IMPROVING DEPRESSIVE SYMPTOM MEASUREMENT IN ADOLESCENTS:
A PSYCHOMETRIC EVALUATION OF THE QUICK INVENTORY OF
DEPRESSIVE SYMPTOMATOLOGY, ADOLESCENT VERSION (QIDS-A17)

APPROVED BY SUPERVISORY COMMITTEE

Beth D. Kennard, Psy.D.

Ira H. Bernstein, Ph.D.

Graham J. Emslie, M.D.

Carroll W. Hughes, Ph.D.

A. John Rush, M.D.

to my family

IMPROVING DEPRESSIVE SYMPTOM MEASUREMENT IN ADOLESCENTS: A PSYCHOMETRIC EVALUATION OF THE QUICK INVENTORY OF DEPRESSIVE SYMPTOMATOLOGY, ADOLESCENT VERSION ($QIDS\text{-}A_{17}$)

by

CHARLOTTE LOUISE HALEY

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2009

ACKNOWLEDGEMENTS

There are a number of people whose guidance and support were instrumental to my successful completion of this dissertation and doctoral program. First, I would like to express my appreciation to the members of my committee whose generous guidance and support allowed me to complete this original research. I am particularly thankful to Dr. Betsy Kennard, my committee chair, mentor and supervisor, for her leadership, balanced guidance, and perseverance to steer this project through to completion. Her skills as a researcher, clinician, and supervisor have left a lasting impression, providing far more than I could have imagined upon beginning my research placement. I am also especially thankful to Dr. A. John Rush, who has been a steady presence in my intended career from the beginning, and without whom this study would not exist. I am grateful for his continued involvement and time investment in spite of his adventurous move to Singapore last year. Dr. Graham Emslie provided guidance and encouragement in creating and carrying out the study, as well as a wonderful research group and work environment. Dr. Ira Bernstein patiently taught me to understand the surface of his depth of knowledge regarding psychometrics, with humor, kindness, and generosity of time. Dr. Carroll Hughes provided a statistical base on which to build, expertise in child and adolescent assessment, and much appreciated support.

Jarrette Moore, Annie Walley, Jessica Jones, Krystle Joyner, Hayley Evans, Jeanne Nightengale-Teresi, and Taryn Mayes provided help in immeasurable ways, including recruitment, strategy and hope. I would not be at this point without their involvement. Additionally, I am grateful to Drs. Paul Croarkin, Rong Rong Tao, and

Kirti Saxena for their willingness to help me reach my recruitment goal in the midst of their demanding schedules.

Dr. Catherine Karni and the wonderful staff of the Children's Medical Outpatient Division of Child and Adolescent Psychiatry generously allowed use of the clinic as well as provided valuable assistance in navigating the clinic routine.

I am also especially grateful to particular supervisors and professors: Drs. Ted Asay, Tim Clark, Rycke Marshall, Sandra Pitts, Richard Robinson, Sunita Stewart, and Frank Trimboli, for providing countless opportunities to learn and grow as a clinician. Additionally, special thanks go to Drs. Melanie Biggs and Cindy Claassen, who helped me begin this process by training me as a research coordinator and providing encouragement and support along the dissertation way.

I am thankful for my classmates, whom I feel lucky to count as dear friends. Their humor, support, and sanity checks provided warmth and light along this journey, as well as bonds that will continue in the years to come. Most of all, I am indebted to my loving, supportive family and friends, with particular thanks to my brothers and cousin for helping turn me in this direction once again..

IMPROVING DEPRESSIVE SYMPTOM MEASUREMENT IN ADOLESCENTS: A PSYCHOMETRIC EVALUATION OF THE QUICK INVENTORY OF DEPRESSIVE SYMPTOMATOLOGY, ADOLESCENT VERSION ($QIDS\text{-}A_{17}$)

CHARLOTTE HALEY

The University of Texas Southwestern Medical Center at Dallas, 2009

Supervising Professor: BETH D. KENNARD, Psy.D.

Major depressive disorder (MDD) in children and adolescents is a common and debilitating psychiatric disorder. Current instruments used to identify the presence of and monitor the treatment of depression in adolescents vary in validity, reliability, appropriateness, cost and ease of administration, such that there is not yet an established instrument that meets all the needs of clinicians working with adolescents. The 16-item Quick Inventory of Depressive Symptomatology ($QIDS_{16}$), developed and successfully validated as an accurate, brief and economical measure of depressive symptom severity in adults, has been modified to an adolescent version ($QIDS\text{-}A_{17}$). Results from recent studies suggest that the $QIDS\text{-}A_{17}$ may meet the need for a freely available, easy to

administer, psychometrically-sound measure of core depressive symptoms for adolescents that can be used both as a screening tool and as a measure of symptom severity in both research studies and clinical practice. The current study aims to validate the QIDS-A_{17} instruments, including the self-report format (QIDS-A-SR), and two clinician-rated formats (QIDS-A-C[Adolescent] and QIDS-A-C[Composite]) in an adolescent outpatient population. The study included 103 outpatient adolescents ranging from 8 to 17 years of age. During a single visit, adolescents completed the QIDS-A-SR. A clinician completed the clinician-rated versions separately for adolescents (QIDS-A-C[Adolescent]) and parents (QIDS-A-C[Composite]) and the Children's Depression Rating Scale-Revised (CDRS-R). Classical Test Theory (CTT) analysis found all three QIDS-A_{17} measures to show strong internal consistency and correlate significantly to the CDRS-R, although the CDRS-R was the most reliable. Factor and parallel analysis found all four measures to be unidimensional. Item Response Theory (IRT) analysis found results that complemented the reliability results found in CTT. All four measures demonstrated diagnostic validity based on univariate and multivariate logistic regression, ANOVA, and MANOVA analyses. Scores on all four measures were equated to create conversion tables to facilitate translation of scores between tests. Although the three clinician-rated measures (CDRS-R, QIDS-A-C[Adolescent], QIDS-A-C[Composite]) were slightly more reliable than the QIDS-A-SR, the QIDS-A-SR demonstrated satisfactory reliability, validity, and discriminate utility such that it can be used effectively in settings that would benefit from a quick, valid, freely available self-report measure of depression in adolescents.

TABLE OF CONTENTS

LIST OF TABLES

APPENDIX C

LIST OF FIGURES

APPENDIX D

APPENDIX E: CATEGORY RESPONSE FUNCTION (CRF) CURVES

APPENDIX F: QIDS-A-C (PARENT) RESULTS

LIST OF ABBREVIATIONS

CGAS – Children's Global Assessment Scale

CGI-I – Clinical Global Impressions—Improvement (scale)

CGI-S – Clinical Global Impressions—Severity (scale)

CDRS-R – Children's Depression Rating Scale-Revised

CMC – Children's Medical Center of Dallas

CTT – Classical Test Theory

DSM-IV – Diagnostic and Statistical Manual of the American Psychiatric Association, 4th edition

DSM-IV-TR – Diagnostic and Statistical Manual of the American Psychiatric Association, 4th edition, Text Revision

HRSD – Hamilton Rating Scale for Depression

$HRSD_{17}$ – Hamilton Rating Scale for Depression, 17-item version

$HRSD_{21}$ – Hamilton Rating Scale for Depression, 21-item version

$HRSD_{24}$ – Hamilton Rating Scale for Depression, 24-item version

ICC – Item Characteristic Curves

IDS_{30} – Inventory for Depressive Symptomatology (30-item)

$IDS\text{-}SR_{30}$ – Inventory for Depressive Symptomatology (30-item, Self-Report format)

$IDS\text{-}C_{30}$ – Inventory for Depressive Symptomatology (30-item, Clinician-Rated format)

IRB – Institutional Review Board

IRT – Item Response Theory

MANOVA – Multivariate Analysis of Variance

MDD – Major Depressive Disorder

MDE – Major Depressive Episode

$QIDS_{16}$ – Quick Inventory of Depressive Symptomatology (16-item)

$QIDS\text{-}SR_{16}$ – Quick Inventory of Depressive Symptomatology (16-item, Self-Report format)

$QIDS\text{-}C_{16}$ – Quick Inventory of Depressive Symptomatology (16-item, Clinician-Rated format)

$QIDS\text{-}A_{17}$ – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item)

$QIDS\text{-}A\text{-}SR_{17}$ – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item, Self-Report format)

$QIDS\text{-}A\text{-}C_{17}$ – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item, Clinician-Rated format)

$QIDS\text{-}A\text{-}C_{17}$(Adolescent) – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item, Clinician-Rated format, Adolescent Response)

$QIDS\text{-}A\text{-}C_{17}$(Composite) – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item, Clinician-Rated format, Composite of Parent and Adolescent Responses)

$QIDS\text{-}A\text{-}C_{17}$(Parent) – Quick Inventory of Depressive Symptomatology, Adolescent Version (17-item, Clinician-Rated format, Parent Response on Adolescent)

ROC – Receiver Operating Curves

TIF – Test Information Function

CHAPTER ONE

INTRODUCTION

Statement of the Problem

Major Depressive Disorder (MDD) in children and adolescents is a common and debilitating psychiatric disorder (Birmaher et al., 2007; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993). Fortunately, clinical studies continue to explore and improve both psychopharmacologic and psychological treatments (Bridge et al., 2007; Hibbs & Jensen, 2004; Kennard, Silva, et al., 2006; March et al., 2004; Treatment for Adolescents With Depression Study Team [TADS], 2003; Weisz, McCarty, & Valeri, 2006). However, for effective treatment, the symptoms of depression must be accurately recognized and monitored, both in research environments and daily clinical settings. Diagnostic tools and instruments that measure symptom severity of depression are essential for successful outcome research and valuable to physicians in facilitating effective patient care. Current instruments commonly used for identifying the presence of and monitoring the treatment of depression in adolescents vary in validity, reliability, appropriateness, and ease of administration (Brooks & Kutcher, 2001). Furthermore, instruments designed for adults are often used in adolescents without the necessary validation in adolescent populations. To date, an economical, valid, reliable, and easily administered tool to diagnose and measure symptom severity in adolescent depression has not been established (Brooks & Kutcher, 2001).

The 16-item Quick Inventory of Depressive Symptomatology ($QIDS_{16}$), developed and successfully validated as an accurate, brief and economical measure of depressive symptom severity in adults, has been modified to an adolescent version ($QIDS\text{-}A_{17}$). Results from recent studies (Bernstein et al., 2008; Moore et al., 2007) suggest that the $QIDS\text{-}A_{17}$ may meet the need for a freely available, easy to administer, psychometrically-sound measure of core depressive symptoms for adolescents that can be used both as a screening tool and as a measure of symptom severity in both research studies and clinical practice. Additional studies, however, are necessary to establish the diagnostic utility and psychometric properties of this instrument.

CHAPTER TWO

LITERATURE REVIEW

Depression in Children and Adolescent Populations

Depression occurs throughout the lifespan, and onset frequently takes place during adolescence (Birmaher et al., 1996; Kessler et al., 1994; Weissman, Bruce, Leaf, Florio, & Holzer, 1991). The peak ages of depressive symptom onset are between 15 to 19 years and 25 to 29 years (Burke, Burke, Regier, & Rae, 1990). In the pediatric age group, depressive disorders are a principal cause of morbidity and mortality (Brent, 1987; Fleming & Offord, 1990; Pfeffer et al., 1991). Eighty-five percent of depressed youth report significant suicidal ideation, and suicide is the third leading cause of death in 10-19 year olds (Heron, 2007). Depression contributes significantly to the increased risk of attempted and completed suicide (Bridge et al., 2007; Gould et al., 1998; Lewinson, Clarke, Seeley, & Rohde, 1994; Reinhertz et al., 1995). The prevalence of depressive disorders in children and adolescents ranges from 0.4 to 8.3% (Burke et al., 1990; Fleming & Offord, 1990; Kashani, Beck, et al., 1987; Kashani, Carlson, et al., 1987; Lewinson, Clarke, et al., 1994; Lewinsohn et al., 1986; Lewinsohn, Hops, et al, 1993; Lewinson, Roberts, et al., 1994; Shaffer et al, 1996), and is greater in adolescents than in children.

Symptoms of MDD disrupt critical developmental processes that occur in adolescence, including social, emotional, cognitive, and even physical development (Cicchetti & Rogosch, 2002; Garber, Keiley, & Martin, 2002; Weisz & Hawley, 2002).

Depression in adolescents often leads to significant functional impairment in school or work, and involvement in the legal system (Kandel & Davies, 1986; Kovacs, Feinberg, Crouse-Novak, Paulauskas, & Finkelstein, 1984; Rao et al., 1995; Rohde, Lewinsohn, & Seeley, 1994). In addition to increased risk of attempted and completed suicide, adolescents with depression are at increased risk for substance abuse and for recurrent depression during adulthood (Brent et al., 1988; Brent et al., 1993; Burke, Burke, Rae, & Regier, 1991; Garrison, Jackson, Marsteller, McKeown, & Addy, 1990; Harrington, Fudge, Rutter, Pickles, & Hill, 1990; Kovacs, Feinberg, Crouse-Novak, Paulauskas, Pollock, et al., 1984; Lewinsohn, Rohde, & Seeley, 1993; Rao et al., 1995; Shaffer et al., 1996). Approximately 70% of children with a single episode of MDD experience another depressive episode within five years (Kovacs et al, 1984). In addition, many depressed youth also have lifetime and concurrent comorbid anxiety disorders and other psychiatric comorbidities such as substance abuse, disruptive behavior disorders, and attention deficit disorder (Hatcher-Kay & King, 2003; Kessler, Avenevoli, & Ries, 2001).

Given the prevalence and profound impact of depression on adolescent functioning and long-term prognosis (Fergusson & Woodward, 2002; Rao et al., 1995; Weissman et al., 1999), an easily used tool to improve the identification of depression and to improve the quality of treatment for depressed adolescents has obvious public health benefits (Coyle et al., 2003; Olfson, Shaffer, Marcus, & Greenberg, 2003). In addition, one instrument that can follow children and adolescents into adulthood using the same measurement system would allow for better continuity of care as well as better methods for longitudinal studies, thus improving understanding of the progression of depression through the lifetime.

Current State of Instruments Used to Measure Depression in Adolescents

Types of Assessment: Self-rated, Parent-rated, Clinician-rated

Measurement instruments used through the years to assess depressive symptoms in adolescents fall into two main categories: self-report and observer-rated scales (Brooks & Kutcher, 2001). Of the observer-rated scales for depressive symptoms, parent-rated and clinician-rated scales are most commonly used, as opposed to teacher-rated scales used more for the school environment (Myers & Winters, 2002b). Each type of rating scale has its strengths as well as limitations, as will be summarized below.

Patient Self Report

The majority of rating scales used to measure depression are of the self-rated report type (Myers & Winters, 2002b). Important strengths of the self-report measure include ease and speed of administration, as well as cost savings from reduced labor. In addition, self-report measures are seen by many as the best way to evaluate depressive symptom severity in adolescents and adults, due to the "internalizing" nature of depression (Braaten et al., 2001; Myers & Winters, 2002b; Reynolds & Mazza, 1998). However, some studies question the ability of children and adolescents to accurately report symptoms and, as a result, produce a valid rating (Braaten et al., 2001; Cantwell, Lewinsohn, Rohde, & Seeley, 1997; McConaughy & Achenbach, 1989). Literature generally supports this concern in children but suggests that adolescents are largely competent in reporting their symptoms, as long as the scale is developmentally

appropriate (Cantwell et al., 1997; Myers & Winters, 2002a, 2002b; Reynolds & Mazza, 1998).

Parent and Clinician Report

Given the questionable validity of child and adolescent self-reports, there has been a long-standing consensus that clinicians should use additional sources, either independently or in addition to the child or adolescent's report, to inform clinical assessment of psychopathology (Cantwell et al., 1997). As such, some instruments use parent-rated scales or clinician-rated scales. Depending on the measure and the observer, some strengths of the observer-rated measure type include greater reliability, validity, and sensitivity to change (Brooks & Kutcher, 2001)

Parent-rating scales offer a unique perspective on the child since they can provide wide-ranging information about the child from observations across time and multiple situations (Myers & Winters, 2002a, 2002b). However, the rates of agreement between parent and child ratings are generally low (Herjanic & Reich, 1982; Welner, Reich, Herjanic, Jung, & Amado, 1987). This discrepancy is considered to be due to a variety of reasons, including the child's development and the type of symptom assessed. For example, parent ratings are considered particularly important for younger children, when the child is developmentally unable to provide a reliable report of their psychological experiences. In general, as the child's age increases, so does the agreement between parent and child ratings in general (Achenbach, McConaughy, & Howell, 1987; Myers & Winters, 2002a, 2002b).

Regarding types of symptoms, parent reports appear to be especially helpful when assessing externalizing disorders, as diagnostic behaviors are clear and more easily recognized by a parent verses the child (Achenbach et al., 1987; Cantwell et al., 1997). For depressive disorders, however, where many of the diagnostic symptoms are internal and not readily observable, parent ratings generally appear to be less accurate than the youth's self-report (Cantwell et al., 1997; Welner et al., 1997; Yule, 1993; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). The general consensus regarding parental reports is that they should be used in addition to child self-reports to provide important collateral information (Myers & Winters, 2002a, 2002b).

Clinician-rated scales offer an interview format during which clinicians can draw upon their experience and expertise to garner the necessary information for rating symptoms. Clinician-rated scales are given in interview formats that vary in structure. Agreement between clinician-rated and child self-rated instruments is varied, with some studies finding high concordance (Shain, Naylor, & Alessi, 1990), and others low concordance (Dorz, Borgherini, Conforti, Scarso, & Magni, 2004; Shemesh et al., 2005). To overcome these discrepancies between self-rated and observer-rated instruments, some scales have combined both parent and child reports into a semi-structured clinician rated scale, which takes into account the parent and child reports and creates a consensus rating. This format appears to be the most reliable type of rating scale to date with children and adolescents, although improvement is still needed (Cantwell et al., 1997; Myers & Winters, 2002a, 2002b).

Other Limitations

In addition to the various limitations of the above types of scales, further restrictions exist among instruments currently used to assess adolescent depression. Despite the consensus that it is often best to use multiple sources for accurate assessment of child psychopathology, most instruments do not provide corresponding clinician and parent reports in addition to the child self reports (Brooks & Kutcher, 2001; Myers & Winters, 2002a, 2002b). This not only limits the ability to gather additional comparative data, but it prevents detailed psychometric analyses to explore the performance of the self report in comparison to other formats.

Moreover, current adolescent measures lack a corresponding measure that is applicable for all age groups to allow for continuity of measurement and longitudinal study from adolescence into adulthood. Some adult measures have been used in children and adolescents, but often without the necessary validation in child/adolescent populations to ensure that the measures are applicable (Brooks & Kutcher, 2001). The validation of a measure for use in adolescents and adults would allow for important comparisons across age groups (Brooks & Kutcher, 2001).

Finally, while various adolescent measures are sound in some areas, they are weak in the other important areas. Some instruments measure a clear construct with generally good psychometric properties, but have weak discriminant validity and are not sensitive to change. Others may be sensitive to change and psychometrically sound but are too time consuming for realistic clinical use (Brooks & Kutcher, 2001; Myers & Winters, 2002b). A summary of current measures of depression commonly used in adolescents is shown in Table 1.

Summary

Although several clinician measures provide adequate measurement of depressive symptoms in adolescents, many limitations exist. There is yet to be an instrument that provides all the necessary qualities of measurement, including ease and speed of administration, full symptom coverage of adolescent depression, and good reliability, validity, internal consistency, and sensitivity to change. Of the currently used instruments, the Children's Depression Rating Scale-Revised (CDRS-R; Poznanski, Cook, & Carroll, 1979; Poznanski, Freeman, & Mokros, 1985; Poznanski & Mokros, 1996) is the present field standard (Cheung, Emslie, & Mayes, 2005) and generally shows good psychometric properties (reliability, validity, and sensitivity to change; Brooks & Kutcher, 2001; Myers & Winters, 2002b). However, the CDRS-R has limitations, including poor inter-informant reliability and lack of precision in how to weight the data from various informants to create the consensus score. In addition, it is time consuming, requires training time, lacks a corresponding self-report, and is not in the public domain. As such, use is often limited to a clinical research setting (TADS, 2003) (see next section, Psychometric Properties of the CDRS-R, for further information).

Psychometric Properties of the CDRS-R

The clinician-rated Children's Depression Rating Scale – Revised (CDRS-R) was developed to measure the severity of depression in children aged 6-12 years, although it has also been used with adolescents. It was originally patterned after the Hamilton

Depression Rating Scale (HRSD; Hamilton, 1960) and subsequently revised to the integrated format used today (Poznanski et al., 1985). The CDRS-R has 17 items, and the score is based on a composite score obtained by the clinician synthesizing responses obtained in clinical interviews with the parent and adolescent. Each item is rated on a 1 to 5 or 1 to 7 point scale, with a "1" describing the absence of the given symptom. The total raw score ranges from 17 to 113. The T-score ranges from 30 to 100. A T-score of 65 to 74 suggests that a Major Depressive Episode (MDE) may be present (Poznanski & Mokros, 1996).

Studies over the years have established the reliability, validity, and sensitivity to change in the CDRS-R (Brooks & Kutcher, 2001). The CDRS-R has shown adequate internal consistency, ranging from moderate (Cronbach's coefficient alpha (α) = 0.70) (Guo et al., 2006) to high internal consistency (α = 0.85) (Poznanski & Mokros, 1995). The scale has demonstrated good test-retest reliability (r = 0.86) and good inter-rater reliability (r = 0.92) (Poznanski et al., 1985). Moderate to good concurrent validity has been found by correlations with a variety of rating scales, including the HRSD (r = 0.92) (Myers & Winters, 2002b; Shain et al., 1990). However, poor levels of inter-informant reliability have been found across several studies, where child and parent total score correlation was 0.38 (Mokros, Poznanski, Grossman, & Freeman, 1987). While the CDRS-R is used in clinical studies to assess change in severity of depression (Emslie et al., 1997), other studies suggest that the CDRS-R may not be as sensitive for pre-adolescents or adolescents (Stark, Reynolds, & Kaslow, 1987).

Regarding the test structure, not all of the CDRS-R items use the same rating scale. Some items are rated on a 7-point scale (e.g., difficulty having fun, social

withdrawal) while others are rated on a 5-point scale (appetite disturbance, sleep disturbance, and listless speech), thus providing differently weighted items. In addition, the CDRS-R does not cover all nine core criterion symptoms for depression, as it leaves out weight change, hypersomnia, and concentration. The scale does rate school performance, which may in part address concentration, but also creates confusion in how to rate adolescents who are not currently in school. Furthermore, three of the items (depressed facial effect, listless speech, hypoactivity) require direct observation in an interview.

Another limitation is that the CDRS-R lacks a precise definition of remission and, instead, defines remission as a binary term (i.e., yes or no) using a cut-off score (Kennard, Silva, et al., 2006; Kennard et al., 2009). In clinical treatment studies, it is important to have precise measures of depression severity to know when remission has been reached, since remission is an essential goal of treatment (American Psychiatric Association, 2000b; Keller, 2004; Rush & Trivedi, 1995). For adolescent depression measures, there has yet to be an established, empirically validated definition of remission based on total rating. However, recent studies suggest a CDRS-R total score of 28 as a possible indicator of remission when combined with a Clinical Global Impression Scale (CGI-I) score of 2 or less (Emslie et al., 2002; Emslie et al., 1997; Kennard, Emslie, et al., 2006; Kennard et al., 2008; Wagner et al., 2004).

The total CDRS-R score includes ratings of both symptoms and function (e.g., poor school performance or associated impaired relationships with peers, teachers, or family members; social withdrawal and continued social isolation or conflicts with peers or family members; and impact on fun activities or withdrawal from activities). This can

create a potential confound in measuring symptom remission (Rush, Kraemer, et al., 2006). As a result, the CDRS-R may be less sensitive to detecting symptom remission, since functional difficulties rated on the CDRS-R may have been present prior to the onset of the MDE or, as in adults, these problems may take longer to resolve than depressive symptoms (Mintz, Mintz, Arruda, & Hwang, 1992).

