analyses were repeated using average scores for each subject (MFQ-Ave), calculated by taking the sum of MFQ-C and MFQ-P and dividing by two. In the overall sample, MFQ-Ave scores discriminated MDE significantly better than scores of either the MFQ-C ($z = 2.91$, $p = .004$) or the MFQ-P ($z = 2.71$, $p = .007$) alone. AUCs for these scores were the same or better than for either the MFQ-C or MFQ-P scores alone in most subgroups, except in youth with anxiety disorders, for whom MFQ-C scores were the most accurate discriminating MDE.

A significantly lower AUC on the MFQ-P was noted in youth from clinic as opposed to non-clinic origins ($z = 3.05$, $p = .002$), and in youth with comorbid anxiety as opposed to those without comorbid anxiety disorders ($z = 2.15$, $p = .03$). A significantly lower AUC on the MFQ-C was noted in youth with comorbid ADHD compared to those without comorbid ADHD ($z = 2.00$, $p = .05$). Otherwise, there were no significant differences between AUCs for the MFQ-C, MFQ-P, or MFQ-Ave across subgroups.

ROC analyses were used to determine optimal cutpoints for discriminating MDE in the overall sample. Figures 1A and 1B summarize ROC analyses in the overall sample using total scores for the MFQ-C and MFQ-P, respectively. In these graphs, points moving toward the lower left corner of the graph represent increasing cutpoints and are associated with progressively lower rates of true positives (sensitivity) and false positives (1 – specificity). In the overall sample, optimal cutpoints were 29 for the MFQ-C (positive screen rate: 21%, sensitivity: 68%, specificity: 88%), 27 for the MFQ-P (positive screen rate: 23%, sensitivity: 61%, specificity: 85%), and 32 for the MFQ-Ave (positive screen rate: 15%, sensitivity: 58%, specificity: 94%).

The optimal cutpoints for the overall sample significantly discriminated MDE in all subgroups (*p*-values of Chi-squared tests < .05), with one exception: the MFQ-P cutpoint of 27 did not significantly discriminate MDE among youths with anxiety disorders (Chi-square = 2.51, $p = .113$). Optimal cutpoints of MFQ-C, MFQ-P or MFQ-Ave were also determined for each subgroup, and varied substantially across subgroups, and across raters within the same subgroups. These are available along with their corresponding ROC curves by emailing the lead author.

Additional ROC analyses were conducted including only the 152 having a mood disorder. Relative to the ROC analysis of the overall sample, the AUC for discriminating 77 cases of MDE among those with any mood disorder was lower on the MFQ-C (AUC = .68, SE = .04, $p < .001$). The optimal cutpoint for discriminating subjects with MDE among those with some mood disorder remained 29 (positive screen rate: 50%, sensitivity: 68%, specificity: 68%). MFQ-P scores did not significantly discriminate cases of MDE among subjects with any mood disorder (AUC = .59, SE = .05, $p = .06$), and an optimal cutpoint was not determined.

ROC analysis discriminating any mood disorder

ROC analyses were repeated to determine the criterion validity of the MFQ-C, MFQ-P, and MFQ-Ave, discriminating youth with any mood disorder ($n = 152$) from others without a mood disorder ($n = 318$),