Another limitation is that the CDRS-R lacks a self-report version. Although the CDRS-R incorporates assessment of the child, self-reports may be especially useful in adolescents. Since depression is an internalizing disorder, important symptoms may not come to parental attention until obvious behavioral problems occur (Herjanic & Reich, 1982; Wu et al., 1999). Self-report instruments designed specifically for adolescents may provide an opportunity for earlier detection and assessment of such problems, as they may reveal important information about symptoms that might otherwise go unnoticed (Flanery, 1990; Kazdin, 1989).

Finally, the CDRS-R has not been subjected to IRT analyses, and it is not in the public domain, limiting its use in routine practice. In addition, recent analyses (Bernstein et al., 2008; Guo et al., 2006; Jain et al., 2007) suggest that the CDRS-R is multidimensional, where unidimensional scales are recommended as they are more sensitive to change than multidimensional scales (Gibbons, Clark, & Kupfer, 1993).

Psychometric Properties of the IDS_{30}. $QIDS_{16}$, and QIDS-A_{17}

The Inventory of Depressive Symptomatology (IDS_{30})

The Inventory of Depressive Symptomatology (IDS_{30}) was developed to address the limitations of previously developed patient and clinician rating scales of depression (Rush et al., 1986; Rush et al., 1996; Trivedi et al., 2004). The IDS_{30} contains 30 items, of which 16 assess the nine core symptom domains required to diagnose a major depressive episode. The additional items assess common symptoms associated with depression (e.g., anxious mood, irritable mood, sympathetic nervous system arousal), as well as melancholic symptom features, as defined by the DSM-IV (e.g., unreactive mood, distinct quality to mood), and atypical symptom features (e.g., leaden paralysis, interpersonal rejection sensitivity). The IDS_{30} is available in clinician-rated (IDS-C_{30}) and self-report (IDS-SR_{30}) formats. High correspondence has been found between IDS-C_{30} and IDS-SR_{30} total scores (Rush, Gullion, Basco, Jarrett, & Trivedi, 1996) and items (Trivedi et al., 2004). The IDS-SR_{30} has been found to produce results in outcome studies in adults comparable to the 17- and 24-item versions of the Hamilton Depression Rating Scale (Rush et al., 2005). High internal consistency (alpha = 0.90-0.92) for both measures has been found (Trivedi et al., 2004). In addition, the IDS-C_{30} and IDS-SR_{30} are in the public domain, and various translations into many languages are available at no cost (www.ids-qids.org). The IDS_{30} was not chosen for application in adolescents because the performance of the $QIDS_{16}$ is highly satisfactory (see section below), faster and easier to use, and more easily adapted to adolescents.

The Quick Inventory of Depressive Symptomatology ($QIDS_{16}$)

The 16-item QIDS, a shorter version of the IDS_{30}, was developed to be more time-efficient in both research and clinical settings (Rush et al., 2000). While the IDS_{30} takes 15 to 20 minutes to complete, the $QIDS_{16}$ takes 5 to 7 minutes. The $QIDS_{16}$ was created in clinician (QIDS-C_{16}) and self-report (QIDS-SR_{16}) formats. The $QIDS_{16}$ is also available in an automated, interactive voice response (IVR) telephone system (QIDS-IVR_{16}) (Rush, Bernstein, et al., 2006). The $QIDS_{16}$ contains 16 items initially extracted from the IDS_{30} that measure the nine criterion symptom domains to establish the diagnosis and to measure the severity of a major depressive episode based on DSM-IV-TR. The nine DSM-IV-TR symptom domains addressed on the $QIDS_{16}$ include sad mood, concentration, self-criticism, suicidal ideation, interest, energy/fatigue, psychomotor agitation/retardation, change in appetite/weight, and sleep disturbances (including initial, middle, or late insomnia and hypersomnia). The $QIDS_{16}$ and IDS_{30} have anchor points that specify the severity and frequency of symptoms and provide equivalent weightings for each symptom using a 0 to 3 value of intensity (Gullion & Rush, 1998; Trivedi et al., 2004). For the sixteen-item QIDS, the total test score ranges from 0-27.

The QIDS-C_{16} and QIDS-SR_{16} have high internal validity (Rush, Bernstein, et al., 2006; Rush et al., 2003) with coefficients alpha ranging between .86 and .87. In addition, Rush, Bernstein, and colleagues (2006) found correlations ranging between .86 and .93 between the $HRSD_{17}$ and each of the $QIDS_{16}$ scales. Concurrent validity has also been established between the QIDS-SR_{16} and the $HRSD_{17}$, $HRSD_{21}$, and $HRSD_{24}$ in adults (Rush et al., 2005). Item total correlation for the $HRSD_{17}$ and the QIDS-SR_{16} and QIDS-

C_{16} found that more $QIDS_{16}$ items correlated more highly with the $QIDS_{16}$ total score than was the case for the $HRSD_{17}$ items. Specifically, several $HRSD_{17}$ items contribute minimally to the total score (e.g., suicide, insight, weight loss). For the $QIDS_{16}$, each of the nine domains contributes substantially to the total score. IRT analyses in adults revealed very similar psychometric properties, indicating that the QIDS-SR_{16} performs analogously and comparably to the clinician-rated QIDS-C_{16} (Bernstein et al., 2007).

In determining response and remission, high correspondence was found between the QIDS-SR_{16} and QIDS-C_{16} with evidence that the self-report alone is an adequate measure of each of these outcomes in adults. The adult $QIDS_{16}$ were found to be unifactorial (unidimensional), which supports use of the $QIDS_{16}$ as a tool by which to measure treatment response in both clinical and research settings (Bernstein et al., 2007; Rush, Bernstein, et al., 2006; Rush et al., 2005; Trivedi et al., 2004).

In summary, the $QIDS_{16}$ for adults demonstrates excellent correspondence and psychometric properties. Furthermore, item response theory (IRT) analyses and classical test theory (CTT) have established item performance in a wide variety of adult populations (Bernstein et al, 2007; Brown et al., 2008; Rush, Bernstein, et al., 2006; Rush, Carmody, et al., 2006; Rush et al, 2005). To date, work with adults has indicated that the self-report version (QIDS-SR_{16}) can replace the more time-consuming clinician rated QIDS-C_{16} (Bernstein et al., 2007; Rush, Bernstein, et al., 2006).

Current Psychometric Study with the Adult $QIDS_{16}$ in Adolescents

Bernstein et al. (2008) recently completed a study using the adult QIDS-SR_{16} and QIDS-C_{16} ratings with 140 adolescent outpatients (ages 12-17) seen at the Child and

Adolescent Psychiatric Clinic at UT Southwestern Medical Center at Dallas and Children's Medical Center of Dallas. These subjects were assessed only once with the following measures: the QIDS-SR_{16} was completed by the adolescent, and a trained clinical evaluator interviewed the adolescent and parent/guardian separately to complete the QIDS-C_{16}(Adol), the QIDS-C_{16}(Par), the composite QIDS-C_{16}(Comp), and the composite CDRS-R. The composite QIDS-C_{16}(Comp) was completed by choosing the most pathological (highest) of the adolescent and parent responses (the QIDS-C_{16}(Adol) and the QIDS-C_{16}(Par), respectively).

Results indicated that the QIDS-C_{16} (all versions) and the QIDS-SR_{16} ratings were unifactorial (unidimensional) in adolescents. The CDRS-R was found to be at least two-dimensional in this population. The reliability of the QIDS-C_{16}(Adol) and QIDS-SR_{16} was 0.80 and 0.86, respectively. The reliability of the QIDS-C_{16}(Comp) measure was 0.77, while the reliability of the CDRS-R Composite was 0.87. The QIDS-C_{16}(Par) was least reliable among the measures (0.71).

The QIDS-C_{16}(Adol) and QIDS-SR_{16} were highly correlated (r= 0.81), while they correlated 0.79 and 0.68 with the CDRS-R, respectively. The QIDS-C_{16}(Comp) correlated highly with the CDRS-R (r = 0.82). Both the QIDS-C_{16}(Adol) and the QIDS-C_{16}(Par) were highly correlated with the QIDS-C_{16}(Comp) (r = 0.85), which was expected as these measures were used to generate the QIDS-C_{16}(Comp) scores. The QIDS-C_{16}(Par) showed low correlations with the QIDS-C_{16}(Adol) (r = 0.55), QIDS-SR_{16} (r = 0.49), and the CDRS (r = 0.69).

Results indicated that the QIDS-SR_{16} was slightly more reliable than the other versions of the $QIDS_{16}$. As noted previously, Rush, Bernstein, et al. (2006) reported a

very high degree of similarity between the QIDS-C_{16} and QIDS-SR_{16}. Taken together, these results provide a downward age extension of the evidence previously presented for the various adult versions of the $QIDS_{16}$ (Bernstein et al., 2007; Bernstein et al., 2008; Rush, Bernstein et al., 2006).

The Quick Inventory of Depressive Symptomatology, Adolescent Version (QIDS-A_{17})

The Quick Inventory of Depressive Symptomatology, Adolescent Version (QIDS-A_{17}) has been adapted from the $QIDS_{16}$ for use with adolescents, including self-report (QIDS-A_{17}-SR) and clinician-rated (QIDS-A_{17}-C) formats, which can be completed by either the adolescent or the parent. A composite version, QIDS-A_{17}-C(Composite) can be completed by the clinician when both the parent and child provide individual reports. The composite is created by using the clinician's best estimate of the most valid response from either respondent. To adapt the $QIDS_{16}$ to adolescents, developmentally appropriate probes were included to place each question in context for either the parent or child respondents. In addition, an irritability item was added to assess both sadness and irritability for the mood domain using DSM-IV-TR criteria for adolescent depression. The scoring of all versions of the 17-item QIDS-A is identical to scoring the adult version, and the items cover the same nine DSM-IV-TR symptom domains, as described above. The total score range on the 17-item QIDS-A is identical to the 16-item adult QIDS, since the highest score of the two mood items (sad or irritable) is used to rate the mood domain, while one item alone (sad mood) is used for the adult $QIDS_{16}$. The language level, which is estimated to be at a 4^{th} or 5^{th}-grade level, allows use

of the QIDS-A with patients 12 years of age or older independently, or with younger patients with clinician assistance.

Pilot Study of an Electronic Version of the QIDS-A_{17} (QIDS-A-IVR)

Interactive, voice response (IVR) measures, although increasingly used and explored (Kobak, Greist, Jefferson, Mundt, & Katzelnick, 1999; Mundt, 1997; Rush, Bernstein, et al., 2006), are beyond the scope of this study. However, it should be noted that Moore and colleagues (Moore et al.., 2007) evaluated the psychometric properties of an electronic, speech enabled version of the QIDS-A_{17} (QIDS-A-IVR) in a pilot study of twenty-seven pairs of adolescents (aged 12 through 17) and caregivers, recruited from a larger study investigating depression measures in a medical setting. The adolescents and parents separately completed the QIDS-A-IVR in a private room via telephone. In addition, a clinician interviewed the adolescents and parents separately to complete the $QIDS_{16}$-C-A (administered to the adolescent), the $QIDS_{16}$-C-P (administered to the parent), and the CDRS-R. Results found the QIDS-A-IVR to be reliable in this sample, with Cronbach's α of .85. Furthermore, the QIDS-A-IVR correlated significantly with the QIDS-C-A and the CDRS-R ($r = 0.95$ and 0.76, respectively). Although further research is needed in a larger, more diverse sample, these results suggest the validity and reliability of the QIDS-A-IVR as a measure of depression in adolescence.

CHAPTER THREE

RATIONALE, AIMS, AND HYPOTHESES

Rationale and Aims

The aim of this study is to validate an accurate, simple, and efficient measure for adolescents that is practical and easy to use in clinical settings with actual patients and clinicians. The 16-item Quick Inventory of Depressive Symptomatology ($QIDS_{16}$), an adult measure of depressive symptom severity, has been successfully developed and validated (Rush, Bernstein, et al., 2006; Rush et. al., 2000; Rush et al., 2003; Trivedi et al., 2004). In addition, the adult $QIDS_{16}$ has been modified for adolescents, the QIDS-A_{17}, using developmentally appropriate probes and by adding an irritability item to assess all nine DSM-IV criterion symptoms for adolescent major depressive episode (American Psychiatric Association, 2000a). The QIDS-A_{17} formats include self-report (QIDS-A_{17}-SR), clinician-rated (QIDS-A_{17}-C) for use with adolescents and parents separately, and a composite score (QIDS-A_{17}-C[Composite]), that is generated when both parents and adolescents provide data to the clinician. In addition, a self-report is available for the parent/guardian to complete regarding the adolescent (QIDS-A_{17}-SR [P]). However, this measure was not used for the purposes of this study due to the previously mentioned limitations of utilizing parent-only reports for internalizing disorders. Similarly, the clinician-rated version based on the parent responses (QIDS-A-C[Parent]) will only be used in this study to generate the QIDS-A-C(Composite) and will not be included individually in the main analyses. An abbreviated summary of statistical results

regarding the QIDS-A-C(Parent) will be included in Appendix F. As noted previously in Psychometric Properties of the CDRS-R, the CDRS-R item scores are created by integrating responses from both the child and the parent. Only the integrated CDRS-R composite scores will be used for the analyses in this study. Thus, when the CDRS-R is mentioned in this study, this will refer to the composite CDRS-R.

The adult $QIDS_{16}$ is a universal, freely available, easy to use, psychometrically-sound measure of core depressive symptoms for all age groups that can be used both as a screening tool and as a measure of symptom severity in both research studies and clinical practice. If validated, the QIDS-A_{17} would serve as the adolescent equivalent, benefiting clinicians and patients, and contributing to public health. The present study was designed to begin formal testing of the QIDS-A_{17} on a well-defined population, using a currently used, well-validated, measure of childhood depression for comparison. The specific aims are as follows:

Aim I: To define the psychometric properties of the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), and QIDS-A_{17}-C(Composite).

Aim II: To define the thresholds, based on total scores on the CDRS-R, the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), and the QIDS-A_{17}-C(Composite), for ascertaining the probable presence of a major depressive episode (MDE) in a clinical sample as determined by a diagnostic checklist for MDE.

Aim III: To determine whether the adolescent self report (QIDS-A_{17}-SR) or clinical interview with the adolescent alone (QIDS-A_{17}-C[Adolescent]) are sufficient to replace the more time consuming QIDS-A_{17}-C(Composite) or the field standard (CDRS-R).

Aim IV: To provide conversion tables by which to translate total scores on the QIDS-A_{17}-SR to total scores on the QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R.

The present study provides data by which to compare the psychometric properties of each measurement approach using classical test theory (CTT) and item response theory (IRT) analyses. As a result, it will be determined whether the present "field standard", the CDRS-R, based on composite scores by trained clinician interviews with adolescents and parents, can be replaced by the QIDS-A_{17}-SR, using the adolescent responses alone to a self report, or to a clinical interview with the adolescent alone (QIDS-A_{17}-C[Adolescent]).

Questions and Hypotheses

Research Question One: What are the psychometric properties of the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and CDRS-R in an adolescent outpatient population, including reliability measures and construct and concurrent validity measures?

Hypothesis One (A): The QIDS-A_{17}-SR will demonstrate good internal consistency, reliability, and related psychometric properties.

Hypothesis One (B): The QIDS-A_{17}-C(Adolescent) will demonstrate good internal consistency, reliability, and related psychometric properties.

Hypothesis One (C): The QIDS-A_{17}-C(Composite) will demonstrate good internal consistency, reliability, and related psychometric properties.

Hypothesis One (D): The CDRS-R will demonstrate good internal consistency, reliability, and related psychometric properties.

Research Question Two: What is the relationship between the total scores on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R?

Hypothesis Two: The QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite), and the CDRS-R will correlate highly.

Research Question Three: What is the relationship between the scores on the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite), and the CDRS-R, and the probable presence of a major depressive episode (MDE) as determined by the MDE checklist?

Hypothesis Three (A): The probable presence or absence of MDE, as determined by the MDE checklist, will be related to the scores on the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite) and the CDRS-R to identify relevant scoring thresholds by which to identify MDE.

Hypothesis Three (B): Higher scores on the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite), and the CDRS-R will relate to an increased probability of having the presence of MDE.

Hypothesis Three (C): Higher scores on the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite), and the CDRS-R will be uniquely correlated to an increased probability of having the presence of MDE.

Research Question Four: What are the equivalency scores between the individual scores on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R?

Hypothesis Four: The relationship between individual items on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R will be established with the creation of inter-test conversion tables.

CHAPTER FOUR

METHODOLOGY

Subjects

Participants for this study were children and adolescents recruited from the Division of Child Psychiatry at UT Southwestern and the Child and Adolescent Psychiatry Outpatient Clinic at Children's Medical Center of Dallas. The data for this study also includes a limited data set incorporated from the ongoing, NIMH-funded study, Pediatric MDD: Sequential Treatment with Fluoxetine and Relapse Prevention CBT (1RO1MH-39188; Emslie & Kennard, principle investigators, 2008).

Inclusion Criteria

Study participants included outpatients between the ages of 8 and 17 years of age who were still attending school. Adolescents who had left school were not included, as school functioning was a major assessment area in this age group and an item on the CDRS-R severity scale. Participants had no restrictions regarding medications or other treatment(s) received outside of this study. Subjects may have had any concurrent general medical condition(s) or Axis I disorder(s) except as noted below in exclusion.

Exclusion Criteria

Patients were excluded from the study if they had concurrent mental retardation, active psychosis, terminal illnesses, or neurological disorders that precluded participation

from completing study questionnaires. Subjects with concurrent acute substance/alcohol intoxication, judged clinically, were excluded. Patients who were delayed more than two years from age-appropriate grade level and patients with Dyslexia/Reading Disorder were excluded if they were unable to clearly understand and complete the self-report instruments without assistance. Such cases were determined via medical chart review as well as guidance from the patient's clinic doctor. Patients unable to speak and read English were excluded, as the primary self-report and parent-report scales required for data collection do not have norms for non-English translations.

Informed Consent

Appropriate approval was obtained from the UT Southwestern Medical Center Institutional Review Board (IRB). The process of informed consent was conducted with participants and their caregiver(s) prior to the collection of any data and included an explanation regarding the purpose, procedures, possible risks and benefits, confidentiality related to the study, and their rights as patients. Potential participants were informed of the alternatives to participation in this study and were given the opportunity to ask questions. Participants were informed that they would receive monetary compensation upon completion of the study. Patients and their caregivers then signed written informed assent and consent (See Appendix A). They also signed the HIPAA Authorization for Use and Disclosure of Protected Health Information form prior to participation in the study. Copies of the signed consent form and the HIPAA Authorization form were provided to the caregivers and also placed in the medical chart.

Procedures

Participants were recruited between February and December 2008 from outpatients who were receiving treatment as usual or participating in other research studies at the Division of Child Psychiatry at UT Southwestern and the Child and Adolescent Psychiatry Outpatient Clinic at Children's Medical Center of Dallas. The daily clinic appointment scheduling sheet and medical records were used to prescreen patients for eligibility using age and diagnoses. Eligible outpatients and their caregivers were asked by their physician or another member of their treatment team if they were interested in speaking to the study coordinator for more information about the study. Those who were interested either remained after their clinic appointment for the informed consent process and completion of all study measures or set an appointment with the study coordinator to return for participation at a later date. Participants were also recruited via IRB-approved flyers available in the clinic.

After informed consent was obtained, a trained clinician interviewed the patient and caregiver separately. The caregiver was asked to complete a demographic questionnaire while the patient was interviewed. During the patient interview, the patient was first asked to complete the QIDS-A_{17}-SR independently. If the patient had difficulty completing the measure independently due to age or grade-appropriate reading ability, the clinician assisted the patient by reading the items verbatim for the patient to answer. No additional assistance or guidance was given for the self report. Following completion of the self report, the clinician interviewed the patient to complete the adolescent portion of the clinician version of the QIDS-A_{17} (QIDS-A_{17}-C [Adolescent]) and the CDRS-R.

Next, the clinician interviewed the caregiver to complete the caregiver portion of the clinician version of the QIDS-A_{17} (QIDS-A_{17}-C [Parent]) and the CDRS-R. Composite scores for the QIDS-A_{17}-C (Composite) and CDRS-R were generated using the clinician's best estimate of the most valid response based on the ratings of each item (domain) from the parent and the adolescent. The clinician then completed the Children's Global Assessment Scale (C-GAS), to measure overall functioning not limited to impairment from depression, and the Clinical Global Impression scale (CGI-S), an additional measure of symptom severity based on the clinician's impression of overall symptom severity. In addition, the DSM-IV Checklist for Major Depressive Episode (MDE), which established the probable presence or absence of a major depressive episode as defined by DSM-IV criteria, was completed by the treating clinician during the regularly scheduled appointment. When possible, separate clinicians were used to complete the MDE checklist versus the other study measures to avoid bias from the clinician due to prior exposure to the subjects' scores on the depressive symptom measures. Subjects were assessed only once. Upon completion of all measures, the participant was compensated monetarily with a $25 gift card to Target.

All data obtained were stored in a locked file cabinet within a locked room at the UT Southwestern Research Center for Pediatric Psychiatry. Data was removed from the locked cabinet for entry into a confidential database and immediately returned to the cabinet after data entry. All data entered into this database was double-entered and checked to ensure accuracy prior to data analysis.

Measures

DSM-IV Checklist for Major Depressive Episode (MDE)

The MDE checklist is a list of ten symptoms included in the diagnosis of a major depressive episode, according to DSM-IV criteria. The clinician indicates the symptoms for which the patient meets criteria, as determined by a clinical interview, and then indicates whether the patient meets criteria for a current major depressive episode (*Definite*), is likely depressed but does not meet full criteria for a major depressive episode at the time of the interview (*Probable*), or does not meet criteria for a major depressive episode (*No*). This is not a structured diagnostic measure, but rather an indicator of clinical diagnosis, based on a clinical interview during a regular clinic appointment. The MDE checklist is included in Appendix B.

Children's Depression Rating Scale-Revised (CDRS-R; Poznanski & Mokros, 1995)

The CDRS-R is a clinician-rated instrument used to measure the presence and severity of depressive symptomatology in children and adolescents. It has 17 items and typically assesses symptoms occurring in the last 7 days. The total raw score ranges from 17 to 113. The T-score ranges from 30 to 100. A T-score of 65 to 74 suggests that a major depressive episode may be present. The CDRS-R is administered to the child and parent separately, and the clinician uses clinical judgment to synthesize the separate responses for a composite score. The psychometric properties of the CDRS-R are discussed beginning on page 9.

Quick Inventory of Depressive Symptomatology for Adolescents (QIDS-A_{17})

The Quick Inventory of Depressive Symptomatology for Adolescents (QIDS-A_{17}) is a 17-item instrument used to measure the presence and severity of depressive symptomatology occurring in the last 7 days. It was adapted from the Quick Inventory of Depressive Symptomatology ($QIDS_{16}$; Rush et al., 2003; Trivedi et al., 2004), an adult measure of depressive symptom severity that is also used for screening. It is available in self-report (QIDS-A_{17}-SR) and clinician-rated (QIDS-A_{17}-C) formats, and includes the following versions: self-report for the adolescent (QIDS-A_{17}-SR), self-report for the parent to complete on the adolescent (QIDS-A_{17}-SR[P]), clinician-rated based on an interview with the adolescent (QIDS-A_{17}-C[Adolescent]), clinician-rated based on an interview with the parent on the adolescent (QIDS-A_{17}-C[Parent]), and clinician-rated composite (QIDS-A_{17}-C[Composite]), based on the clinician's separate interviews with the adolescent and parent. The composite scores are created using the clinician's clinical judgment to synthesize the separate responses. In addition, the clinician versions are available in a semi-structured interview format that provides semi-structured prompts for each question.

The scoring of all versions of the QIDS-A_{17} is identical to scoring the adult version and the items (now 17 instead of 16 since an irritability item was added to further assess the mood domain) cover the same nine DSM-IV TR symptom domains. For symptom domains requiring more than one item (i.e. appetite/weight change, sleep disturbance, and psychomotor agitation/retardation), the response to the highest scored item in the domain is included in the total score. The total score range on the QIDS-A_{17} is 0-27. The

psychometric properties of the $QIDS_{16}$ and $QIDS\text{-}A_{17}$ are discussed beginning on page 14.

The Clinical Global Impression Scale (CGI; Guy, 1976)

The CGI is used as a clinician assessment of overall symptom severity and improvement, each with a seven point scale, with lower values being more favorable and healthy. The CGI has three scales which assess severity of illness, global improvement, and efficacy index. This study will use the severity scale (CGI-S) to reflect current symptom severity. The Severity of Illness item requires the clinician to rate the severity of the patient's illness at the time of assessment, relative to the clinician's past experience with patients who have the same diagnosis (i.e., depression). The patient is assessed on severity of mental illness at the time of rating according to: normal (not at all ill); borderline mentally ill; mildly ill; moderately ill; markedly ill; severely ill; or extremely ill. The items on the CGI are considered universal and appropriate for use in pediatric as well as adult populations. The intraclass correlation coefficient for CGI improvement was found to be 0.93 as a continuous variable and .95 as a categorical variable.

The Children's Global Assessment Scale (CGAS; Shaffer et al., 1985)

The CGAS was adapted from the Global Assessment Scale for Adults and provides a rating of adaptive functioning. This is a measure of the overall level of functioning, not limited to impairment from depression. The subject is rated by a single number, equal to the most impaired level of general functioning over a specified time period. The CGAS is scored on a continuum from 1 to 100, with a low score indicating greater dysfunction.

CHAPTER FIVE

STATISTICS

Psychometric Theory and Statistical Methods

A brief overview of psychometric theory will be presented to provide perspective for the statistical analysis performed in this study. For a more extensive account of psychometric theory, the following references may be reviewed: Nunnally (1978), Nunnally & Bernstein (1994), Clark & Watson (1995) and Embretson & Reise (2000).

Classical Test Theory and Item Response Theory are two distinct theoretical models of measurement used to evaluate the psychometric properties of a test instrument. Both frameworks offer distinctive information about the psychometric properties of a scale and may be used in conjunction to complement one another (Clark & Watson, 1995).

Classical Test Theory (CTT) has been used for many years to evaluate measurement tools and is often referred to as a theory of true and error scores. The basic theory is that test scores result from two sources: the true score of a measured attribute (i.e., depression) and the error in measurement (i.e., the discrepancies between true depression scores and obtained scores). The true scores contribute to the consistency within a test, while the error scores lead to inconsistency. CTT emphasizes reliability, which is a manifestation of the influence of true score variance and error variance on the actual test scores.

Item Response Theory (IRT) is a modern test theory that is increasingly used as a standard for analyzing test data. IRT is also called latent trait theory and allows for more extensive inferential testing than CTT. IRT analyzes individual item scores on a test to determine the probability of a particular response with respect to a subject's level of a measured trait (i.e., depression). IRT theorizes that there is an underlying latent trait, such as depression, and the relationship between the latent trait and an individual's item responses can be estimated, using a variety of methods. IRT assumes that a test is unidimensional, meaning that the test measures a single ability or trait, such as depression. The specific IRT methods are detailed below with respect to the purposes of this study.

CTT measures of scale consistency, including Cronbach's alpha (Cronbach, 1951), item (symptom) means, total scale score means, item score-total score correlations, and intercorrelations among measures were computed for all four depression rating scales (QIDS-A-C(Adolescent), QIDS-A-C(Composite), QIDS-A-SR, and CDRSR).

In order to determine whether the scales met the unidimensionality assumption required for IRT (i.e., the measures assess only depression), exploratory factor analysis and principle component analyses were conducted on the three QIDS-A_{17} scales and the CDRS-R. First, an unrotated common factor analysis was performed on each scale, after which parallel analysis was used to infer dimensionality. Parallel analysis (Horn, 1965; Humphreys & Ilgen, 1969; Humphreys & Montanelli, 1975; Montanelli and Humphreys, 1976) for each measure was computed by randomly generating normally distributed correlation matrices that had the same number of variables/domains as each measure (9

for the QIDS-A_{17} measures and 17 for the CDRS) and the same number of subjects as the actual data contained (n = 100). The 9 X 9 and 17 X17 correlation matrices were each factored 50 times and then averaged to produce a set of 9 random, simulated eigenvalues for the QIDS-A_{17} measures and 17 simulated eigenvalues for the CDRS-R. These simulated eigenvalues were then compared to the eigenvalues obtained from the real data in the initial factor analysis described above. The number of real eigenvalues that were larger than the randomly generated eigenvalues indicated the dimensionality. For a unidimensional scale, the first real eigenvalue should exceed the first randomly generated eigenvalues while the remainder of the obtained eigenvalues should fall below the remaining simulated eigenvalues. Scree plots were generated by plotting the obtained and simulated eigenvalues.

Samejima's graded IRT model (Samejima, 1997) was used to complement the CTT analyses. Item/domain parameter estimates were obtained for the three versions of the QIDS-A measures and the CDRS-R. In Samejima's model, each scale is divided into a series of categories relative to the number of response choices per item/domain in a given scale. For example, all items on the QIDS-A measures use a 4-point scale (i.e., scored on a 0 to 3 scale), which translates to three categories, also known as locations or boundary response functions. The first location, denoted b_0, is the point that separates a response of "0" from responses "1" through "3"; the second location, denoted b_1, is the point that separates responses "0" or "1" from responses "2" and "3"; the third, b_2, is the point that separates responses "0", "1", or "2" from a response of "3".

In order to compare the QIDS-A measures, which have 4 categories of responses for each item (responses 0-3) with the CDRS-R, which has 5 to 7 categories (1-5 or 1-7

response choices, depending on the item), the CDRS-R categories were reduced into 4 categories to allow for equal comparison with the QIDS-A measures. Specifically, any responses coded 5-7 were collapsed into response category four.

Each item/domain is fit with three S-shaped curves along the boundaries b_0, b_1, and b_2, and share a common, parallel slope designated *a*. In this study, slope *a* represents the relationship between the symptom domain and level of depression. Depression is symbolized θ and is presented on a continuum that is scaled to a mean of 0 and a standard deviation of 1 in the sample. The *a* parameter is similar to item-total correlation in CTT.

Using Raju's (Raju, van der Linden, & Fleer, 1995) approach for this interpretation, boundary response function curves describe the probability of choosing a response greater than the given response category, relative to level of depression (i.e., as a function of θ). Figure 1 provides a generic example of boundary function curves for the item sad mood. Category response function curves are similar to the boundary response function curves but represent the probability of a response equaling a given category. Figure 2 presents examples of category response functions for two distinct items, sad mood and appetite, to illustrate the difference in pattern for an item with a strong relationship to depression (sad mood) versus an item with a weaker relationship (appetite). In addition, each curve provides a threshold that represents the half-way cutoff point (i.e., probability of .50) on the curve that is similar to the item mean of CTT. *Multilog for Windows* was used to obtain the Samejima parameter estimates (Thissen, 2003).

In instances where there are multiple groups or conditions, such as the three versions of the QIDS-A measures, IRT enables comparison among the parameters of the

versions by testing the fit of various models. For example, a model in which parameter estimates (i.e., a, b_0, b_1, b_2) of all QIDS-A measures are allowed to vary freely may be compared to a model in which the slope a (i.e., the ability of items/domains to discriminate among levels of depression) is constrained to equality among all measures. The difference in fit between the two models can be expressed as a form of chi-square, denoted G^2, with *df* equal to the difference in number of parameters. If G^2 is significant, this indicates that the items/domains relate to overall depression to a different extent among the four measures. If G^2 is not significant, this implies that there is no difference, such that the domains relate to overall depression in a similar way among the four measures.

The fit of five different models were compared to the three QIDS-A measures to determine if the parameter estimates varied across the versions (i.e., QIDS-A-C[Adolescent], QIDS-A-C[Composite], and QIDS-A-SR). The models were as follows:

- Model 1: All parameters were allowed to vary freely
- Model 2: The thresholds (*b*) were constrained while the slopes (*a*) varied freely
- Model 3: The thresholds (*b*) varied freely while the slopes (*a*) were constrained
- Model 4: The thresholds (*b*) and slopes (*a*) were constrained to equality
- Model 5: The error variances were constrained across all four measures, but varied freely across the 9 domains

Analyses were conducted using Mplus Version 5 statistical software (Muthén & Muthén, 2007). The Bentler-Satorra chi-square (Satorra & Bentler, 1994) was used to compare these models, using the weighted least squares to estimate the parameters. In addition to the difference of fit testing among the different models, the acceptability of model fit was

assessed for each model using the mean square error of approximation (RMSEA), the weighted root mean square residual (WRMR), the comparative fit index (CFI), and the Tucker Lewis Index (TLI). These goodness-of-fit indexes establish whether there is a relatively good fit between the hypothesized model and the observed data. Empirical cutoffs (Hu & Bentler, 1999) were used for the RMSEA (good models < .06), the WRMR (< .90), the CFI (> .90) and the TLI (> .90).

The Samejima model was also used to obtain test information functions (TIF) for each of the three QIDS-A measures and the CDRS-R. The TIF characterize the sensitivity of the scale to slight changes in level of depression. The higher the TIF curve, the more precise the estimate of depression severity is at the corresponding level of depression. This function is somewhat similar to coefficient alpha in measuring the internal consistency of a scale, although coefficient alpha assesses the scale as a measure of overall depression, as opposed to TIF that measures sensitivity detection in levels of depression as a function of the amount of depression.

Finally, the graded IRT model was used to equate scores for all four depression measures. Taking the parameter estimates *a* and *b* for each scale, the procedure of Orlando, Sherbourne, & Thissen (2000) and related software was used to produce a latent trait score (θ) for each possible total score on the QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), QIDS-A_{17}-SR, and CDRS-R. These scores were then equated for each pair of scales by matching the total scores whose values of θ were most similar (Orlando et al., 2000). In instances when the values of θ did not line up exactly, the corresponding total scores were equated using best judgment, taking into consideration the matched scores immediately before and after the total score in question.

Diagnostic validity was explored using five separate analyses: univariate logistic regression, multivariate logistic regression, receiver operating curve (ROC) analysis, univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA). Each analytic approach tested the ability of each of the four measures (the QIDS-A_{17}-C[Adolescent], QIDS-A_{17}-C[Composite], QIDS-A_{17}-SR, and CDRS-R) to discriminate between two groups: depressed subjects and non-depressed subjects. These two groups were originally identified by the MDE checklist which classified subjects as definitely in a current MDE, probably in an MDE (but not meeting all DSM-IV criteria), or not in an MDE at the time of the evaluation. For purposes of analysis, those subjects classified as *probably* in an MDE were pooled with the subjects classified as *definitely* in an MDE so that two distinct groups (depressed and non-depressed) were established.

Univariate logistic regression is a statistical analysis technique that assesses the impact of a predictor variable (the independent variable, i.e., a depression scale) on a criterion variable (the dependent variable, i.e., depression). The dependent variable is dichotomous, such as presence and absence of depression. For example, this study examined whether each of the four measures (the three QIDS-A measures and CDRS-R) predicted the presence or absence of depression. Each measure was examined independently for the univariate logistic regression.

Multivariate logistic regression is similar to univariate logistic regression with the exception that there can be more than one predictor variable (independent variable). For this study, all four depression scales were used as predictor variables to establish the relative predictive importance of each scale (independent variable).

ROC analysis provides the sensitivity and specificity of a test in the form of a graph (Portney & Watkins, 2000). ROC curves (Kraemer, 1992) were created by plotting the successive "obtained" values of sensitivity and specificity for the scores on each of the three QIDS-A measures and the CDRS-R. As an example, consider a score of 5 on the QIDS-A-SR. The proportion of depressed patients falling at that score or higher defines the sensitivity (also called hit rate) and the proportion of nondepressed patients falling at that score or higher defines 1 minus the specificity (also called false alarm rate). Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that curves through the upper left corner (100% sensitivity, 100% specificity). Thus, the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test. This process is included as a logistic measure because most theories of the process that gives rise to the ROC curve assume either underlying logistic distributions or normal distributions, which are closely related.

ANOVA and MANOVA are linear evaluations that assess the ability of the scales to discriminate depressed versus non-depressed groups. ANOVA is a univariate evaluation that is used to see the main and interaction effects of categorical dependent variables, such as depressed and nondepressed subject groups, on a dependent variable, such as one of the four depression scales (the three QIDS-A measures and the CDRS-R). As a univariate evaluation, the test ignored the covariances among the four measures to determine the ability of each measure individually to discriminate between the two groups, separate from the other measures.

MANOVA is a multivariate evaluation that is used to see the main and interaction effects of categorical dependent variables (depressed and non-depressed) on multiple dependent variables (the three QIDS-A measures and the CDRS-R). In other words, MANOVA evaluates what additional information might be obtained about the ability of the four measures, when evaluated together, to discriminate between depressed and non-depressed groups.

Statistical Analysis

Descriptive statistics will be reported for the following demographic variables: patient age, gender, race/ethnicity, and annual income.

Research Question One: What are the psychometric properties of the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and CDRS-R in an adolescent outpatient population, including reliability measures and construct and concurrent validity measures?

Hypothesis One (A): The QIDS-A_{17}-SR will demonstrate good internal consistency, reliability, and related psychometric properties.

Standard Classical Test Theory (CTT; Nunnally & Bernstein, 1994) analyses were used to infer the properties of each rating scale and version (the CDRS-R, QIDS-A_{17}-SR, and QIDS-A_{17}-C for adolescent only and Composite versions). Cronbach's coefficient alpha for the scale as a whole, item means, and item-total correlations was used. Pearson's correlation between versions was used to characterize the difference between total scores for all tests. Mean differences for each item were computed. Factor analysis using parallel analysis (Carmody et al., 2005; Nunnally & Bernstein, 1994) was conducted for all tests to determine if the scales are unidimensional.

In addition, Samejima's item response theory (IRT; Samejima, 1997) model was used to compare each rating scale and version with respect to item characteristic curves (intercepts and slopes) for each item. These are the IRT equivalents of the item means and item-total correlations described above in CTT. In addition, test information function

(TIF) were computed for each version of each rating to determine and compare which areas of the test are most sensitive.

Hypothesis One (B): The QIDS-A_{17}-C(Adolescent) will demonstrate good internal consistency, reliability, and related psychometric properties.

CTT and IRT were used as described in Hypothesis One (A).

Hypothesis One (C): The QIDS-A_{17}-C(Composite) will demonstrate good internal consistency, reliability, and related psychometric properties.

CTT and IRT were used as described in Hypothesis One (A).

Hypothesis One (D): The CDRS-R will demonstrate good internal consistency, reliability, and related psychometric properties.

CTT and IRT were used as described in Hypothesis One (A).

Research Question Two: What is the relationship between the scores on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R?

Hypothesis Two: The QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), the QIDS-A_{17}-C(Composite), and the CDRS-R will correlate highly.

Concurrent validity was calculated by correlation coefficient *r*.

Research Question Three: What is the relationship between the scores on the QIDS-A_{17}-SR, the QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R, and the presence of a major depressive episode (MDE) as determined by the MDE checklist?

Hypothesis Three (A): The presence or absence of MDE, as determined by the MDE checklist, will be related to the scores on the three versions of the QIDS-A_{17} and the CDRS-R to identify relevant scoring thresholds by which to identify MDE.

Logistic regression was used to examine the scales' abilities to differentiate depressed from nondepressed subjects and measure incremental validity.

Hypothesis Three (B): Higher scores on the QIDS-A_{17} versions and the CDRS-R will relate to an increased probability of being depressed.

Receiver operating characteristic (ROC) curves (Rush & Trivedi, 1995; Tanner & Swets, 1954) were used to determine the optimum threshold to determine the presence of MDE for each rating. The presence or absence of a major depression episode was determined by the MDE checklist (as noted above in Measures).

Hypothesis Three (C): Higher scores on the QIDS-A_{17} versions and the CDRS-R will be uniquely correlated to an increased probability of being depressed.

A multivariate analysis of variance (MANOVA) was applied to all scales to explore unique correlations and degrees of discrimination between depressed and nondepressed subjects.

Research Question Four: What is the relationship between the total scores on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R?

Hypothesis Four: The relationship between individual items on the QIDS-A_{17}-SR, QIDS-A_{17}-C(Adolescent), QIDS-A_{17}-C(Composite), and the CDRS-R will be established with the creation of inter-test conversion tables.

Standard IRT procedures for test equating (as noted above in Statistical Methods), were used to construct tables of equivalent scores for each pair of rating scales (CDRS-R and QIDS-A_{17}-SR, QIDS-A_{17}-C[Adolescent], and QIDS-A-C[Composite]). This facilitated the creation of correspondence tables that determined what score on each rating corresponded to a given score on the other rating. Such tables enabled translation of thresholds for mild, moderate, and severe depression established for the CDRS-R to the QIDS-A_{17}-C (Adolescent and Composite) and QIDS-A_{17}-SR.

CHAPTER SIX

RESULTS

Among 103 subjects included in the QIDS-A study, 68 were recruited directly from the clinic and 35 were included as a limited data set from a separate, NIMH-funded study, noted previously. Of the 35 in the limited data set, 3 subjects did not complete all measures and were excluded from some of the analysis, detailed below.

From February 20, 2008 until December 29, 2008, approximately 359 eligible patients came to the clinic. Of these, 112 were approached for possible participation, 98 agreed to participate, and 68 completed the study. Of the 30 patients who agreed to participate but did not enter the study, the majority were not able to be reached to follow up. Eleven patients refused to participate upon approach, mainly indicating that they were not interested. Three patients expressed uncertainty about participating, took a flyer for consideration, and did not follow up. Of the 247 eligible patients who were not approached for possible participation, the majority were missed due to limited staffing. Some of these may include patients who did not show for an appointment.

Descriptive Statistics

The sample consisted of 103 outpatient clinic and research participants (8-17 years of age). Sixty-eight participants were recruited during a scheduled outpatient clinic visit, and 35 were included as a limited data set from a separate, ongoing treatment study for pediatric depression. The ethnic racial breakdown of the sample was found to be 70%

Caucasian, 16% African-American, 10% Hispanic/Latino, 3% American Indian/Alaskan Native, and 1% Asian. Females comprised 51% of the sample. The ages ranged from 8 to 17 years old with a mean age of 13.8 ± 2.4. Among the 103 subjects, 21 (20%) were between 8 and 11 years old. Based on a diagnostic checklist for MDE, as detailed previously in Methods and Measures, 40 (39%) subjects met criteria for a current Major Depressive Episode, 55 (53%) were not depressed, and 8 (8%) had some symptoms of depression but did not meet full criteria for a current Major Depressive Episode. Table 2 shows the demographic and clinical characteristics of the sample.

Classical Test Theory Analysis

Tables 3-5 summarize the CTT analyses for the QIDS-A_{17} measures and the CDRS-R.

Internal Consistency

Medium to high internal consistency was found for all measures, in this particular population, using Cronbach's alpha. Internal consistency was equal for the QIDS-A_{17}-C(Adolescent) and the QIDS-A_{17}-C(Composite) ($\alpha = .84$). The CDRS-R was the most reliable ($\alpha = .92$). The QIDS-A_{17}-SR showed the weakest reliability of the measures ($\alpha = .78$), although it was at an acceptable level for reliability.

The Spearman-Brown prophecy formula (Nunnally & Bernstein, 1994) was computed on the Cronbach's alphas for all QIDS-A_{17} measures to compare the difference in scale length between the QIDS-A_{17} (9 domains) and the CDRS-R (17 domains) to

make sure length was not a factor in reliability. In other words, this correction removed any difference due to test length to allow for equal comparison of reliability across measures. Such correction increased the reliabilities of the QIDS-A_{17} measures in this particular population to Cronbach's alpha ≥ .87. Coefficient alpha was .87 for the QIDS-A_{17}-SR and .91 for both the QIDS-A_{17}-C(Adolescent) and the QIDS-A_{17}-C(Composite), indicating a high degree of internal consistency for all measures. In comparison to the CDRS-R, the difference in uncorrected reliability for the two QIDS-A_{17}-C versions was mostly due to the difference in length (QIDS-A_{17}-C is less reliable only because it is a shorter scale). The QIDS-A_{17}-SR was still slightly less reliable, but the difference was minimal.

CTT analyses were also computed on the QIDS-A measures with the irritability item removed to examine the contribution of the irritability item. Results found that introducing the irritability item reduced reliability by .01 in all QIDS-A_{17} measures, although this change was not significant.

Item Means and Item-Total Correlations

Item means and item-total correlations (r_{it}) were computed for all measures (Tables 3 and 4). The item means measured the tendency of subjects in this setting to endorse particular symptoms. The item-total correlations (domain-total correlations) measured how robustly a symptom relates to overall depression as indicated by the total scale score.

Across QIDS-A_{17} measures, sleep disturbance was the most commonly endorsed symptom, followed by sad or irritable mood. In addition, appetite disturbance was

frequently endorsed on the QIDS-A_{17}-SR. Although these symptoms were most often reported, sad/irritable mood and loss of general interest were symptoms most strongly related to overall depression across QIDS-A_{17} measures. Symptoms least reported across measures included thoughts of death or suicide and loss of general interest.

Irritability and low self-esteem were most frequently endorsed on the CDRS-R, while difficulty having fun, depressed feelings, and social withdrawal were most highly related to overall depression. Morbid ideation and listless speech were endorsed least frequently. Note that symptoms with high item-total correlations (i.e., strongly related to depression) on the CDRS-R had corresponding items that were highly correlated on the QIDS-A_{17} measures.

Removal of the irritability item for comparison showed only one observable difference across the QIDS-A_{17} measures. The sad mood item (without irritability) was endorsed slightly less frequently while relating more robustly to overall depression.

Correlations Among Measures

As seen in Table 5, total scores on both QIDS-A-C measures (Adolescent and Composite) showed high correlations with the CDRS-R (r = .78 and .89, respectively). The QIDS-A-SR correlated moderately with the QIDS-A-C(Adolescent) (.69), the QIDS-A-C(Composite) (.66), and the CDRS-R (.63). As one would expect, the QIDS-A-C(Composite) correlated highly with the QIDS-A-C(Adolescent) (r = .88).

Disattenuated intercorrelations were computed by correcting the above intercorrelations for unreliability due to measurement error. Such correction strengthened the correlation between QIDS-A-SR and QIDS-A-C(Adolescent) from .69

to .85. The disattenuated correlations between the QIDS-A-C(Composite) and two measures (QIDS-A-C[Adolescent] and CDRS-R) exceeded 1.0, reflecting essentially perfect correlations when corrected for unreliability.

Exploratory Factor Analysis

Scale dimensionalities

Figures 3 and 4 present the scree plots that were created using the factor analysis and parallel analysis described above in Statistical Methods. For all four measures (the three QIDS-A_{17} versions and the CDRS-R), the first real obtained eigenvalue far exceeded the first randomly generated eigenvalue, while the remainder of the real eigenvalues were smaller than the random eigenvalues. The first few obtained and simulated eigenvalues are listed in Table 6. These results indicate that all three versions of the QIDS-A_{17} as well as the CDRS-R were unidimensional for this sample.

The sample sizes of each measure varied slightly due to missing data. The actual sample size ranged from 99 to 103, so n = 100 was used in the parallel analysis. To make equal comparisons, parallel analysis was also run on n = 99 and n = 103 to compare eigenvalues to actual results. Unidimensionality in all measures was confirmed in these comparisons.