Table 2 ROC analysis of Mood and Feelings Questionnaire

		Major depressive episode			Any mood disorder		
Group	Rater	Prev (%)	AUC	SE	Prev (%)	AUC	SE
Overall $n = 470$	MFQ-C	16.4	.85	.02	32.3	.83	.02
	MFQ-P		.86	.02		.90	.01
	MFQ-Ave		.89	.02		.90	.01
Clinic $n = 210$	MFQ-C	33.8	.82	.03	58.6	.83	.03
	MFQ-P		.75	.03		.81	.03
	MFQ-Ave		.83	.03		.86	.03
Non-clinic $n = 260$	MFQ-C	2.3	*.82*	*.10*	11.2	.82	.05
	MFQ-P		.90	.04		.91	.03
	MFQ-Ave		.90	.06		.90	.03
Boys $n = 223$	MFQ-C	10.3	.82	.05	28.7	.78	.04
	MFQ-P		.88	.03		.89	.02
	MFQ-Ave		.89	.03		.88	.02
Girls $n = 247$	MFQ-C	21.9	.84	.03	35.6	.87	.02
	MFQ-P		.84	.03		.90	.02
	MFQ-Ave		.88	.02		.93	.02
<13 y.o. $n = 212$	MFQ-C	6.6	.85	.06	20.8	.85	.03
	MFQ-P		.90	.03		.93	.02
	MFQ-Ave		.92	.03		.93	.02
≥13 y.o. $n = 258$	MFQ-C	24.4	.86	.03	41.9	.83	.03
	MFQ-P		.84	.02		.88	.02
	MFQ-Ave		.88	.02		.89	.02
Nonwhite $n = 80$	MFQ-C	12.5	.86	.05	27.5	.84	.05
	MFQ-P		.86	.04		.84	.05
	MFQ-Ave		.90	.04		.87	.05
White $n = 390$	MFQ-C	17.2	.85	.03	33.3	.83	.02
	MFQ-P		.86	.02		.91	.01
	MFQ-Ave		.90	.02		.91	.01
ADHD $n = 135$	MFQ-C	13.3	.74	.07	40.7	.74	.04
	MFQ-P		.81	.04		.85	.03
	MFQ-Ave		.82	.05		.84	.03
External $n = 85$	MFQ-C	22.4	.76	.07	56.5	.72	.06
	MFQ-P		.77	.06		.77	.05
	MFQ-Ave		.82	.06		.79	.05
Anxiety $n = 68$	MFQ-C	20.6	.82	.06	52.9	.84	.05
	MFQ-P		*.73*	*.07*		.86	.05
	MFQ-Ave		.81	.06		.88	.04

Prev = prevalence; External = conduct or oppositional defiant disorder; Anxiety = any anxiety disorder. AUCs are significant to $p < .001$, except those in italics which are significant to $p < .01$.

and are also summarized in Table 2. AUCs in the overall sample and various demographic subgroups were in the same approximate ranges as AUCs in analyses discriminating MDE. However, AUCs for both the MFQ-P and MFQ-Ave were in the highly accurate range (AUC > .90) in the overall sample and in many subgroups, while AUCs for the MFQ-C were only in the moderately accurate range (AUCs between .70 and .89). Differences between AUCs of the MFQ-C and the MFQ-P ($z = 3.63$, $p = .0003$) and between MFQ-C and MFQ-Ave ($z = 5.16$, $p < .0001$) were significant in the overall sample. Once again, AUCs for all MFQ scores were lower in all subgroups either from clinic origins or with comorbid diagnoses, but remained significant ($p < .001$) and reached at least moderate accuracy. In the overall sample, optimal cutpoints were lower for discriminating any mood disorder, as follows: MFQ-C ≥ 20 (positive screen rate: 36%, sensitivity: 70%, specificity: 81%), MFQ-P ≥ 21 (positive screen rate: 35%, sensitivity: 77%, specificity: 86%), and MFQ-Ave ≥ 22 (positive screen rate: 31%, sensitivity: 73%, specificity: 89%).

These optimal cutpoints for the overall sample significantly discriminated subjects with any mood disorder in all subgroups (*p*-values for all Chi-squared tests < .05). Once again, optimal cutpoints on the MFQ-C, MFQ-P or MFQ-Ave were also determined for the subgroups and varied substantially across subgroups, and across raters within subgroups. These are also available along with corresponding ROC curves by emailing the lead author.

Validity of MFQ relative to other measures

AUCs of other measures completed by the same raters were compared to those of MFQ-C and MFQ-P regarding their discrimination of MDE. AUCs varied from those of the overall sample because only subsets of subjects also completed these other measures. Relative to the BDI (AUC = .78, SE = .05), the MFQ-C performed equivalently (AUC = .78, SE =