Item Response Theory Analysis

Item Response Theory Comparisons

Figures 5 and 6a-c contain the Samejima (1997) IRT parameter estimates for the QIDS-A-C(Adolescent), QIDS-A-C(Composite), and QIDS-A-SR. For all measures, the estimates were obtained from a model in which parameters were allowed to vary freely. Figure 5 illustrates the pattern of influence of depression on each domain response (i.e., sad mood). As noted in Statistical Methods, this is similar to item-total correlation in CTT. This relationship is represented by the slope *a*, and larger slope parameters indicate greater depiction of depression. For example, general interest, self-view, and sad/irritable mood domains most represented depression in this sample as inferred from the three QIDS-A measures given that their slope (*a*) parameters are generally the largest among the domains. In contrast, appetite and sleep were the least representative of depression among all three QIDS-A measures, indicated by the small slope (*a*) parameters.

There was some variability in the domain slopes among the three QIDS-A measures. For example, general interest was most influenced by depression (i.e., had the highest slope) for both QIDS-C versions (Adolescent and Composite), while sad/irritable mood was most influenced by depression on the QIDS-A-SR. Note that adolescents tended to endorse more pathology on the self view and suicidal ideation domains when interviewed by a clinician versus self report.

Figures 6a, 6b, and 6c contain the b_0, b_1, and b_2 parameter estimates for the three QIDS-A measures. These reflect the likelihood of choosing a particular response (i.e., 0-3 on a QIDS-A_{17} domain item) in each domain, regardless of how well the domain relates to depression. As described in Statistical Methods, these are similar to the item means in

CTT. A lower threshold reflects a higher probability that one would choose a more pathological answer. For example, the suicidal ideation domain shows higher thresholds than the other domains, indicating that the item mean for suicidal ideation is lower (i.e., endorsed infrequently). Figure 6a shows lower thresholds for sad/irritable mood than for appetite, indicating that people are more likely to choose a more pathological response (i.e., a response of 1, 2, or 3, as opposed to 0) for sad/irritable mood than for appetite. Notice that sleep consistently shows a higher likelihood of a pathological response (Figures 6a-6c), even though it was found to be least influenced by depression (Figure 5). This is similar to the CTT results, noted above, in that frequency of a domain response did not always reflect a strong relationship to depression.

Corresponding Samejima (1997) estimates were obtained for the CDRS-R. Difficulty having fun, hypoactivity, and depressed feelings were the most discriminating items for depression, while sleep disturbance and morbid ideation were the least discriminating for the CDRS-R in this sample.

Although the domains of the CDRS-R could not be directly compared with the QIDS-A domains due to the difference in number and content overlap of many of the items, the sad mood domain was similar enough (sad/irritable mood in QIDS-A measures; depressed feelings in CDRS-R) to provide an IRT example of one item across all measures. Figure 7 depicts the category response functions of all four measures for the sad mood item, corresponding to the lowest response category on each measure (i.e., responses of “0” on the QIDS-A measures and “1” on the CDRS-R). As described above in Statistical Methods, category response functions represent the probability of response equaling a given response category for a particular level of depression, as derived from

the a, b_0, b_1, b_2 parameters. Thus, Figure 7 shows the probability of a subject choosing a "0" (or "1" on the CDRS-R) on the sad mood item (x-axis) in relation to the continuum of level of depression, or F (y-axis). Note that the probability of choosing "0" is high when the level of depression is extremely low or nonexistent (i.e., the far left of the y-axis). The slopes decrease as the level of depression increases towards the right on the y-axis continuum, indicating the low probability of choosing "0" on sad mood when depressive pathology increases. Figure 7 also nicely illustrates the difference of sad mood reporting among the four depression measures, particularly the difference between the CDRS-R versus the three QIDS-A measures. Notice that the curve for the CDRS-R is shallower than the QIDS-A measures, particularly the QIDS-A-C(Adolescent). Although the difference is slight, this demonstrates the ability to explore the variation in item responses in different ways.

Category response function curves were generated for all four measures separately, for all four response categories (i.e., 0-3), to illustrate the varying response probabilities as related to depression. These are provided in Appendix E. As an example, Figure 8 illustrates the category response function slopes of two extreme items, sad/irritable mood and appetite, on the QIDS-A-C(Adolescent), for response category "0". Note the steep slope on sad/irritable mood, indicating the strong relationship between level of depression and probability of choosing a response of "0" for sad/irritable mood. As depression increases, the probability of choosing a response of "0" sharply decreases. In contrast, appetite has a relatively shallow slope, demonstrating

a weaker relationship to level of depression. The slopes for response category "1" will be the same but will be shifted to the right on the depression continuum.

As noted above in Statistical Methods, IRT was used to establish whether the parameter estimates (i.e., a, b_0, b_1, b_2 parameters) differed across the three versions of the QIDS-A_{17} measures. This was determined by comparing the fit of data from the model in which all parameters were allowed to vary freely (i.e., Model 1) with four other models with varying constrained parameters. First, Model 1 was tested for fit by comparing it to a baseline model (also called a null model) that has no structure and stipulates that all items are independent of one another (i.e., assumes no relationship between the parameters and depression). The chi-square test statistic for Model 1 was 175.07 (df = 65), while the chi-square test statistic for the baseline model was 1637.42 (df = 51), ps < .00001. The lower chi-square value indicates better fit. Thus, Model 1 fits much better than the baseline model, indicating that depression significantly influences the domain responses on the QIDS-A measures. In other words, there is a relationship between depression and the domain responses. Furthermore, each of the three QIDS-A_{17} measures was tested to evaluate their individual contributions to the fit within Model 1. All three QIDS-A_{17} measures contributed approximately equally to the goodness of fit (χ^2s = 41.57, 44.09, and 46.17 for QIDS-A-C[Composite], QIDS-A-C[Adolescent], and QIDS-A-SR, respectively, at ps < .001).

As described above in Statistical Methods, additional measures of fit were examined, including the CFI, TLI, RMSEA, and WRMR. Mixed results were found, as some indexes found this model to be a reasonable fit (CFI = .93, TLI = .95) while others indicated a poor fit (RMSEA = .13, WRMR = 1.61). Since Model 1 was a significantly

better fit than the baseline model, and all three QIDS-A measures contributed equally to the model fit, and since the CFI and TLI supported a good fit for Model 1, it was determined that Model 1 was acceptable to use, despite mixed results.

After Model 1 was found to be acceptable, comparisons in fit were made to Model 2, 3, 4, and 5. Model 2, where the thresholds (*b*) were constrained while the slopes (*a*) were allowed to vary freely, tested for intercept differences. Model 2 is a nested model under Model 1 as it is a possible outcome for Model 1. The chi-square differed significantly ($G^2(29) = 54.15$, $p < .01$), indicating that the thresholds differ among the three QIDS-A measures. Therefore, constraining the thresholds made the model fit worse, so this model was rejected.

Model 3, where thresholds (*b*) varied freely while the slopes (*a*) were constrained, tested for slope differences. The chi-square differed significantly ($G^2(37) = 56.78$, $p < .05$), indicating that the patterns (i.e., slopes) differ among measures. Although the thresholds (Model 2) and slopes (Model 3) both differ significantly, Model 2 has a higher chi-square/degrees of freedom ratio (54/29 = 1.86) than Model 3 (56/37 = 1.51), indicating that the thresholds differ more than the slopes across the three QIDS-A measures.

Model 4, where both thresholds (*b*) and slopes (*a*) were constrained to equality, differed significantly ($G^2(44) = 72.94$, $p < .01$). This was to be expected as Model 2 and 3 already indicated that the thresholds and slopes differ.

Additional models were run, including Model 5, where the error variances were constrained across all three measures but allowed to vary freely across the 9 domains.

However, these models were not identified (i.e., not capable of producing a unique solution) and therefore were not able to be computed.

Test Information Functions

A test information function represents the ability of the scale as a whole to discriminate differences in the magnitude of depression (θ). It is a measure of reliability and can be compared to coefficient alpha in CTT. Figure 9 includes these information functions for the three QIDS-A measures and the CDRS-R. Results indicate that the CDRS-R was the most sensitive from about -1.5 z-score units below the mean up to 3 z-score units above the mean (i.e., from mild depression to severe depression). The QIDS-A-SR and QIDS-A-C(Composite) were both slightly more sensitive in discriminating from no depression to mild levels of depression (from $\theta = -3$ to -1.5). Among the three QIDS-A measures, the QIDS-A-C(Composite) performed better from $\theta = -1.5$ to 0 (i.e., lower to moderate levels of depression), after which the QIDS-A-C (Adolescent) performed slightly better up to $\theta = 1.5$, i.e., moderate depression. From this point, the QIDS-A-SR performs just slightly better. These results are consistent with the previously presented coefficient alphas, where the CDRS-R was the most reliable (i.e., the highest coefficient alpha), followed by both QIDS-A-C versions, and then the QIDS-A-SR. The test information functions elaborate on the CTT reliability by indicating not only the strength of reliability but also the points at which it is most sensitive to various levels of depression. As such, although the CDRS-R was most reliable, TIF results clarified that it was most reliable in detecting moderate depression, while the QIDS-A-SR is actually more reliable in detecting mild depression. Individually, each measure peaked at

approximately $\theta = 0.5$, indicating that all measures are more discriminating for moderate depression, regardless of the level of sensitivity noted above.

Diagnostic Validity

Table 7 shows the effect sizes for all QIDS-A measures and the CDRS-R for univariate and multivariate logistic regression and linear statistical measures, detailed below. All reported results were standardized to control for differences in measurement among the scales.

Univariate Logistic Regression Analysis

Univariate logistic regression analysis was conducted to examine the ability of each scale individually to classify depressed and non-depressed subjects, as measured by the MDE checklist. A logistic regression with only the intercept entered produced a "-2 Log Likelihood" (-2LL) criterion (residual chi-square) of 137.63. A lower -2LL value is produced when the intercept plus a covariate (one of the four measures, i.e., one of the QIDS-A scales or the CDRS-R) is entered. The reduction from the higher value (intercept only) to the lower value (intercept and covariate) determines if the individual covariate is significant for detecting depression (i.e., classifying depressed from non-depressed). The amount of reduction is known as the Likelihood Ratio, which is the test statistic for -2LL. Individually, each measure reduced the residual chi-square by the following Likelihood Ratio chi-square values: CDRS-R (80.59), QIDS-A-C(Composite) (65.70), QIDS-A-C(Adolescent) (47.20), and QIDS-A-SR (38.85). These decreases were

significant on 1 *df*, *p*s < .0001. This indicates that the CDRS-R was the most discriminating of depressed versus non-depressed as determined by the MDE checklist, and the QIDS-A-SR was the least discriminating of the four measures. They were all significant for detecting depression.

The regression weight, which is the log of odds ratio favoring being depressed, was also examined as another way to determine the ability of each measure to estimate depressed subjects versus non-depressed. These estimates are listed in Table 7 as the effect size estimates for univariate logistic regression. All estimates were significant on 1 *df*, *p*s < .0001. The QIDS-A-C(Composite) was the best estimate of all four measures, while the QIDS-A-C(Adolescent) and QIDS-A-SR were slightly less precise. Interestingly, the CDRS-R was the poorest estimate. However, examination of the Wald chi squares for each measure show this difference to be minimal (χ^2s [1, N = 100] = 27.17, 26.92, 24.14, and 22.74 for the QIDS-A-C[Composite], QIDS-A-C[Adolescent], QIDS-A-SR, and CDRS-R, respectively, *p*s < .0001).

Multivariate Logistic Regression Analysis

The multivariate logistic regression computed the same statistics detailed above for univariate logistic regression, but instead of examining each measure separately, all four measures (the three QIDS-A measures and CDRS-R) were entered together as covariates. The Likelihood Ratio was 85.23 and was significant on 1 *df*, *p*s < .0001. The regression weights are listed in Table 7. Of the four measures, the CDRS-R was the only significant estimate. This is due to the fact that the QIDS-A measures are so similar that

they become redundant when run collectively as covariates such that any significance is lost.

ROC Analysis

Figure 10 displays receiver operating characteristic (ROC) curves that were created using data obtained in the logistic regression analysis regarding sensitivity and specificity of each scale's ability to classify depressed and non-depressed groups. Note that the CDRS-R is most sensitive to depression from a false alarm rate of 0.0 up to a false alarm rate of .15. At this point, the QIDS-A-C(Composite) is most sensitive up to a false alarm rate of .45. The CDRS-R is once again more sensitive from this point up to a false alarm rate of .80, after which the curves begin to converge.

Another way to evaluate the performance of each measure is to compare the areas under each curve (c-statistic). Greater area under the curve indicates better overall performance. The areas under these curves were .952, .946, .880, .870 for the CDRS-R, QIDS-A-C(Composite), QIDS-A-C(Adolescent), and QIDS-A-SR, respectively. Thus, the CDRS-R and QIDS-A-C(Composite) were almost equal in best overall performance in classifying depressed from nondepressed groups, while the QIDS-A-C(Adolescent) and QIDS-A-SR were similar and still robust in overall performance.

Table 8 provides the thresholds, sensitivities, and specificities at four particular locations along the depression continuum for each measure. These thresholds were chosen based on scores that reflected at least 30% sensitivity (low), 50% sensitivity (medium), 70% sensitivity (high), and 90% sensitivity (very high). Such cutoff points

were based on criteria used in a previous study that employed similar statistical analyses with the adult QIDS-SR_{16} (Bernstein et al., in press).

Analysis of Variance (ANOVA)

ANOVAs were conducted for each measure (the three QIDS-A measures and CDRS-R) to test whether each measure independently was able to differentiate between the depressed and non-depressed groups, using the MDE checklist clinical diagnosis to define the groups. Table 9 provides the means and standard deviations for the depressed and nondepressed groups. The F-test found that the difference in means between the groups (depressed and not depressed) was significant at *p*s < .0001 for all four measures, with F(1, 98) = 135.99 for the CDRS-R, 102.76 for the QIDS-A-C(Composite), 66.13 for the QIDS-A-C(Adolescent), and 49.39 for the QIDS-A-SR.

The effect size for each ANOVA was calculated by dividing the model sum of squares by the corrected total sum of squares. The effect size values are listed in Table 7 and follow the same order of strength of discrimination as listed above for the F-test results. These results suggest that the CDRS-R is the most discriminating of depressed versus non-depressed groups, followed by the QIDS-A-C(Composite), QIDS-A-C(Adolescent), and finally the QIDS-A-SR. Regardless of the order, all measures were found to discriminate between the two groups at a significant level (*p*s < .0001).

Multivariate Analysis of Variance (MANOVA)

A multivariate analysis of variance (MANOVA) was conducted employing all four scales (the three QIDS-A measures and CDRS-R). The Wilkes' lambda multivariate

test of overall differences among groups, which measures the collective ability of the measures to discriminate depression, was statistically significant, as expected, with $F(4, 95) = 35.35$, $p < .0001$. This supports the logistic regression and ANOVA results that found all four measures to significantly differentiate the depressed and non-depressed patients. More importantly, the MANOVA provides the weight of each measure on the discriminant axis, to show the maximally discriminating way to combine all four measures. The weights indicate the incremental discriminating importance of each scale, i.e., the relative strength of each scale's ability to discriminate depressed and non-depressed when all four measures are combined. The corresponding discriminant weights are listed in Table 7. Results indicate that the CDRS-R increments prediction of depression the most, while controlling for the other three scales, and the QIDS-A-C(Adolescent) increments prediction least of the four measures.

Equated Scale Scores

Table 10 shows the conversion between equivalent levels of symptom severity in this sample for all four measures. By constructing tables of equivalent scores between each rating scale, this allowed for determining what score on each rating scale corresponded to a given score on the other rating scale as compared to a particular level of depression.

CHAPTER SEVEN

DISCUSSION

The present study was designed to evaluate the psychometric properties of the three QIDS-A measures (clinician-adolescent, clinician-composite, and adolescent self-report) and determine whether either the adolescent self-report (QIDS-A-SR) or the clinician interview with the adolescent alone (QIDS-A-C[Adolescent]) are acceptable to replace more time consuming and expensive measures to meet the need for an accurate, quick, affordable measure of depression in adolescents. The results indicate that all measures are of acceptable reliability and validity. Although the CDRS-R and QID-A-C(Composite) had the highest overall reliability, the QIDS-A-SR was reliable enough to be considered for use in lieu of a more time consuming interview combining both adolescent and parent output. The QIDS-A-C(Adolescent), found to be slightly more reliable than the QIDS-A-SR, could also effectively be used in place of a composite interview. The initial discussion will address the findings within each major aim.

The primary aim of the study was to define the psychometric properties of the QIDS-A-C(Adolescent), QIDS-A-C(Composite), and QIDS-A-SR, using the CDRS-R as a field standard measure of comparison.

Classical Test Theory

Internal Consistency

As hypothesized, each measure demonstrated strong internal consistency in this particular population. The CDRS-R and both clinician versions of the QIDS-A_{17} (Adolescent and Composite) showed the highest reliability, and coefficient alphas were essentially comparable once the differences in test length were accounted for (.90-.92). The QIDS-A-SR_{17} showed slightly lower internal consistency, although it was well within the acceptable range.

These results differ somewhat from the Bernstein et al. (2008) study, described on page 25, which compared the adult $QIDS_{16}$ measures (without an irritable mood item) to a similar outpatient adolescent population. Bernstein and colleagues found the QIDS-SR_{16} to be as reliable as the CDRS-R (α = .86 and .87, respectively) even without correcting for the difference in test length. Furthermore, the QIDS-A-SR showed slightly higher internal consistency than the parent, adolescent, and composite versions of the adult QIDS-C_{16}. The Bernstein et al. study results, paired with the data from the current study that indicated removing the irritability item on the QIDS-A_{17} data increased reliability by .01, suggests that the irritability item is not essential for strong reliability in this sample.

Item Means and Item-Total Correlations

As detailed in Results, there were similarities across the three QIDS-A_{17} measures for tendencies to endorse symptoms, particularly sleep and sad or irritable mood, and also for symptoms that related most strongly to overall depression (sad or

irritable mood and loss of general interest). When the irritability item was removed from the data, the sad mood item (without irritability) was endorsed slightly less frequently while relating more robustly to overall depression. It might be that the irritability item, in this particular sample, is also endorsed as a symptom of other disorders separate from depression (e.g., Bipolar Disorder, Attention-Deficit/Hyperactivity Disorder, Oppositional Defiant Disorder/Conduct Disorder), such that it slightly decreases the loading to depression. It is interesting to note that although irritability was the most frequently endorsed symptom on the CDRS-R, it related to overall depression only half as strongly as the depressed feelings symptom. This might further suggest, as with the QIDS-A_{17}, that irritability is an important clinical symptom of depression but not necessarily an essential factor in measuring overall depression.

Correlations Among Measures

Intercorrelations of total scores among measures showed variability. All measures that were given by a clinician (CDRS-R and both QIDS-A-C_{17} versions) were highly correlated (.78-.89). In contrast, the QIDS-A-SR_{17} showed only moderate correlations (.63-.66) with most measures, although it shared a slightly stronger correlation with the QIDS-A-C(Adolescent) (.69). The stronger correlations among clinician measures may be in part due to the fact that the information is gathered by a clinician, such that clinically-minded probes were asked to garner similarly appropriate information. Furthermore, although multiple clinicians were used in the study, each clinician administered all measures to a particular subject which may have increased the relationship among measures for each subject.

These results differed slightly from the Bernstein et al. (2008) study that examined the adult $QIDS_{16}$ in an adolescent population. In the Bernstein et al. study, the $QIDS\text{-}SR_{16}$ and the $QIDS\text{-}C_{16}$(Adolescent) were more highly correlated (.81) than in the present study (.69). Similarly, the $QIDS\text{-}SR_{16}$ was more highly correlated to the $QIDS\text{-}C_{16}$(Composite; .73) in the Bernstein et al. study than the present study (.66).

One difference between the Bernstein et al. (2008) study and the current study is that the Bernstein et al. study administered all measures in a randomized order, while the current study always administered the $QIDS\text{-}A\text{-}SR_{17}$ to the adolescent first, before the clinician administered the other measures. The current study was designed this way to avoid any bias on the $QIDS\text{-}A\text{-}SR_{17}$ scores that might result from begin probed about the same symptoms by the clinician prior to completing the self report. Thus, it is possible that the different results between these two measures might be in part due to methodology differences and resulting bias.

Item Response Theory

Item Response Theory Comparisons

The IRT analyses provided additional support with regard to the reliability as well as comparability of the QIDS-A measures. Specifically, the same items shown in CTT to relate most strongly to depression (loss of general interest, self view, and sad/irritable mood) were also found in the IRT analyses to be the most discriminating for depression. The same was true for items with low correlations to depression (appetite and sleep). Similarly, IRT and CTT results corresponded with regard to items that were most frequently endorsed (i.e., sleep and sad/irritable mood).

Although the three QIDS-A measures were similar in response pattern, there were some interesting differences. On the QIDS-A-SR, sad mood was most related to depression while the clinician-rated versions found loss of general interest to correlate most with depression. This may be related to the clinician's ability to probe for pertinent information that one may not think of when completing a self report. Similarly, adolescents tended to report more pathology on self view and general interest when interviewed by a clinician than on self report, possibly due to querying on the clinician's part.

While the domains of the CDRS could not be directly compared to the QIDS-A measures in IRT, some similarities were found in items most influenced by depression. The depressed feelings item on the CDRS-R corresponds to the QIDS-A sad/irritable mood item, while the difficulty having fun item on the CDRS-R is somewhat similar to the loss of general interest item on the QIDS-A measures. Sleep disturbance and morbid ideation were the least discriminating on the CDRS-R. It is possible that the weak relationship between depression and morbid ideation in this sample is simply due to the fact that it was not a common response, which restricts the possible threshold range necessary for accuracy in IRT.

Multiple models were applied to the QIDS-A data to examine whether the three versions of the QIDS-A varied across domain responses and influence of depression. Model 1 allowed for the most variation among parameters. Although Model 1 did not show a strong fit, it was the best fit of all the models and was good enough to use for purposes of IRT analyses in this study. Use of this model specified that the parameters are different among the three versions of the QIDS-A measures in this sample.

Test Information Functions

The test information functions found all measures to be most discriminating at moderate levels of depression, with the CDRS-R the most sensitive of the measures for moderate depression. Interestingly, the QIDS-A-SR and the QIDS-A-C(Composite) were slightly more reliable than the other measures in detecting mild depression, at almost equal sensitivity levels. The QIDS-A-C(Adolescent) was slightly more discriminating than the other QIDS-A measures at a moderate level of depression. The current results differ from the test information functions generated in the Bernstein et al. study (2008) that investigated the adult $QIDS_{16}$ in an adolescent population. Bernstein et al. found all measures to be more discriminating at all levels of depression than in the current study, even on the extreme ends (i.e., no depression and severe depression). Furthermore, the Bernstein et al. study found the $QIDS\text{-}SR_{16}$ to be more sensitive than the $QIDS\text{-}C_{16}$ measures at moderate levels of depression, while the opposite was found for the $QIDS\text{-}A_{17}$ measures in the present study. These discrepancies might be due to sample differences, particularly regarding ratio of depressed and non-depressed patients. The Bernstein et al. study had a limited representation of depressed patients (9% [12 subjects]; based on clinical diagnoses) as compared to the current study (39% [40 subjects]; based on the MDE checklist). Although the methods of diagnosis were different between the studies, both studies considered a patient to be currently depressed if they met DSM-IV criteria for an MDE. Bernstein et al. classified the remainder of their subjects as either in remission from depression (50% [71 subjects]) or never depressed (41% [57 subjects]).

It is also possible that some differences in results between the Bernstein et al. (2008) study and the current study are due to age differences. The Bernstein et al. study used subjects aged 12 to 17, while the current study used a wider age range (8 to 17). However, the mean ages of each sample were similar (14.4 ± 1.5 years for Bernstein et al., and 13.8 ± 2.4 years for the current study).

Scale dimensionalities

As described in Results, all four measures were unidimensional. With regard to the QIDS-A_{17}, this was expected and consistent with previous literature on the adult $QIDS_{16}$ in adult populations (Bernstein et al., 2007; Rush, Bernstein, et al., 2006; Rush et al., 2003; Trivedi et al., 2004) as well as an adolescent population (Bernstein et al., 2008). It was somewhat surprising that the CDRS-R was unidimensional in this sample as it has been found to be at least two dimensional in numerous studies (Bernstein et al, 2009; Guo et al, 2006; Jain et al, 2007). This has been explained as due to the CDRS-R measuring a combination of symptoms of depression, signs of depression (i.e., observed depressed facial affect, observed listless speech), and level of functioning (i.e., impaired school work). One possible explanation for the current finding of unidimensionality is that, in this sample, patients with the most depressive symptoms also provided the most depressive signs and functions, thus pushing what might otherwise be two or three factors (signs, symptoms, and function) into one factor.

A second aim of the study was to define the thresholds for determining probable presence of MDE, thus exploring diagnostic validity.

Univariate Logistic Regression Analysis

When measuring the predictive ability of each measure independent from the others, the QIDS-A-C(Composite) was the best estimate of all four measures, possibly because it pooled interviews from both adolescent and parent. Interestingly, the CDRS-R was the poorest estimate. Although this suggests that the CDRS-R is the least predictive of depression compared to the QIDS-A measures, comparison of the chi squares indicated that the difference is minimal. As noted in the ANOVA results, the ANOVA analysis found the CDRS-R to be the most predictive. Results often vary with the type of test used, which is why a variety of tests were run to examine the discriminating abilities of each measure.

Multivariate Logistic Regression Analysis

When all four measures were used collectively as predictor variables, the CDRS-R was the only significant estimate. This did not imply that the QIDS-A measures were not predictive, simply that they are so similar that they wash out any significance among each other when examined together in this manner. This demonstrates that giving all four measures to a patient would not provide enough unique information regarding depression to make it useful, as opposed to using one or two of the measures.

ROC Analysis

ROC analysis found the CDRS-R and the QIDS-A-C(Composite) to be the most sensitive in discriminating depression overall. This is not surprising since both measures compile data from both the adolescent and the parent, thus maximizing the information contributing to detection of depression. Although the remaining two QIDS-A measures differed slightly in sensitivity at various points, they were very similar (and robust) in overall performance, based on the area under the curve. This indicates that both the QIDS-A-C(Adolescent) and QIDS-A-SR show a high level of accuracy in discriminating depressed patients from non-depressed patients. This supports the case that the QIDS-A-SR is accurate enough for the purposes of correctly classifying a depressed patient. As such, the QIDS-A-SR could be used as a simple self-report screening measure to increase detection for depression in clinic settings while minimizing time and staff burden and maximizing clinician efficiency.

Analysis of Variance (ANOVA)

The ANOVA results further substantiated that all four measures significantly discriminated between the depressed and non depressed groups. The CDRS-R was found to be the most predictive using ANOVA, although it was the least predictive using univariate logistic regression (as noted previously). While this difference was minimal, it is important to note because it indicates that all four measures are similar enough in predictive ability that the order varies depending on the type of analysis, such that the order is likely inconsequential. Thus, each measure shows definitive discriminative validity.

Multivariate Analysis of Variance (MANOVA)

The MANOVA results, although similar to the multivariate logistic regression, found the QIDS-A-SR to be slightly more predictive than the QIDS-A-C(Composite). This suggests that the QIDS-A-SR has more incremental validity, however slight, which further supports the independent use of the QIDS-A-SR to discriminate depression. Generally, the results confirmed the overall diagnostic validity that was found in the other analyses as well. This is not surprising, as these various analyses are in a sense asking the same question in slightly different ways. Given the similarities of the scales and good psychometric properties, it is expected that they would produce similar "answers" to these questions.

The diagnostic validity results of the QIDS-A_{17} found in the current study compare favorably to the results of a recent study that used similar analyses for the adult QIDS-SR_{16} in an adult population (Bernstein et al., in press). The Bernstein et al. study compared the QIDS-SR_{16} to two other adult self-report rating scales for depression, using a structured diagnostic measure to classify depressed and non-depressed patients. The QIDS-SR_{16} was found to be the most valid in the Bernstein et al study, such that it was recommended above the other measures for utility in private practice settings. As such, the definitive validity of the QIDS-A_{17} measures found in the current study supports similar utility in an adolescent population. Furthermore, the age range of the current study (ages 8 to 17) indicates that the QIDS-A_{17} can be used accurately in a wide age range, thus further increasing its value. In general, it is promising that the QIDS-A_{17} measures showed similarly strong discriminative validity as the adult $QIDS_{16}$, as this

further substantiates the similarity and utility of the QIDS measures across formats as well as age groups.

A third aim of the study was to determine whether the QIDS-A-SR or the QIDS-A-C(Adolescent) is sufficient to replace the more time consuming QIDS-A-C(Composite) or CDRS-R.

All four measures, the CDRS-R, QIDS-A-C(Composite), QIDS-A-C(Adolescent), and QIDS-A-SR, showed sound psychometric properties in this population, including reliability, validity, sensitivity, and specificity. The CDRS-R and QIDS-A-C(Composite) correlated the highest with each other and showed similar sensitivity and specificity, which was not surprising since both measures incorporate information from both the adolescent and parent. The QIDS-A-C(Adolescent) was generally the next in overall performance, correlating highly with the QIDS-A-C(Composite) and CDRS-R, and showing good discrimination comparable to the QIDS-A-C(Composite), particularly at moderate levels of depression. The strong overall performance of the QIDS-A-C(Adolescent) and the favorable comparisons to the two composite measures (QIDS-A-C[Composite] and CDRS-R) indicate that the QIDS-A-C(Adolescent) could sufficiently replace either of the composite measures.

More importantly, the QIDS-A-SR showed satisfactory reliability, validity and correlation with the other measures. The QIDS-A-SR was not far from the QIDS-A-C(Adolescent) in performance and demonstrated similar sensitivity and specificity,

particularly at lower and higher levels of depression. It was not surprising that the QIDS-A-SR was slightly less reliable than the other measures since these other measures had the advantage of clinician experience and judgment to help probe for salient information. In spite of this difference, the QIDS-A-SR consistently performed at levels indicative of a reliable and valid measure. Additionally, the QIDS-A-SR was slightly more discriminating than the CDRS-R and the QIDS-A-C(Adolescent) at very low levels of depression. Based on this evidence, the QIDS-A-SR is suitable for use on its own.

Furthermore, it is important to consider the intended purpose of the QIDS-A measures. The QIDS-A is intended to measure the amount of symptomatology (i.e., severity and frequency of symptoms) an individual is experiencing related to a major depressive episode, based on the nine DSM-IV symptom domains for depression. This is not a structured diagnostic instrument, but rather a measure of depressive symptomatology that may also be used to screen for the likely presence of an MDE, based on severity and frequency of symptoms. As such, it is not essential that the QIDS-A measures be perfectly sensitive in classifying depressed from nondepressed. What is important is that it shows good psychometric properties in measuring the symptoms of depression, as results from this study demonstrate.

Given the intended purpose of the QIDS-A measures, the satisfactory psychometric properties of the QIDS-A-SR, and the ease and affordability of the self report format, it may be concluded that the QIDS-A-SR is sufficient to use in place of the clinician interview of adolescent (QIDS-A-C[Adolescent]), and especially over the composite versions (QIDS-A-C[Composite] and CDRS-R), particularly in environments where time, staffing and health care costs are an issue.

Methodological Considerations

This study has a number of limitations. The sample size was small, particularly for statistical analyses that involve estimating many parameters, such as IRT and factor analysis. Thus, related results should be confirmed in larger samples. In addition, the sample was representative of a pediatric psychiatric outpatient clinic population within a university hospital setting. As such, these findings may not generalize to other clinic settings. On the other hand, the liberal inclusion criteria, in terms of comorbid disorders and use of any medications, increase the likelihood that this population may be representative of other clinical settings. By including the subset of subjects from the depression research study, this sample likely overrepresented the percent of patients that would typically present with MDE in an outpatient clinic. However, the increased percentage of depressed subjects allowed for greater comparison between the depressed and nondepressed subjects.

The diagnosis of MDE was not obtained from a structured diagnostic interview, but rather from a checklist of DSM-IV symptoms required to meet criteria for a diagnosis of MDE, filled out by a physician. Furthermore, the MDE checklist and the depression rating scales were occasionally completed by the same physician, such that the rater was not always blind to the diagnosis. Additionally, since the QIDS-A-C measures and the CDRS were always completed by the same evaluator, they had data from both measures to guide their scores, which may have increased the correlation between these measures.

Due to the variety of clinicians assisting on the study, the order of administration was not randomized. The order of measures could not be monitored when completed by

a clinician other than the study coordinator. The study coordinator always completed the QIDS-A-SR first, then the QIDS-A-C and CDRS measures (adolescent first, then parent). For data gathered by other IRB-approved clinicians, the adolescent was almost always interviewed first.

Although every effort was made to recruit consecutive outpatients for participation in the study, privacy laws and limited study manpower restricted this endeavor. Privacy rules did not allow research personnel to directly solicit potential subjects, so that the burden of initial recruitment was placed on a member of the patient's health care team. At times these team members forgot to mention the research study, particularly when the clinic was quite busy, resulting in missed potential subjects. It is also possible that the manner in which the different team members presented the study either positively or negatively affected the patient's decision to participate, although this was not monitored. Occasionally, multiple patients expressed interest in participating in the study at the same time. If there were not enough research coordinators available, some of the patients agreed to participate at a later date, either during a separate research appointment or before or after the next clinic appointment. Very few of these patients followed through or responded to phone calls.

Clinical Implications

A primary finding of this study is that all three QIDS-A measures demonstrated strong psychometric properties in this population, indicating that the QIDS-A is an appropriate measure for depression in adolescents. Furthermore, results indicate that all three of the QIDS-A measures can be used adequately in place of the CDRS-R, the

current field standard, depending on the purpose. For example, the three QIDS-A measures were clearly comparable enough to the CDRS to be used effectively for symptom measurement and depression screening. On the other hand, the CDRS-R might be used over the QIDS-A measures in a precise research study since the CDRS-R was more sensitive to detecting slight differences in level of depression. However, all four measures demonstrated comparable discriminative validity, further supporting the use of the QIDS-A. Since the QIDS-A is available at no cost and covers all nine DSM-IV domain symptoms for depression, the validation of this measure meets an established need. More importantly, the self report (QIDS-A-SR) has acceptable psychometric properties and is the most time and cost effective, making it a realistic, effective and useful option in clinical practice as well as research environments.

Results from this study provide strong evidence for the use of the QIDS-A-SR. Although the QIDS-A-SR was slightly less reliable than the clinician measures, it still demonstrated satisfactory psychometric properties and diagnostic utility. If one balances the slight sacrifice in reliability with the need for this type of affordable, valid, efficient tool, particularly given rising health care costs, the loss is minimal. Considering that it is reliable, valid, free, easily available on the internet, and only takes 5 to 7 minutes to complete, the QIDS-A-SR would be particularly useful in busy clinical environments, such as a pediatrician's office.

Another important finding is that the addition of the irritability item to the QIDS-A_{17} (as opposed to the $QIDS_{16}$ that does not measure irritability) did not make much of a difference in the performance of the rating scales. Irritability is an important symptom diagnostically in adolescents, but it appears to have relatively no impact on the reliability

of the QIDS scale overall to measure depression. Thus, this study suggests that the $QIDS_{16}$ versions that do not include an irritability item are not lacking in measuring depression in adolescents. In fact, it is possible that the $QIDS_{16}$ version may be acceptable for use in adolescents, possibly eliminating the need for a separate adolescent version.

Issues for Future Research

Several areas indicate further exploration. As noted above, additional research should explore the performances of the $QIDS_{16}$ and QIDS-A_{17} in adolescent populations to determine whether separate adult and adolescent versions are necessary. Age should also be examined to determine the minimum age for which the QIDS-A_{17} can be used effectively.

Furthermore, this study did not explore the ability of the QIDS-A_{17} measures to reflect sensitivity to change in symptoms, which is important in monitoring response to treatment, both in research and clinically. The adult $QIDS_{16}$ measures, including the QIDS-SR_{16}, have shown satisfactory sensitivity to change in determining treatment response and remission (Brown et al., 2008; Rush et al., 2006). Given that the adult $QIDS_{16}$ and the QIDS-A_{17} are almost identical in structure and both demonstrate acceptable psychometric properties, it is likely that the QIDS-A_{17} measures would also show sensitivity to symptom change. However, additional research is needed to evaluate this potential use for the QIDS-A_{17} measures.

Similarly, future research should examine the ability of the QIDS-A_{17} measures to indicate remission as compared to the CDRS-R to determine agreement and

equivalency scores for remission between the two measures. This would provide additional information by which to examine the extent that the QIDS-A_{17} may effectively replace the CDRS-R in various settings.

APPENDIX A

Consent Form

The University of Texas Southwestern Medical Center at Dallas
Parkland Health & Hospital System
Children's Medical Center
Texas Scottish Rite Hospital for Children

CONSENT TO PARTICIPATE IN RESEARCH

Title of Research:	Improving Depression Measurement in Adolescents
Funding Agency/Sponsor:	UT Southwestern Medical Center
Study Doctors:	Beth Kennard, Psy.D. A. John Rush, M.D. Graham J. Emslie, M.D. Ira Bernstein, Ph.D. Carroll W. Hughes, Ph.D. Paul Croarkin, M.D.
Research Personnel:	Charlotte Haley, B.A. Hayley Evans, B.A. Krystle Joyner, M.S. Alex Simmons, M.S. Tabatha Hines, B.A. Jessica Jones, M.A. Taryn L. Mayes, M.S. Jarrette Moore, M.A. Annie Walley, L.M.S.W.

You may call these study doctors or research personnel during regular office hours at 214-648-4333. At other times, you may call them at 214-648-5555.

Note: If you are a parent or guardian of a minor and have been asked to read and sign this form, the "you" in this document refers to the minor.

Instructions:

Please read this consent form carefully and take your time making a decision about whether to participate. As the researchers discuss this consent form with you, please ask him/her to explain any words or information that you do not clearly understand. The purpose of the study, risks, inconveniences, discomforts, and other important information about the study are listed below. If you decide to participate, you will be given a copy of this form to keep.

Why is this study being done?

Many teens have some symptoms of depression, but these symptoms may go unrecognized, even by healthcare providers. This study is being done to see if a questionnaire can be used as a standard measure for depression in a wide range of adolescent patients.

This research is being done because many questionnaires and assessment scales used to monitor symptoms of depression (both improvement and worsening) are often difficult to administer and may not be reliable in adolescents.

Why am I being asked to take part in this research study?

You are being asked to take part in this study because you are seeking psychiatric care. You may or may not have any symptoms of depression. Medical research involves offering a plan of care to a group of patients, collecting and studying information about each patient's experience, and using that information to develop the best possible care for future patients.

How many people will take part in this study?

About 200 people will take part in this study through UT Southwestern/Children's Medical Center.

What is involved in the study?

If you agree to be in this study, you will be asked to sign this consent form and will have the following tests and procedures.

Procedures and Evaluations during the Research:

You and your parent will be interviewed by a trained evaluator. This evaluator will ask you and your parent questions about your current medications and psychiatric symptoms using questionnaires and an interview. You and your parent will be asked to complete several brief paper and pencil tests. The assessments will be conducted one time only, and will take approximately 60 minutes to complete. In addition, your clinic physician will ask questions about depression during your normal visit.

How long can I expect to be in this study?

Your participation in this study involves a one-time visit that will last approximately 60 minutes. Upon completion of all data during this visit, there will be no further contact related to this study. All efforts will be made to collect the study data at the time of this visit. If for some reason all data is not collected during this visit, we will gather the remaining data over the telephone. You can choose to stop participating for any reason at any time.

What are the risks of the study?

Psychological Stress

Some of the questions we will ask you as part of this study may make you feel uncomfortable. You may refuse to answer any of the questions, take a break or stop your participation in this study at any time.

Loss of Confidentiality

Any time information is collected; there is a potential risk of loss of confidentiality. Every effort will be made to keep your information confidential; however, this cannot be guaranteed.

What are the possible benefits of this study?

If you agree to take part in this study, there may not be direct benefits to you. The researchers cannot guarantee that you will benefit from participation in this research. However, you will receive an evaluation specifically designed to examine depressive symptoms. The information obtained through this study will be provided to your physician at Children's Medical Center who will be able to help you with treatments to reduce these symptoms.

We hope the information learned from this study will benefit others with depression in the future. Information gained from this research could lead to better recognition of depression in adolescents.

What options are available if I decide not to take part in this research study?

This is not a treatment study. You do not have to be part of it to get treatment for your condition.

Will I be paid if I take part in this research study?

Yes. You will be given a $25.00 gift card to Target at the end of the study if you take part in this research.

There are no funds available to pay for parking expenses, transportation to and from the research center, lost time away from work and other activities, lost wages, or child care expenses.

Will my insurance provider or I be charged for the costs of any part of this research study?

No. Neither you, nor your insurance provider, will be charged for anything done only for this research study (i.e., the Screening Procedures, Experimental Procedures, or Monitoring/Follow-up Procedures described above).

However, the standard medical care for your condition (care you would have received whether or not you were in this study) is your responsibility (or the responsibility of your insurance provider or governmental program). You will be charged, in the standard manner, for any procedures performed for your standard medical care.

What will happen if I am harmed as a result of taking part in this study?

It is important that you report any illness or injury to the research team listed at the top of this form immediately.

Compensation for an injury resulting from your participation in this research is not available from the University of Texas Southwestern Medical Center at Dallas or Children's Medical Center at Dallas.

You retain your legal rights during your participation in this research.

Can I stop taking part in this research study?

Yes. If you decide to participate and later change your mind, you are free to stop taking part in the research study at any time.

If you decide to stop taking part in this research study, it will not affect your relationship with the UT Southwestern staff or doctors. Whether you participate or not will have no effect on your legal rights or the quality of your health care.

If you are a medical student, fellow, faculty, or staff at the Medical Center, your status will not be affected in any way.

Your doctor is a research investigator in this study. S/he is interested in both your medical care and the conduct of this research study. At any time, you may discuss your care with another doctor who is not part of this research study. You do not have to take part in any research study offered by your doctor.

Will my information be kept confidential?

Information about you that is collected for this research study will remain confidential unless you give your permission to share it with others, or if we are required by law to release it. You should know that certain organizations that may look at and/or copy your medical records for research, quality assurance, and data analysis include:

- UT Southwestern Medical Center
- Representatives of government agencies, like the Food and Drug Administration (FDA), involved in keeping research safe for people, and
- The UT Southwestern Institutional Review Board.

In addition to this consent form, you will be asked to sign an "Authorization for Use and Disclosure of Protected Health Information." This authorization will give more details about how your information will be used for this research study, and who may see and/or get copies of your information.

Whom do I call if I have questions or problems?

For questions about the study, contact Charlotte Haley at 214-648-4333 during regular business hours and at 214-648-5555 after hours and on weekends and holidays.

For questions about your rights as a research participant, contact the UT Southwestern Institutional Review Board (IRB) Office at 214-648-3060.

SIGNATURES:

YOU WILL BE GIVEN A COPY OF THIS CONSENT FORM TO KEEP.

Your signature below certifies the following:

- You have read (or been read) the information provided above.
- You have received answers to all of your questions and have been told who to call if you have any more questions.
- You have freely decided to participate in this research.
- You understand that you are not giving up any of your legal rights.

_____

Participant's Name (printed)

_____ _____

Participant's Signature Date

_____

Legally authorized representative's Name (printed)

_____ _____

Legally authorized representative's Signature Date

_____

Name of person obtaining consent (printed)

_____ _____

Signature of person obtaining consent Date

ASSENT OF A MINOR:

I have discussed this research study with my parent or legal guardian and the researchers, and I agree to participate.

_____ _____

Signature of participant (age 8 through 17) Date

APPENDIX B

Measures

MDE Checklist

Study ID Number:_____ **Today's Date:____/_____/____**

Physician Name: _____

MDE DIAGNOSTIC CHECKLIST

Current Major Depressive Episode (MDE)?

☐ Definite ☐ Probable ☐ No

Check ALL symptoms that apply:

☐ 1a. Depressed mood most of the day, nearly every day

☐ 1b. Irritable mood most of the day, nearly every day

☐ 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day

☐ 3. Significant weight loss, weight gain, or change in appetite nearly every day

☐ 4. Insomnia or hypersomnia nearly every day

☐ 5. Psychomotor agitation or retardation nearly every day (observable by others)

☐ 6. Fatigue or loss of energy nearly every day

☐ 7. Feelings of worthlessness or excessive or inappropriate guilt nearly every day

☐ 8. Diminished ability to concentrate, or indecisiveness, nearly every day

☐ 9. Recurrent thoughts of death, suicidal ideation or a suicidal attempt or plan

For a diagnosis of MDE, one of first 3 symptoms must be present. Five total symptoms must be present, lasting for at least 2 weeks. Differential diagnoses must be ruled out, and symptoms must cause significant distress or impairment.