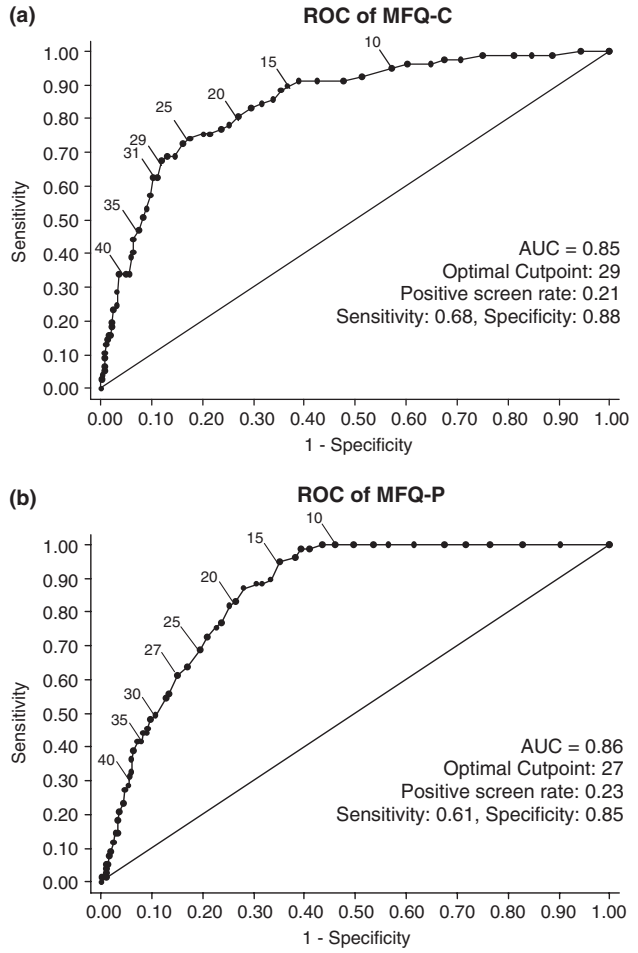

Figure 1 Shown are ROC analysis curves for discriminating MDE in the overall sample (n = 470) using total scores of either the MFQ-C (in 1A) or the MFQ-P (in 1B). Numbers above the curves are various potential cutpoints

.05) ($z = .05$, $p = .96$, $n = 96$). Relative to the CBCL Anxious/Depressed scale (AUC = .70, SE = .13), the MFQ-P performed significantly better (AUC = .90, SE = .05) ($z = 2.06$, $p = .04$, $n = 241$). Relative to the CDRS-R (AUC = .86, SE = .03), which also combines child and parent ratings, the MFQ-Ave also performed equivalently (AUC = .87, SE = .03) ($z = .22$, $p = .83$, $n = 142$).

Discussion

In contrast to previous studies that were exclusively in psychiatric clinics (Kent et al., 1997; Wood et al., 1995), this study examined the criterion validity of the MFQ in a larger, more clinically and demographically diverse sample of youth from both clinic and non-clinic settings. Thirty-two percent of the sample had some mood disorder, and half of those were severe enough to meet DSM criteria for a MDE. Unlike previous studies of the MFQ, some youths in the current sample (about 6%) had bipolar rather than unipolar mood disorders. A large number, especially from the clinic subgroup, had other non-mood psychiatric disorders.

In spite of these sample differences, our findings largely extend findings from these previous studies suggesting the MFQ's validity (Angold et al., 1995; Kent et al., 1997; Wood et al., 1995). Mean scores of the MFQ-C and MFQ-P, respectively, were progressively higher in youth having no mood disorders (12 and 10, $n = 318$), mood disorders not meeting criteria for current MDE (24 and 28, $n = 75$), and unipolar or bipolar MDE (33 and 32, $n = 77$). In the overall sample, areas under the ROC curves for discriminating MDE and any mood disorder, respectively, were .85 and .83 on the MFQ-C, .86 and .90 on the MFQ-P, and .89 and .90 on the MFQ-Ave. These suggest that the MFQ had moderate to high criterion validity for discriminating MDE or any mood disorders in this sample. Similar findings were noted in subgroups determined by age, sex, race, comorbid psychopathology, and clinic or non-clinic origins. AUCs of these MFQ scores compared favorably with those of commonly used measures such as the Beck's Depressive Inventory, the Child Behavior Checklist's Anxious/Depressed scale and the Children's Depressive Rating Scale–Revised completed by the same raters.