APPENDIX C

Tables

Table 1. *Summary of current measures of depression for adolescents*

Scale	Type	Ages	Time needed	Reliability	Validity	Critique	Availability
BDI	Self-report	adolescents	5-10 min	IC: 0.79 to 0.91	Concurrent: 0.49 to 0.73	High false positives; cognitive symptom bias	Cost
CDI	Self-report (parent version)	7-18	10-20 min	IC: 0.59 to 0.88 TR: 0.38 to 0.87	Concurrent: moderate to high Convergent: moderate to high Discriminant: poor to variable	Poor construct validity; better for children	Cost
CDRS-R	Clinician interview	6-12; also used in adolescents	30-45 min	IC: "adequate" IR: 0.80 to 0.96 TR: 0.81	Concurrent: 0.75 to 0.92	Overpredicts depression; over-inclusion of somatic symptoms	Cost
CES-D & CES-DC	Self-report (parent version)	children/ adolesents	5-10 min	IC: 0.75 to 0.89 TR: 0.51 to 0.57	Discriminant: poor to low moderate	High false positives; poor performance with children	Free of charge
HRSD	Clinician interview	adults/ adolescents	10-30 min	IC: 0.90 IR: "excellent"	Concurrent: 0.56	Over-emphasis on somatic symptoms; limited data in adolescents	Free of charge
MADRS	Clinician interview	adults/ adolescents	10-20 min	Good in adults; very limited data in adolescents	Good in adults; very limited data in adolescents	Very limited data in adolescents; lacks somatic and psychomotor items	Free of charge
MFQ & SMFQ	Self-report	8 to 18	5-10 min	IC: 0.84 TR: 0.80	Concurrent: parent version excellent; child version good Discriminant: good	SMFQ does not assess suicidality	Free of charge (with permission from authors)
PHQ-A	Self-report	13-18	10-15 min	No data	MDD: 73% (sens), 94% (spec)	Extensive scoring; no parent version; no validation to other instruments	Free of charge
RADS-2	Self-report (parent version)	11-20	5-10 min	IC: 0.92 to 0.96 TR: 0.80 to 0.86	Concurrent: 0.70 to 0.89 Convergent: 0.70 to 0.89	Limited clinical data; limited sensitivity	Cost

Note. IC = internal consistency; IR = inter-rater reliability; TR = test-retest reliability; BDI = Beck Depression Inventory (Beck & Steer, 1993); CDI = Children's Depression Inventory (Kovacs, 1992); CDRS-R = Children's Depression Rating Scale--Revised (Pozanski & Mokros, 1999); CES-D = Center for Epidemiological Studies--Depression scale (Radloff, 1977); CES-DC = Center for Epidemiological Studies--Depression scale-Child version (Weismann et al., 1980); HRSD = Hamilton Rating Scale for Depression (Warren, 1997); MADRS = Montgomery-Asberg Depression Rating Scale (Montgomery & Asberg, 1979); MFQ = Mood and Feelings Questionnaire (Angold et al., 1995); SMFQ = short version of the MFQ; PHQ-A = Patient Health Questionnaire for Adolescents (Johnson et al., 2002); RADS-2 = Reynolds Adolescent Depression Scale, Second Edition (Reynolds, 2002); From Brooks & Kutcher, 2001; Hughes & Melson, 2008; Myers & Winters, 2002; Pavuluri & Birmaher, 2004

90

Table 2. *Demographic and Clinical Characteristics of Study Participants*

	All Subjects (N = 103)
Mean age ± SD	13.83 ± 2.4
Age group	
Child (≤11 years)	21 (20.0%)
Adolescent (≥12 years)	82 (80.0%)
Gender	
Male	51 (49.5%)
Female	52 (50.5%)
Race/Ethnicity	
Caucasian	73 (70.9%)
African-American	16 (15.5%)
Asian	1 (1.0%)
American Indian	3 (2.9%)
Hispanic	10 (9.7%)
Annual Family Income (n = 98)	
Under $15,000	9 (0.9%)
$15,000-$35,000	20 (20.0%)
$35,000-$75,000	34 (34.7.0%)
Over $75,000	35 (35.7%)
Presence of MDE	
Definite presence	40 (38.8%)
Probable presence	8 (7.8%)
No presence	55 (53.4%)
Mean CGI-S ± SD	3.78 ± 1.2
Mean C-GAS ± SD	55.92 ± 10.4

Table 3. *CTT Analysis of the QIDS-A-C_{17}(Adolescent), QIDS-A-C_{17}(Composite), and QIDS-A-SR_{17}*

	QIDS-A-C_{17} (Adolescent) (n = 101)		QIDS-A-C_{17} (Composite) (n = 102)		QIDS-A-SR_{17} (n = 102)	
Domain	Mean	$\underline{r}_{it}$	Mean	$\underline{r}_{it}$	Mean	$\underline{r}_{it}$
Sleep	1.74	.52	1.82	.47	2.10	.28
Sad or Irritable Mood	1.51	.70	1.78	.66	1.37	.61
Appetite	1.04	.28	.99	.35	1.41	.33
Concentration/Decision Making	1.12	.55	1.25	.62	1.07	.50
Self View	1.05	.67	1.27	.65	.99	.52
Thoughts of Death or Suicide	.36	.49	.37	.44	.39	.41
General Interest	.72	.68	.92	.71	.78	.54
Energy Level	.91	.63	.99	.64	.90	.51
Restlessness/Agitation	1.00	.45	1.14	.48	1.19	.53
Scale Mean	9.46		10.55		10.21	
Scale SD	5.68		5.67		5.09	
α	.84		.84		.78	

Table 4. *CTT Analysis of the $CDSR_{17}$*

	(n = 103)	
Item	Mean	r_{it}
Impaired schoolwork	2.95	.66
Difficulty having fun	2.69	.81
Social withdrawal	2.56	.75
Appetite disturbance	2.88	.47
Sleep disturbance	2.20	.64
Excessive fatigue	2.75	.71
Physical complaints	1.95	.50
Irritability	3.50	.40
Excessive guilt	1.84	.61
Low self-esteem	3.14	.72
Depressed feelings	2.88	.80
Morbid ideas	1.43	.38
Suicidal ideas	1.72	.58
Excessive weeping	2.21	.68
Depressed facial affect	1.94	.61
Listless speech	1.43	.53
Hypoactivity	1.45	.62
Scale Mean	39.53	
Scale SD	15.72	
α	.92	

Table 5. *Observed intercorrelations among measures (above diagonal), coefficients alpha reliabilities (diagonal), and disattenuated intercorrelations among measures (below diagonals)*

	QIDS-A-C_{17} (Adolescent)	QIDS-A-SR_{17}	QIDS-A-C_{17} (Composite)	CDRS-R
QIDS-A-C_{17} (Adolescent)	**.84**	.69	.88	.78
QIDS-A-SR_{17}	*.85*	**.78**	.66	.63
QIDS-A-C_{17} (Composite)	*>1*	*.82*	**.84**	.89
CDRS-R	*.89*	*.74*	*>1*	**.92**

Note. Bolded values are coefficients alpha; Italicized values are corrected (disattenuated) correlations

Table 6. *Dimensionality: Obtained and Simulated Eigenvalues for the QIDS-A-C(Adolescent), QIDS-A-C(Composite), QIDS-A-SR, and CDRS-R*

		Obtained Eigenvalues		
Component	Simulated Eigenvalues	**QIDS-A-C (Adolescent)**	**QIDS-A-C (Composite)**	**QIDS-A-SR**
1	1.48	4.05	4.08	3.40
2	1.31	1.13	1.09	1.21
3	1.20	1.00	.98	1.03

		Obtained Eigenvalues
Component	Simulated Eigenvalues	**CDRS-R**
1	1.79	7.83
2	1.62	1.38
3	1.49	1.12

Table 7. *Effect Sizes for Diagnostic Validity Analyses*

	Univariate		Multivariate	
	Logistic Regression	ANOVA	Logistic Regression	MANOVA
Measure	Estimate	Model/Corrected Sum of Squares	Estimate	Weight on Discriminant Axis
QIDS-A-C (Adolescent)	.3258**	.4029**	-.0286	-.0013
QIDS-A-C (Composite)	.4628**	.5119**	.0060	.0042
QIDS-A-SR	.3176**	.3351**	.1688	.0061
CDRS-R	.2317**	.5812**	.2306*	.0077

* $p < .001$. ** $p < .0001$

Table 8. *Threshold scores, sensitivities, and specificities at four levels of severity for the QIDS-A-C(Adolescent), QIDS-A-C(Composite), QIDS-A-SR, and CDRS-R*

	QIDS-A-C (Adolescent)			**QIDS-A-C (Composite)**			**QIDS-A-SR**			**CDRS-R**		
Level	***Thresh***	***Sens***	***Spec***	***Thresh***	***Sens***	***Spec***	***Thresh***	***Sens***	***Spec***	***Thresh***	***Sens***	***Spec***
Low	5	.38	1.00	6	.33	1.00	6	.31	.96	28	.31	1.00
Medium	8	.56	.96	9	.53	.96	8	.51	.93	33	.51	1.00
High	12	.76	.87	12	.71	.91	10	.73	.87	40	.71	.98
Very High	17	.93	.33	15	.96	.73	15	.91	.45	49	.91	.75

Note. Thres = threshold; Sens = sensitivity; Spec = specificity

Table 9. *Depressed and Nondepressed: Total Score Means and Standard Deviations*

	Depressed N = 45		Nondepressed N = 50	
	M	(SD)	M	(SD)
QIDS-A-C(Adolescent)	13.3	(5.5)	6.2	(3.3)
QIDS-A-C(Composite)	15.0	(4.5)	6.8	(3.5)
QIDS-A-SR	13.4	(4.6)	7.5	(3.8)
CDRS-R	52.2	(13.0)	28.4	(7.1)

Table 10. *Equated Scale Scores on the QIDS-A-C(Adolescent), QIDS-A-C(Composite), QIDS-A-SR, and CDRS-R*

Θ Range	QIDS-A-C (Adolescent)	QIDS-A-C (Composite)	QIDS-A-SR	CDRS-R
-2.00/-2.10		0	0	17
-1.80/-1.90	0	1	1	18
-1.50/-1.60	1	2	2	19
-1.30/-141	2	3	3	20-21
-1.00/-1.10	3	4	4	22-23
-.82/-.92	4	5	5	24-25
-.67/-.73	5	6		26
-.48/-.66	6	7	6	27-29
-.30/-.46	7	8	7	30-31
-.13/-.28	8	9	8	32-34
-.10/.03	9	10	9	35-36
.04/.20	10	11	10	37-40
.23/.36	11	12	11	41-43
.39/.52	12	13	12	44-46
.54/.62	13	14	13	47-48
.67/.77	14	15	14	49-51
.83/.97	15	16	15	52-54
.98/1.00	16	17	16	55-56
1.10/1.20	17	18	17	57-59
1.30	18	19	18	60-61
1.40/1.50	19	20	19	62-64
1.60	20		20	65-66
1.70/1.80	21	21	21	67-69
1.90/2.00	22	22	22	70-71
2.10/2.20	23	23	23	72-73
2.30/2.40	24	24	24	74-75
2.50/2.60	25	25	25	76-77
2.70/2.80	26	26	26	78
2.90/3.20	27	27	27	79-81

APPENDIX D

Figures

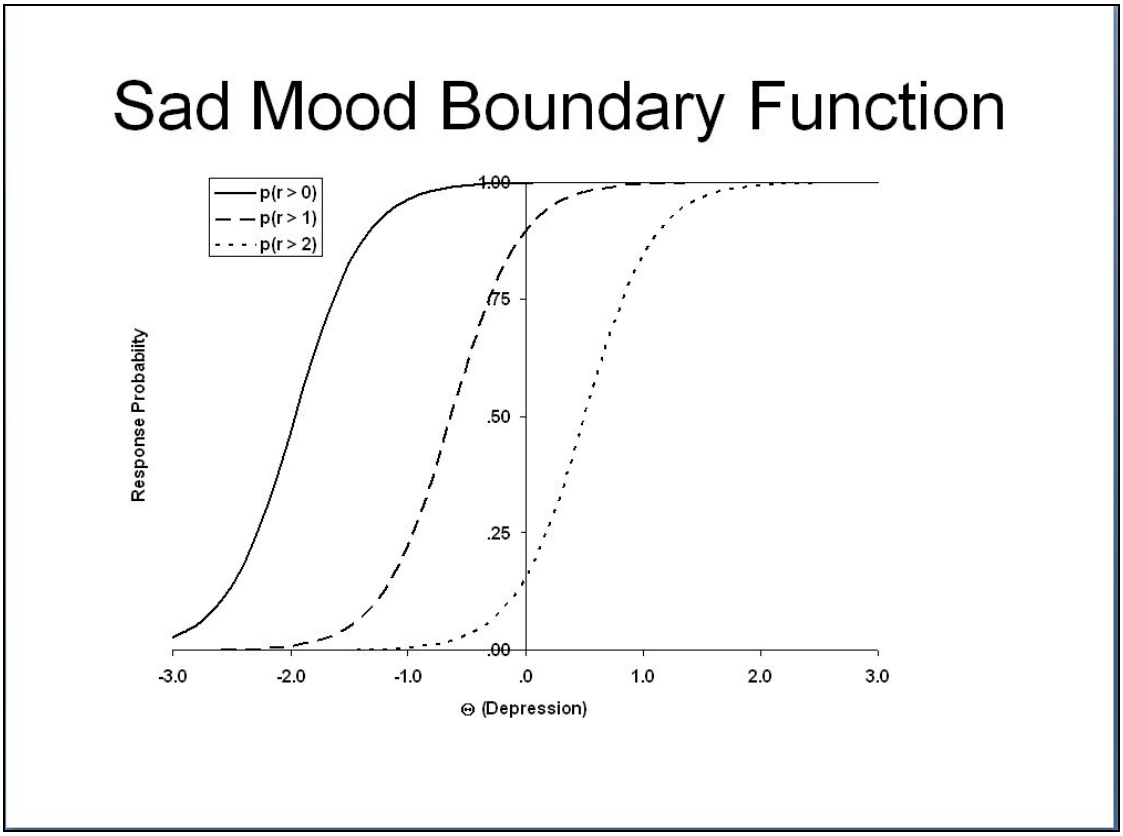
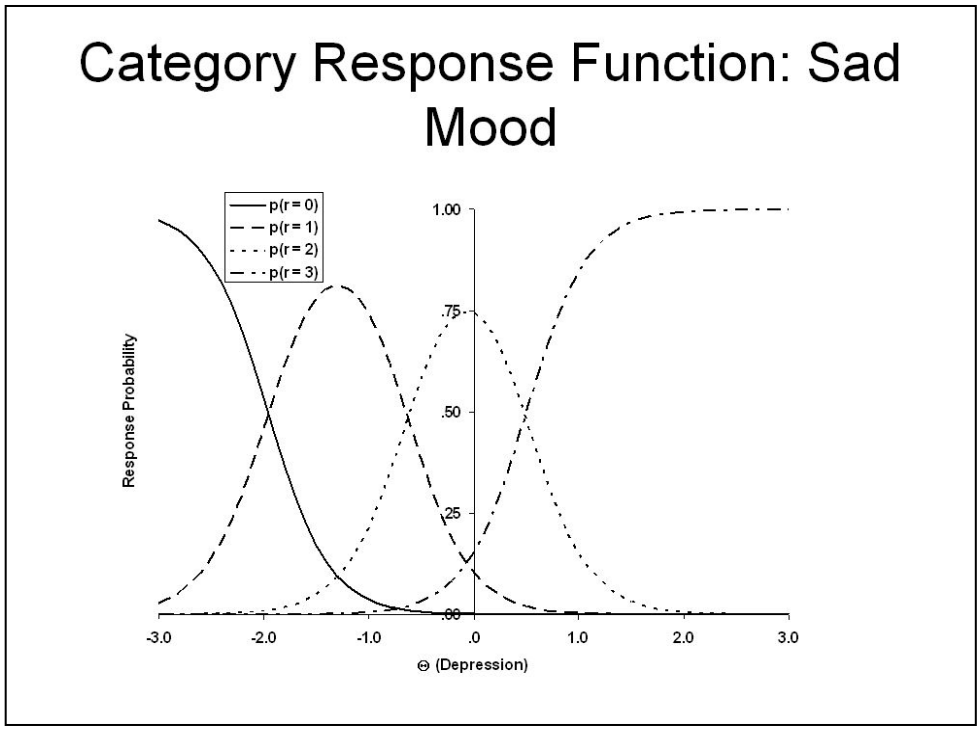


Figure 1. *Generic example of boundary function curves for sad mood item*



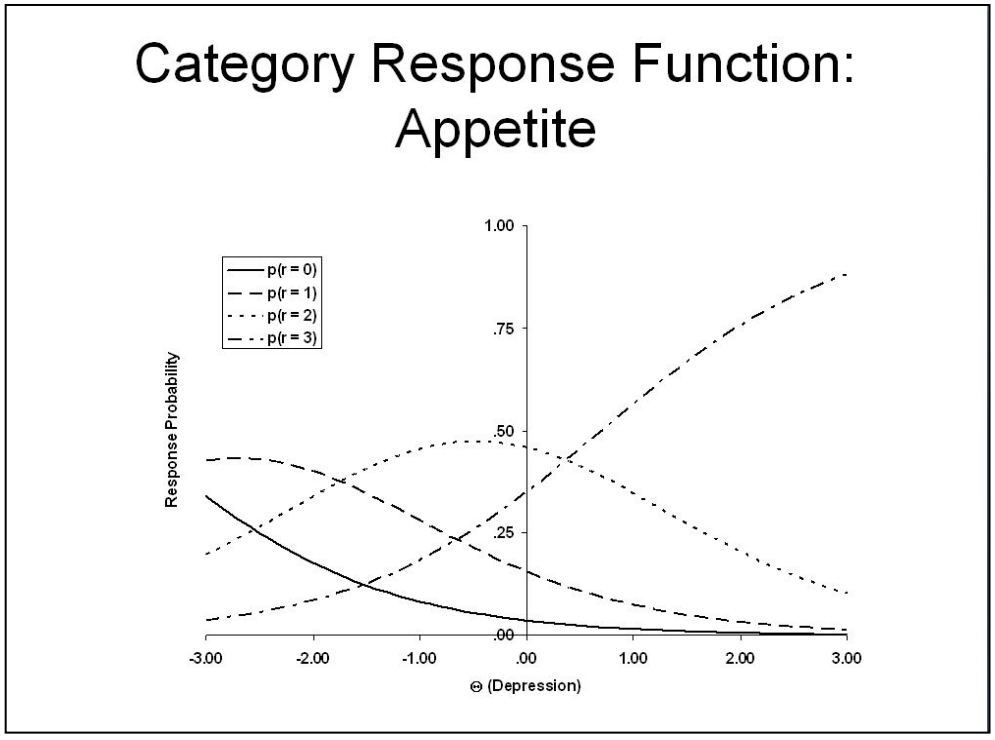


Figure 2. *Generic example of category response functions for two distinct items, sad mood and appetite*

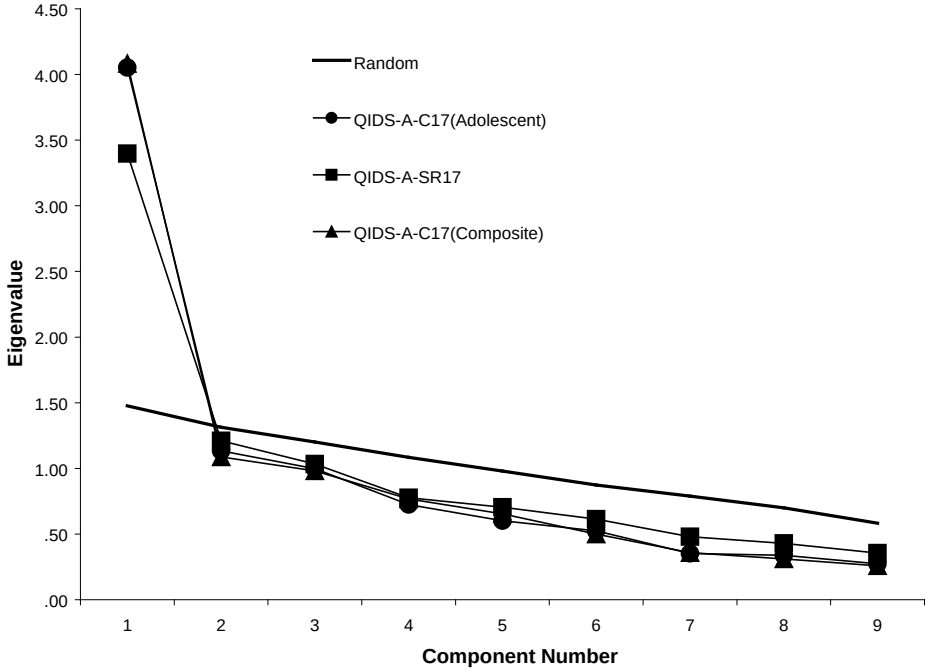


Figure 3. *Scale Dimensionalities: QIDS-A$_{17}$ Scree*

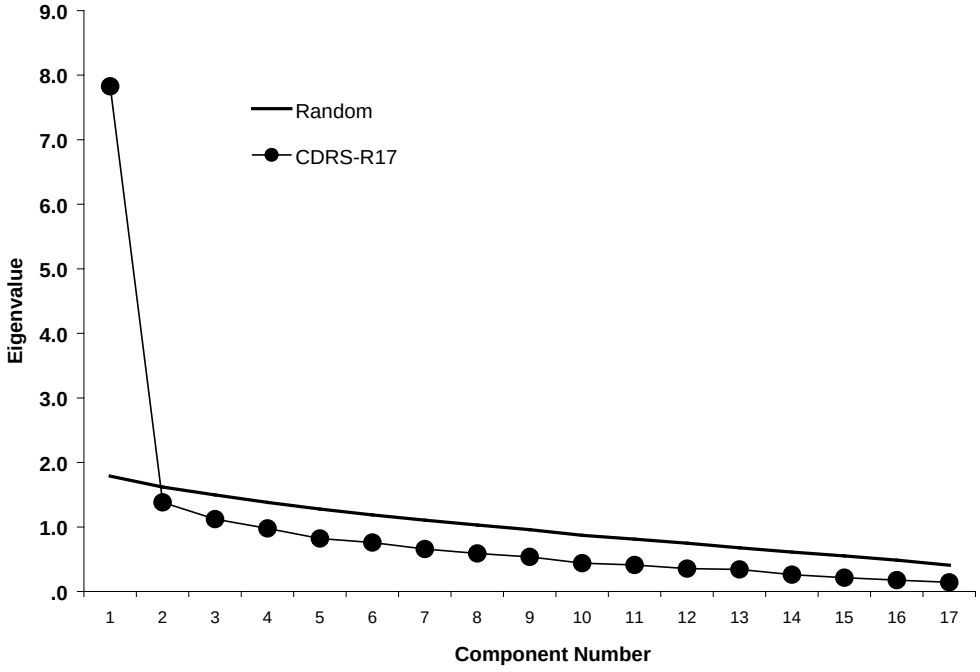


Figure 4. *Scale Dimensionalities: CDRS-R Scree*

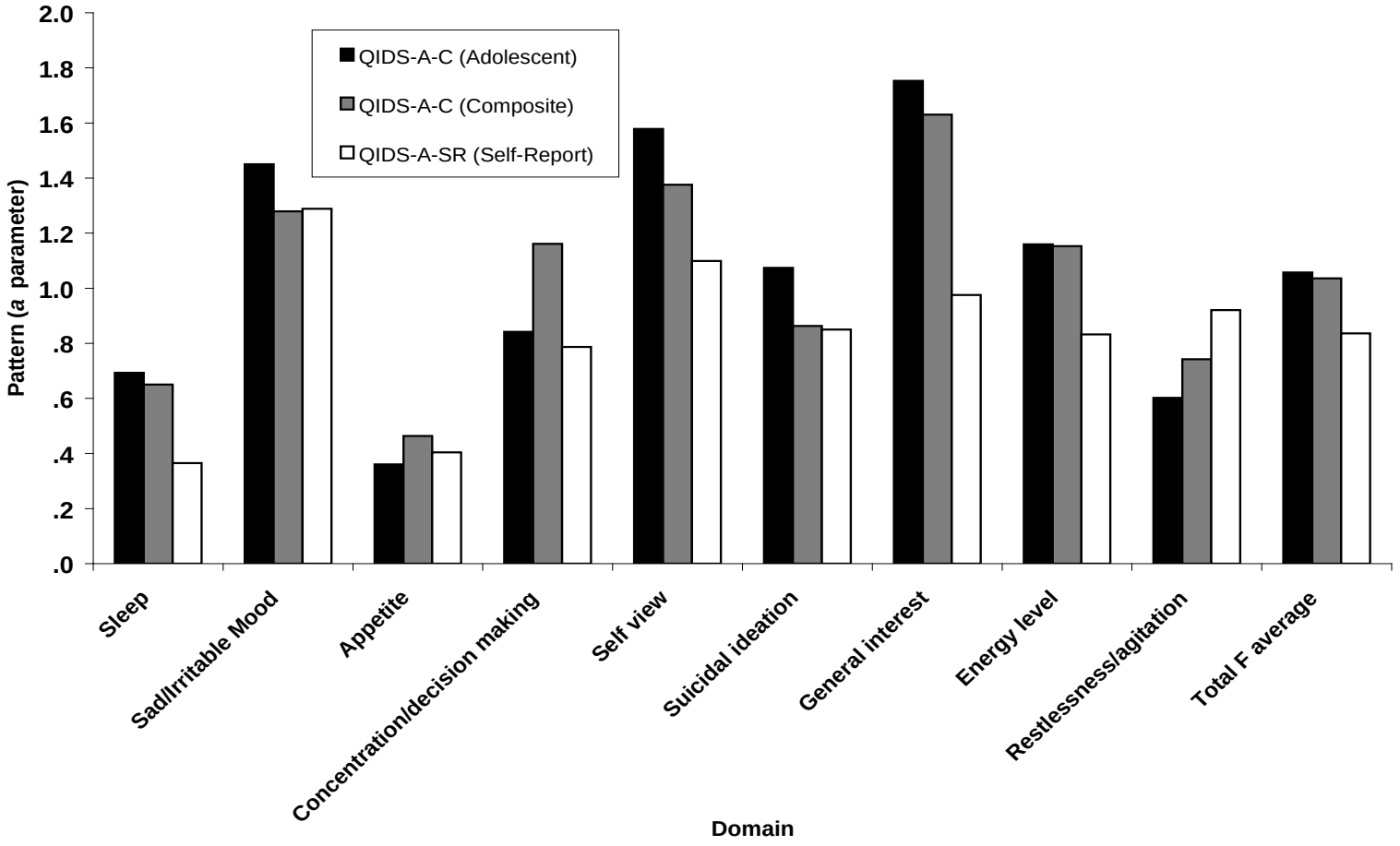
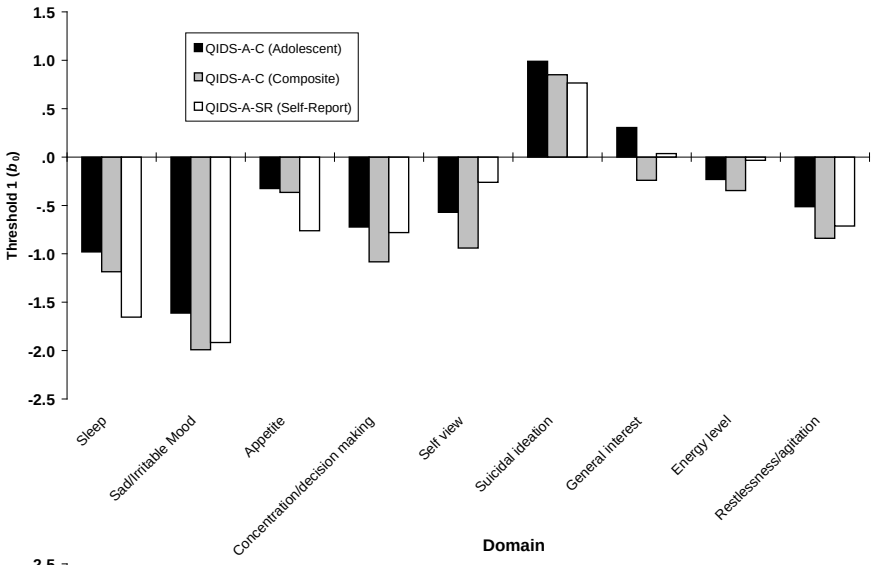
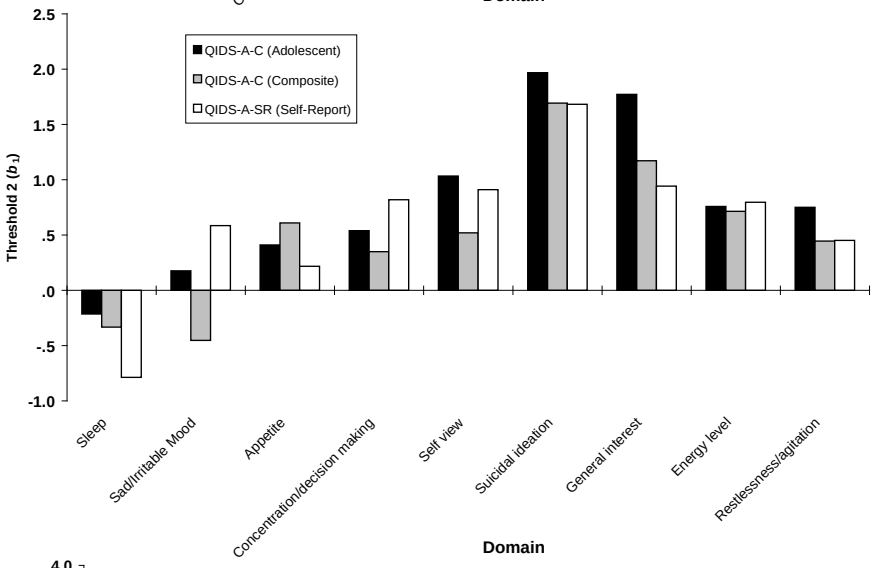


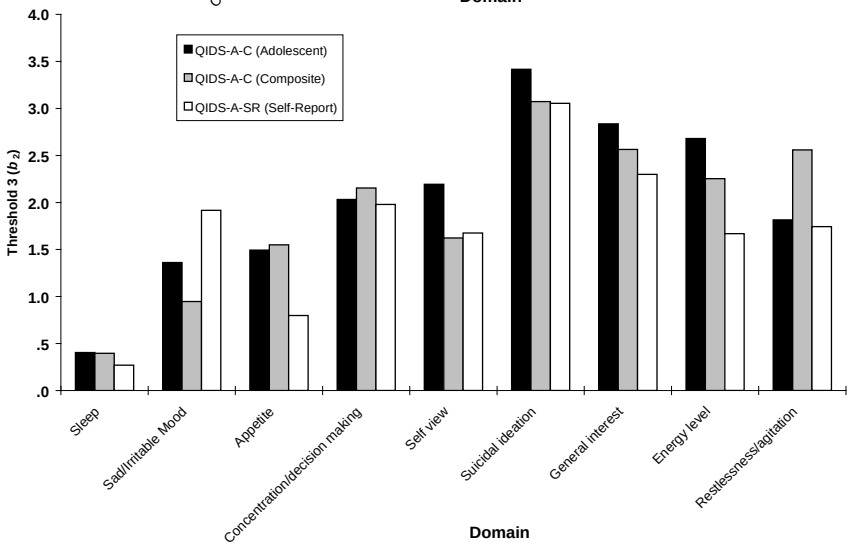
Figure 5. *Influence of Depression on Domain Response*



(6a)



(6b)



(6c)

Figure 6a-c. *IRT Domain Response: QIDS-A_{17} Threshold 1-3*

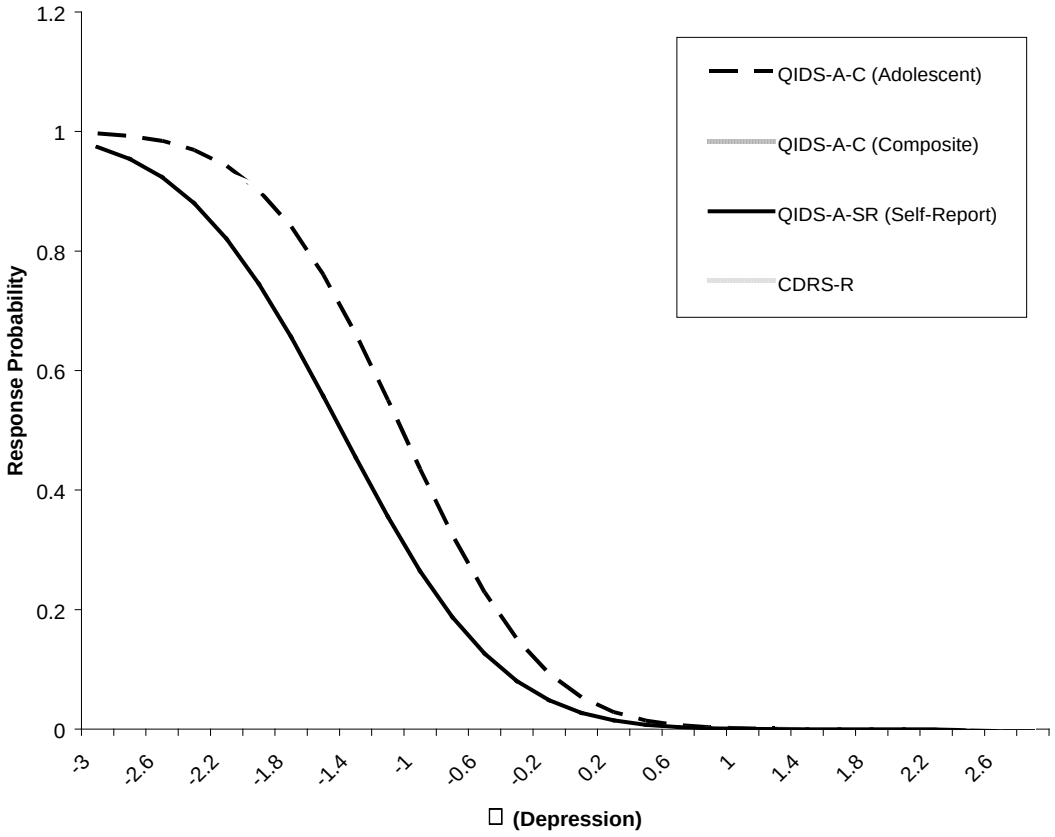


Figure 7. *CRF Curves: All Measures, Sad/Irritable Mood (Response Category 0)*

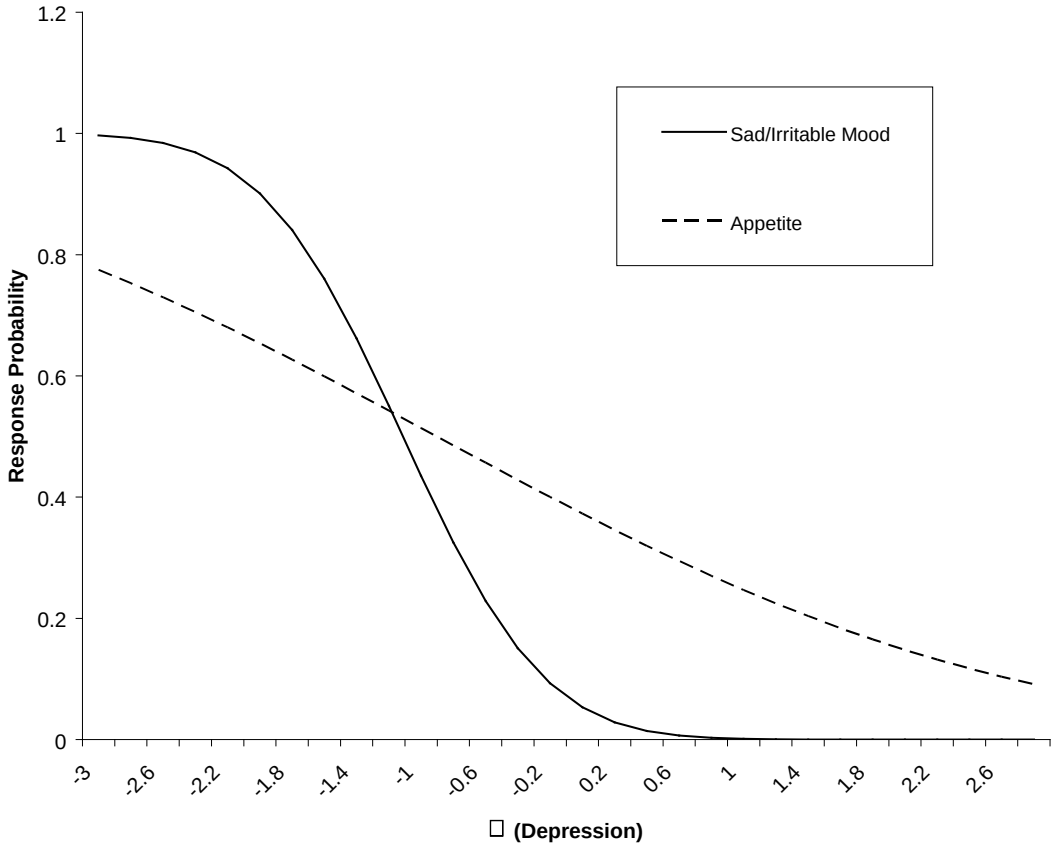


Figure 8. *CRF Curves: QIDS-A-C(Adolescent) Sad/Irritable Mood and Appetite (Response Category 0)*

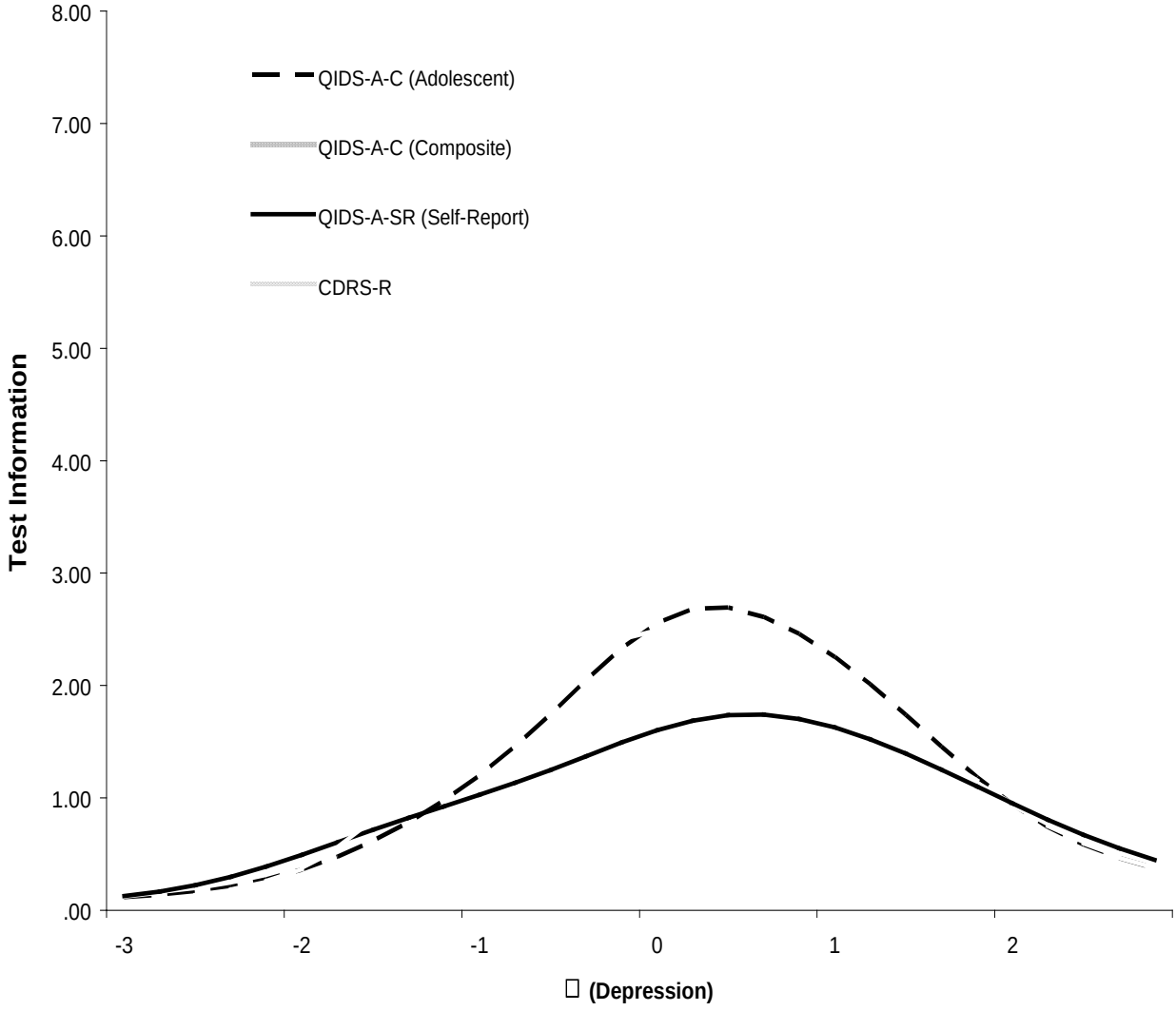


Figure 9. *Test Information Functions*

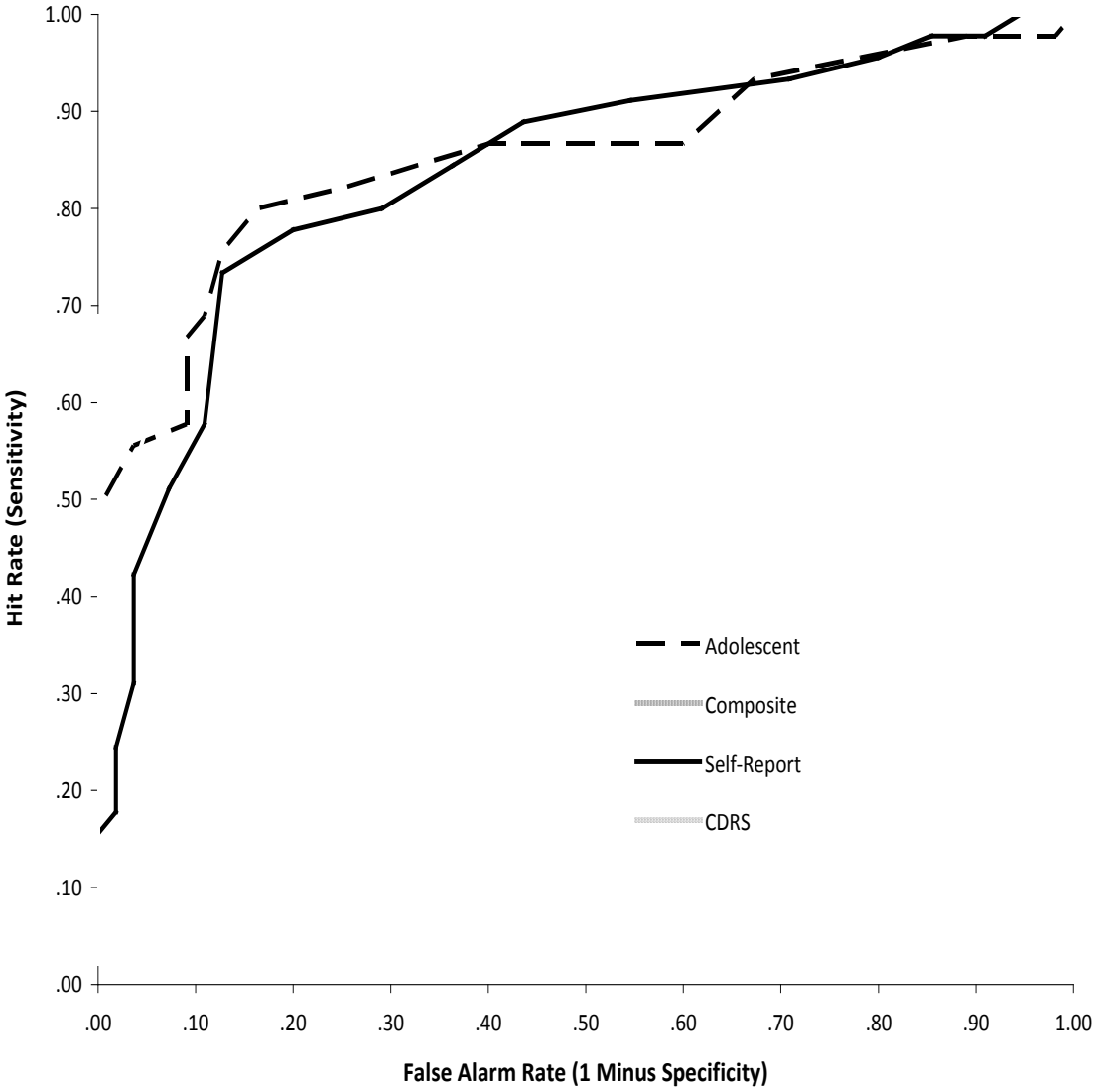


Figure 10. *ROC Curves*

APPENDIX E

Category Response Function Curves

Figure E1.

Category Reponse Functions: Adolescent QIDS-A-C (Response Category 0)

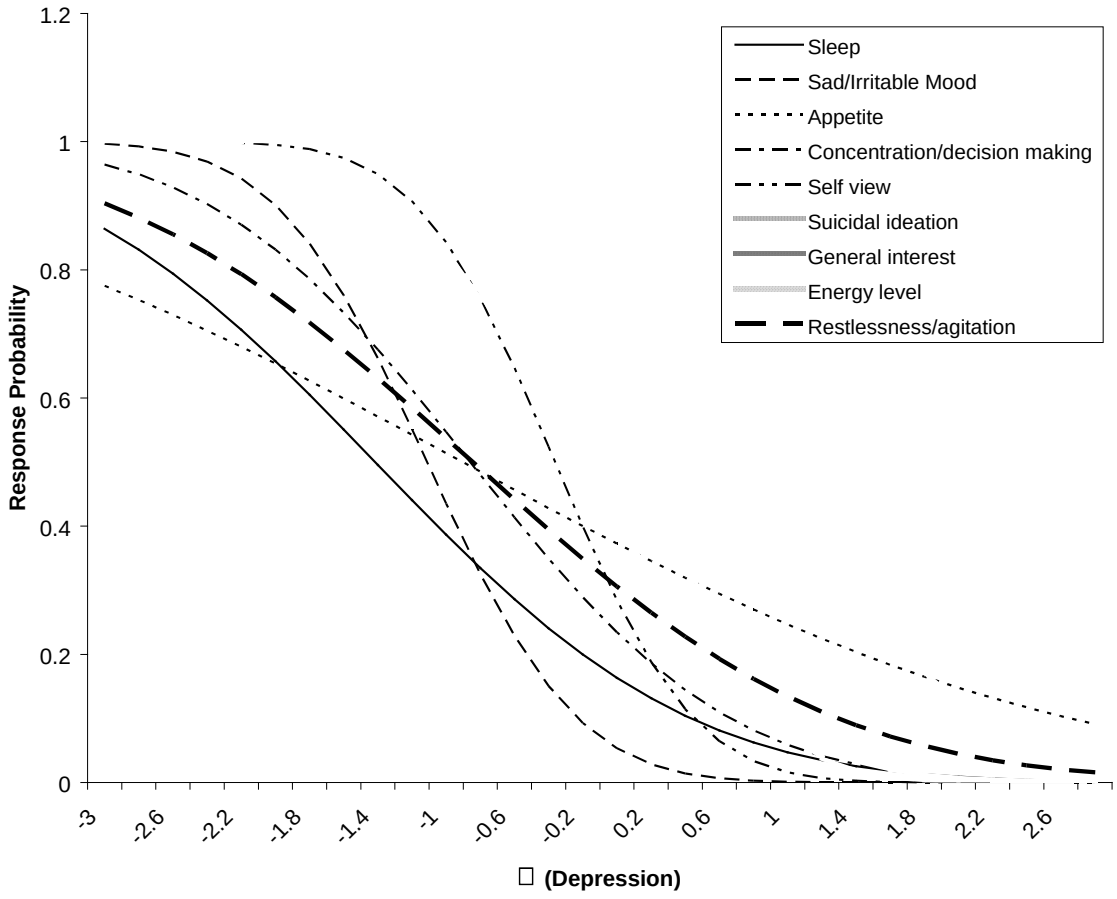


Figure E2.