This study suggests that the MFQ has potential advantages over other available rating scales of pediatric depression, whose limitations were recently reviewed (Myers & Winters, 2002). First, in our sample, the MFQ-C performed as well in children 7 to 12 years old as it did in adolescents 13 to 19 years old (all AUCs > .85), reflecting the MFQ-C's simple wording and format that make it appropriate across a wide age range. Second, although most depressive measures lack a complementary parent-report with equivalent validity, the MFQ-P here discriminated MDE at levels comparable to the MFQ-C, both in the overall sample and most subgroups, and performed even better discriminating any mood disorder. Third, while the MFQ-C and MFQ-P had somewhat less validity discriminating MDE in subjects with comorbid disorders, as would be expected based on studies of other depressive measures (Myers & Winters, 2002), all three MFQ scores nevertheless demonstrated at least 'moderate' validity in the comorbid subgroups.

A final potential advantage of the MFQ-C and MFQ-P is their parallel structure and content, which allow clinicians to consider potentially divergent child and parent ratings of depressive symptoms simultaneously. In most subgroups, a simple average of the MFQ-C and MFQ-P scores for each subject was more valid than either the MFQ-C or MFQ-P score alone discriminating youth with MDE. Our results thus echo those of a previous study of the MFQ's short-version (Angold et al., 1995), which reported that scores combining child and parent ratings were more valid discriminating mood disorders than those of either measure alone. Given

significant differences in AUCs between MFQ-C and MFQ-P scores in some subgroups, our findings suggest the clinician should weigh child-reports somewhat more strongly when discriminating MDE in a clinic setting, or among children likely to have an anxiety disorder or a mood disorder. Only the MFQ-C achieved a significant AUC discriminating MDE among youth with any mood disorder, consistent with findings by Kent and colleagues (1997) who reported that child ratings on the MFQ are more sensitive to differences in symptom severity between major and minor and depressive disorders. On the other hand, parent ratings on the MFQ may be more helpful discriminating minor mood disorders, or discriminating MDE among youth with ADHD or within non-clinic settings.

Differences between our findings and those of a previous study by Wood and colleagues (1995) may be due in part to sampling and methodological differences. The previous study examined the MFQ-C's and MFQ-P's validity exclusively in a clinic sample of adolescents, determined diagnoses by semi-structured interviews of only youth and not their parents, and reported greater validity in the MFQ-C than the MFQ-P for discriminating MDE. Our study, in contrast, had a broader age range of children and adolescents, included clinic and non-clinic subjects, and determined diagnoses by semi-structured interviews of both children and their parents. Our study found similar criterion validity comparing AUCs for MFQ-C and MFQ-P scores in the overall sample. However, when ROC analyses were restricted to clinic subjects, or subjects with any mood disorder, our study noted greater validity in the child measure than the parent measure, consistent with the previous study.

Another important difference was our finding that higher optimal cutpoints of 29 on the MFQ-C and 27 on the MFQ-P optimally discriminated MDE, compared to the previous study's recommended cutpoints of 27 and 21, respectively (Wood et al., 1995). Optimal cutpoints were selected in the previous study to provide the closest approximations of sensitivity and specificity, and in the current study to maximize Cohen's Kappa. Given the low prevalence of mood disorders in our sample, particularly in the non-clinic subgroup, the method used here would tend to favor higher, more specific cutpoints (Kraemer, 1992). The optimal cutpoints for our overall sample generally proved more specific than sensitive, relative to the previous study's, but performed well discriminating MDE or any mood disorders in various subgroups. These provide a reasonable general starting point for using this measure in clinical or non-clinical populations to dichotomously differentiate youths with or without mood disorders.

However, clinicians using these MFQ-C, MFQ-P, or MFQ-Ave cutpoints should also be aware of their potential limitations (Swets, 1992). First, one should keep in mind the increased likelihood of false positive tests when the rate of a mood disorder is low (as in non-clinic settings), or false negatives when the rate of a disorder is high (as in clinic settings). Second, the optimal cutpoints in our sample varied substantially across individual subgroups, underscoring a key point: that optimal cutpoints, sensitivities, and specificities will also vary across different populations. Third, the most mathematically optimal cutpoint on the MFQ optimizing Kappa may not always be the best from a clinical or research standpoint. Clinicians using the MFQ should weigh the relative costs of false negative and false positive tests, examining the ROC curve to evaluate alternative cutpoints. If the MFQ is being used primarily for screening purposes, and the relative cost of a false negative test outweighs that of a false positive test, one might consider using a lower (more sensitive) cutpoint that lies closer to the upper right-hand corner of the ROC graph. On the other hand, if the test is being used as the final criterion for determining youths to be enrolled in a clinical trial or to receive treatments that carry added risk, a higher (more specific) cutpoint may be more appropriate.

Our findings should be considered cautiously in light of other limitations of this study. First, findings from our study may not extend to other clinic or community populations, as suggested by the variability of AUCs and cutpoints across subgroups. The sample was recruited from tertiary mental health clinics and research samples, and although we included the largest subgroup of non-white youths among psychometric studies to date of the MFQ, ethnic minorities generally were under-represented. Moreover, some of the subjects in the non-clinic subgroup were offspring of bipolar parents, and may have had a higher rate of mental illness than the general population. Second, while our interviewers were blind to total scores on child and parent MFQ, they may have been biased in their diagnostic assessments by their knowledge of whether subjects were being assessed in a clinic or non-clinic setting.

Nevertheless, our study provides new evidence that the MFQ validly discriminates MDE and other mood disorders in a heterogeneous sample, especially if child and parent measures are combined. Our findings suggest that both the MFQ-C and MFQ-P deserve strong consideration as screening measures for pediatric mood disorders in clinic and non-clinic settings. However, findings of somewhat lower validity in the clinic subgroup, and in subjects with comorbid diagnoses, underscore the conventional wisdom that questionnaires like the MFQ are no substitute for a thorough clinical interview of both the child and parent.

Acknowledgements

Support for this project came from a NARSAD Young Investigator's Award (Dr. Daviss), and NIMH grants

including K23 MH065375 (Dr. Daviss), R01 MH060952 (Dr. Birmaher), P30 MH066371 (Dr. Brent), and K23 MH001878 (Dr. Axelson). We thank Rebecca Munnell, Giovanna Porta, Ellie Kanal, Travis Brewer, and Mary Ehmann for help with data entry. We also thank Drs. Jeff Bridge, Satish Iyengar, and Reuben Muchemedzi for advice on statistical analyses.

Correspondence to

W. Burleson Daviss, Department of Psychiatry, University of Texas Health Science Center of San Antonio, 7703 Floyd Care Drive, San Antonio, TX 78229-3900, USA; Tel: 210-567-5484; Fax: 210-567-5677; Email: davissw@uthscsa.edu

References

Achenbach, T.M. (1991). *Manual for Child Behavior Checklist/4–18 and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry.

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, *40*, 57–87.

Angold, A., Costello, E.J., Messer, S.C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, *5*, 237–249.

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.

Biederman J., Monuteaux, M.C., Kendrick, E., Klein, K.L., & Faraone, S.V. (2005). The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. *Archives of Disease in Childhood*, *90*, 1010–1015.

Birmaher, B., Ryan, N.D., Williamson, D.E., Brent, D.A., Kaufman, J., Dahl, R.E., Perel, J.M., & Nelson, B. (1996). Childhood and adolescent depression: A review of the past 10 years, Part I. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*, 1427–1439.

Cantwell, D.P., Lewinsohn, P.M., Rohde, P., & Seeley, J.R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 610–619.

Costello, E.J., & Angold, A. (1988). Scales to assess child and adolescent depression: Checklists, screens, and nets. *Journal of the American Academy of Child and Adolescent Psychiatry*, *27*, 726–737.

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.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). 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.

Kent, L., Vostanis, P., & Feehan, C. (1997). Detection of major and minor depression in children and adolescents: Evaluation of the Mood and Feelings Questionnaire. *Journal of Child Psychology and Psychiatry*, *38*, 565–573.

Kraemer, H.C. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage Publications, Inc.

Myers, K., & Winters, N.C. (2002). Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 634–659.

Poznanski, E., Freman, L., & Mokros, H. (1985). Children's Depression Rating Scale–Revised. *Psychopharmacology Bulletin*, *21*, 979–989.

Swets, J.A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, *47*, 522–532.

Swets, J.A., & Pickett, R.M. (1982). *Evaluation of diagnostic systems*. Orlando, FL: Academic Press.

Wood, A., Kroll, L., Moore, A., & Harrington, R. (1995). Properties of the Mood and Feelings Questionnaire in adolescent psychiatric outpatients: A research note. *Journal of Child Psychology and Psychiatry*, *36*, 327–334.

Manuscript accepted 5 April 2006