Category Reponse Functions: Adolescent QIDS-A-C (Response Category 1)

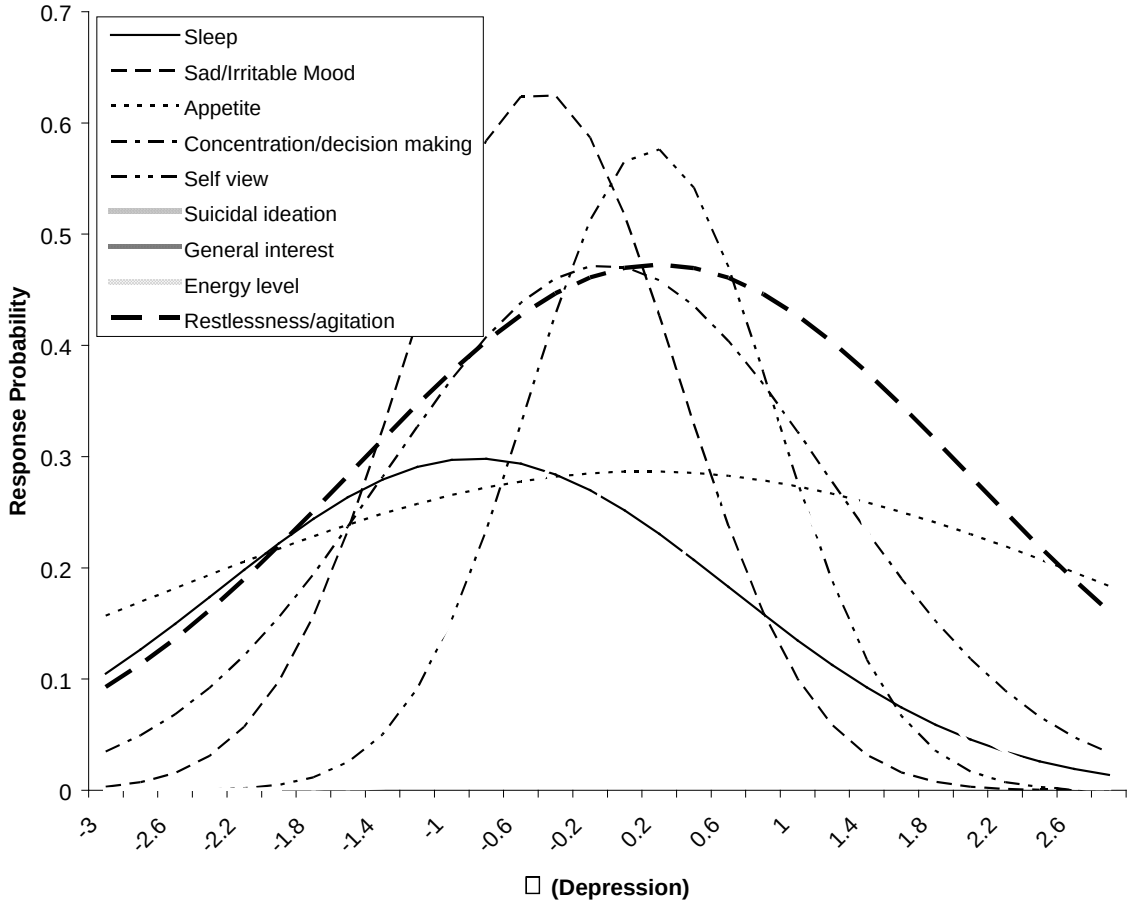


Figure E3.

Category Reponse Functions: Adolescent QIDS-A-C (Response Category 2)

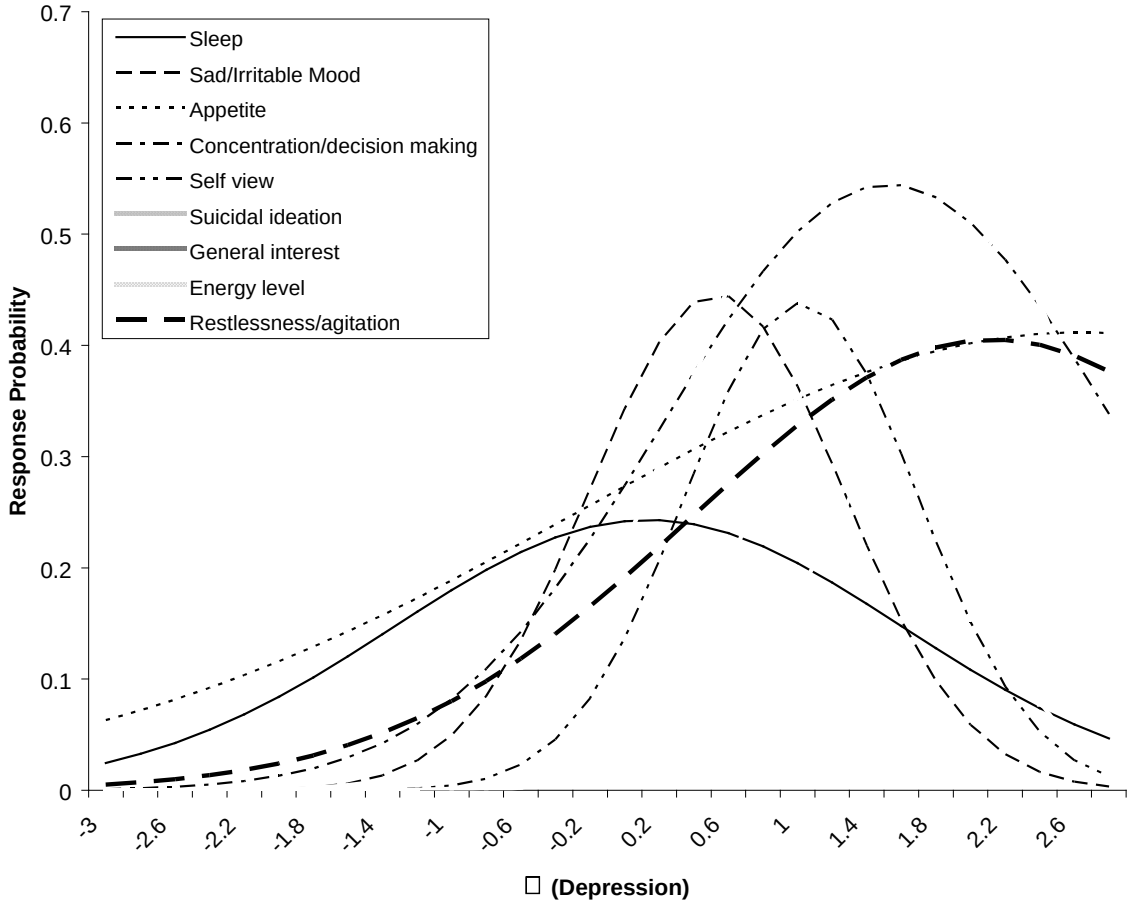


Figure E4.

Category Reponse Functions: Adolescent QIDS-A-C (Response Category 3)

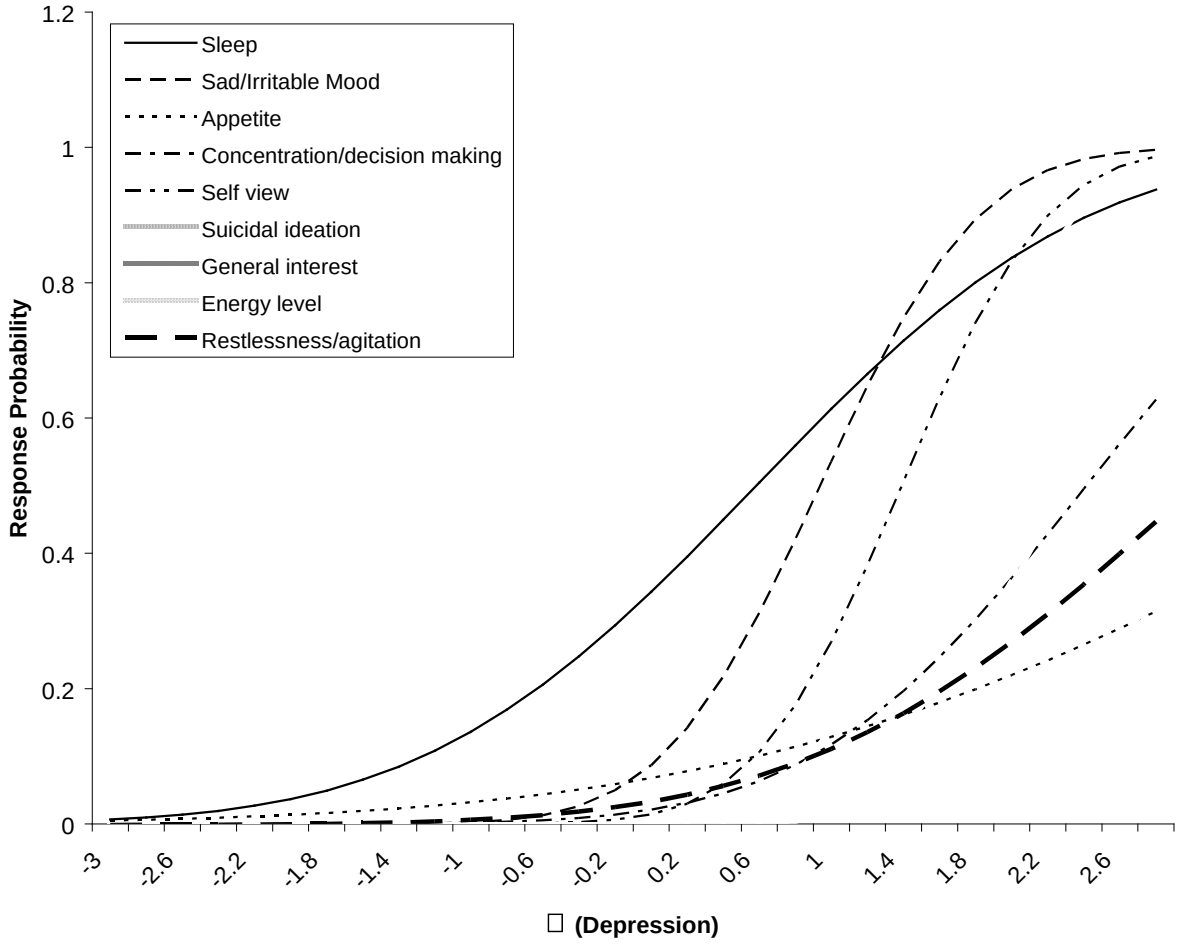


Figure E5.

Category Response Functions: Composite QIDS-A-C (Response Category 0)

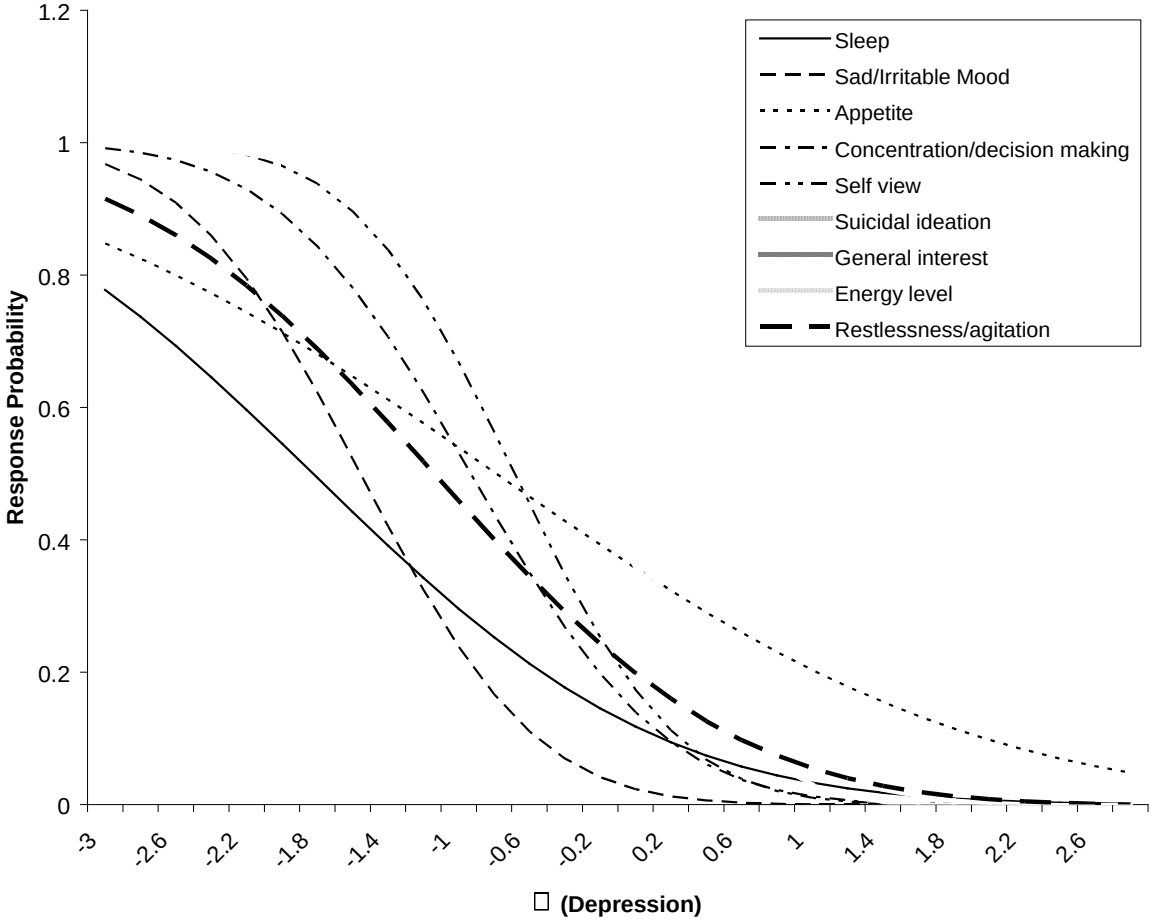


Figure E6.

Category Response Functions: Composite QIDS-A-C (Response Category 1)

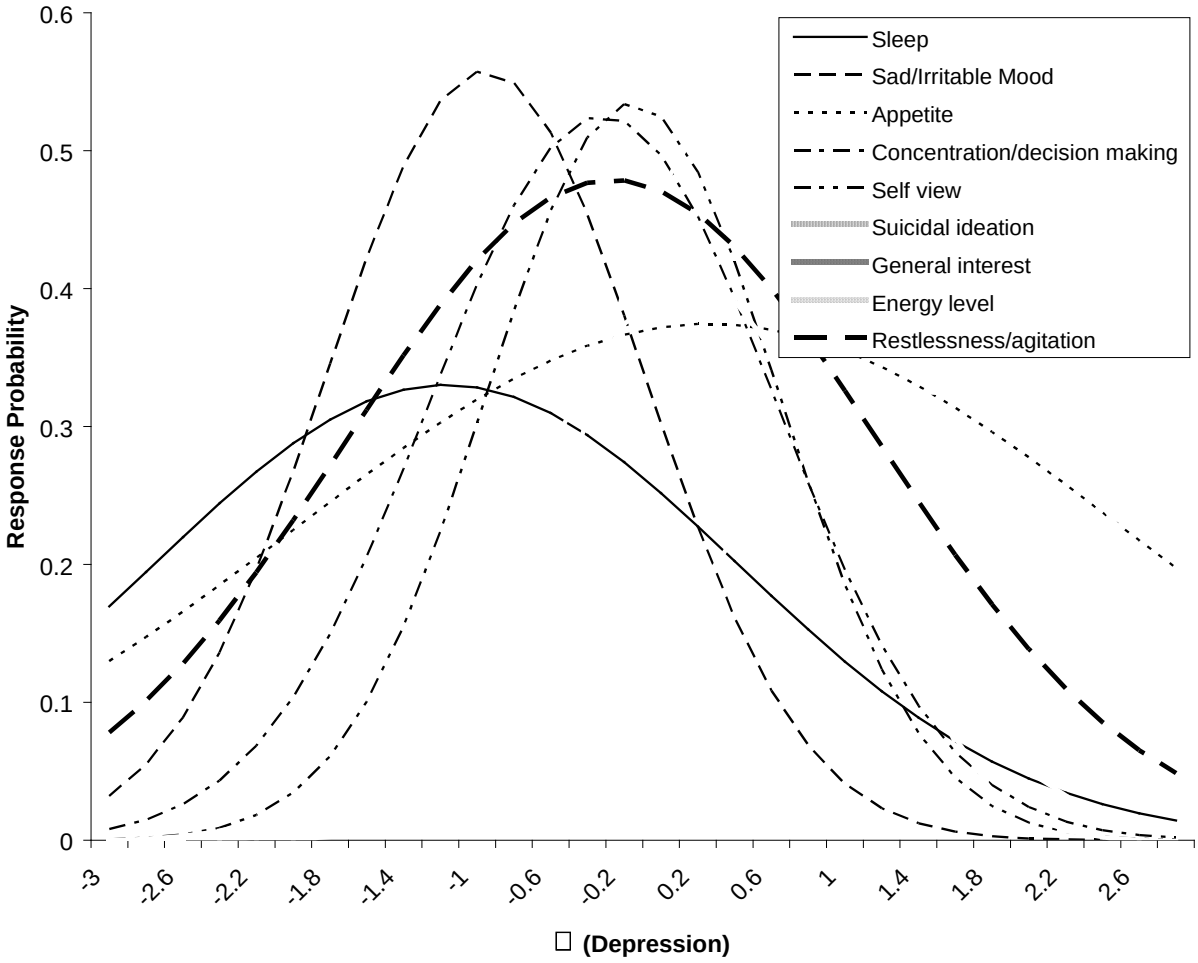


Figure E7.

Category Response Functions: Composite QIDS-A-C (Response Category 2)

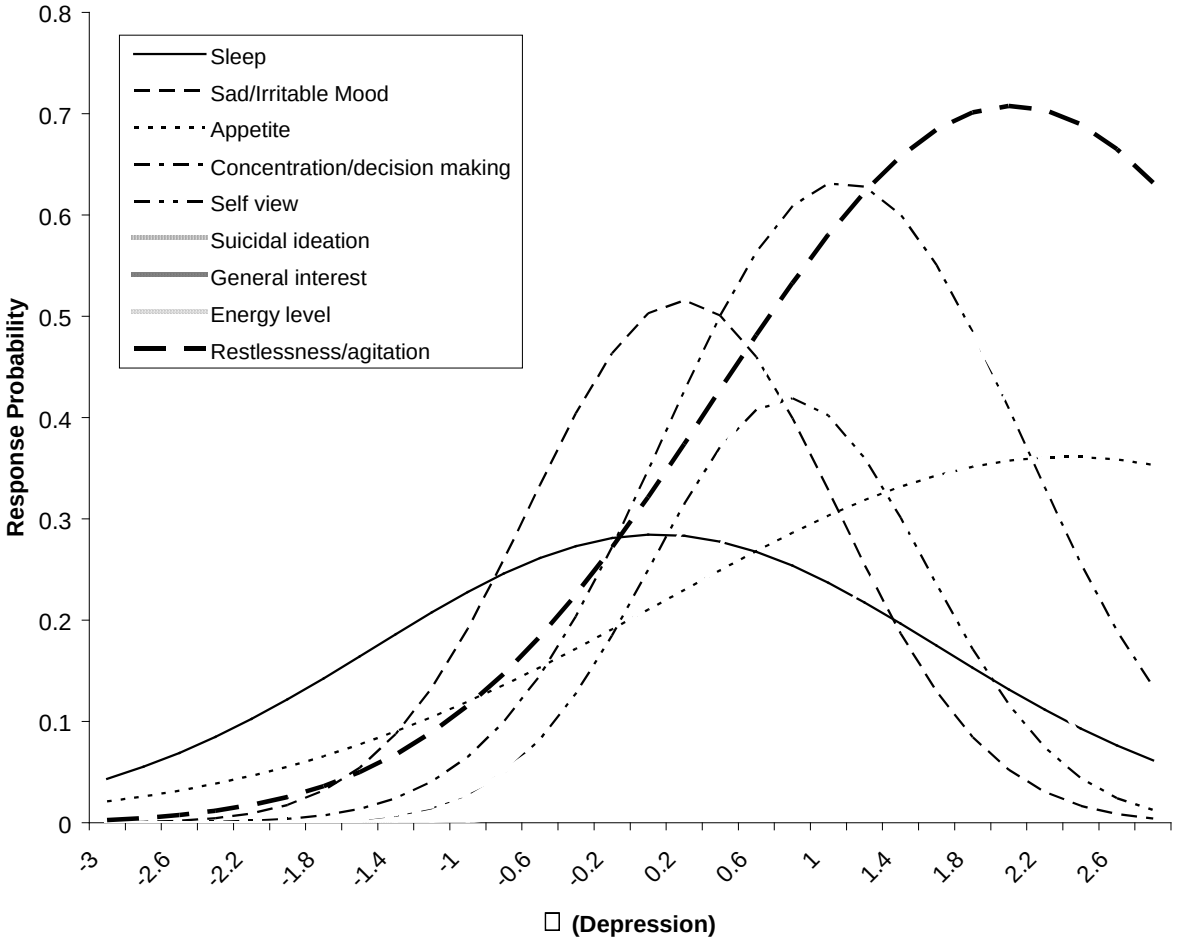


Figure E8.

Category Response Functions: Composite QIDS-A-C (Response Category 3)

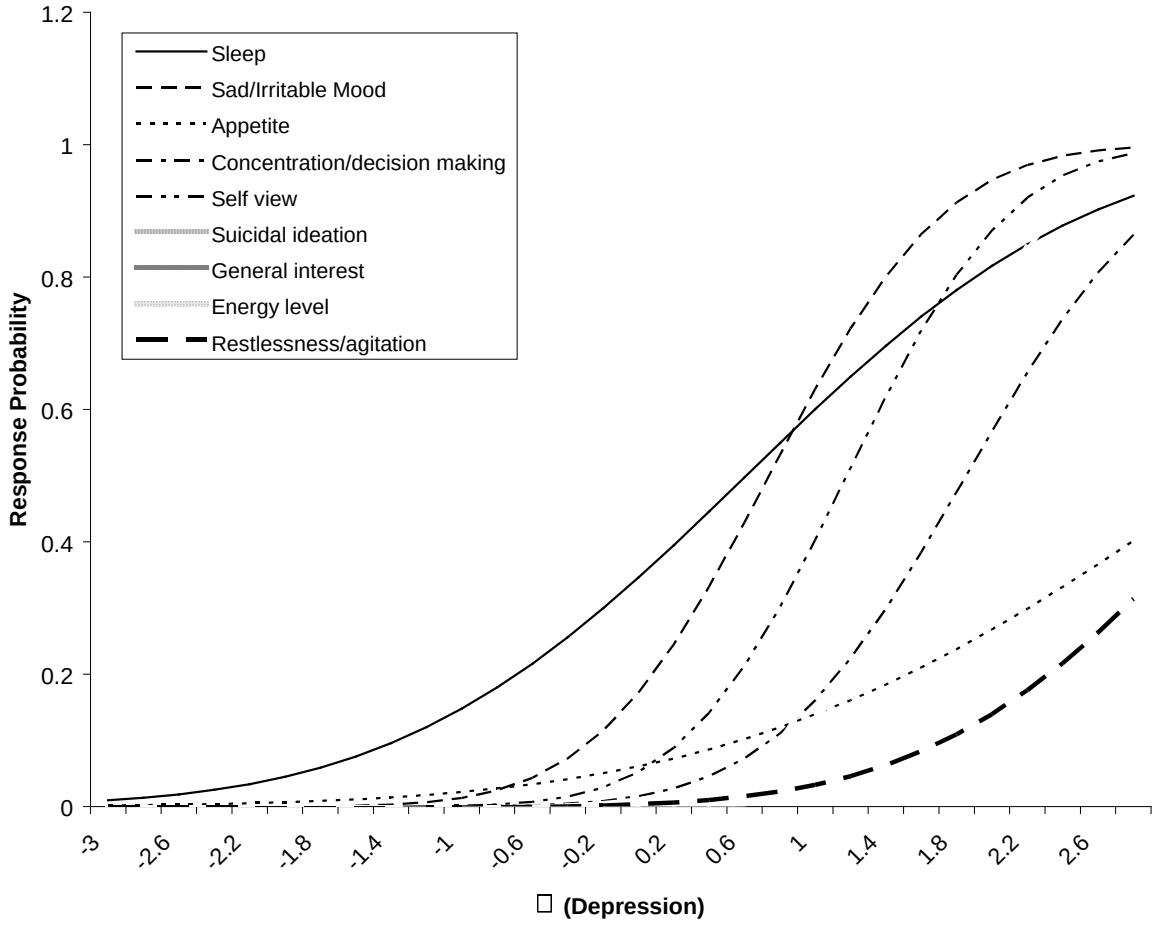


Figure E9.

Category Response Functions: Self-Report QIDS-A-SR (Response Category 0)

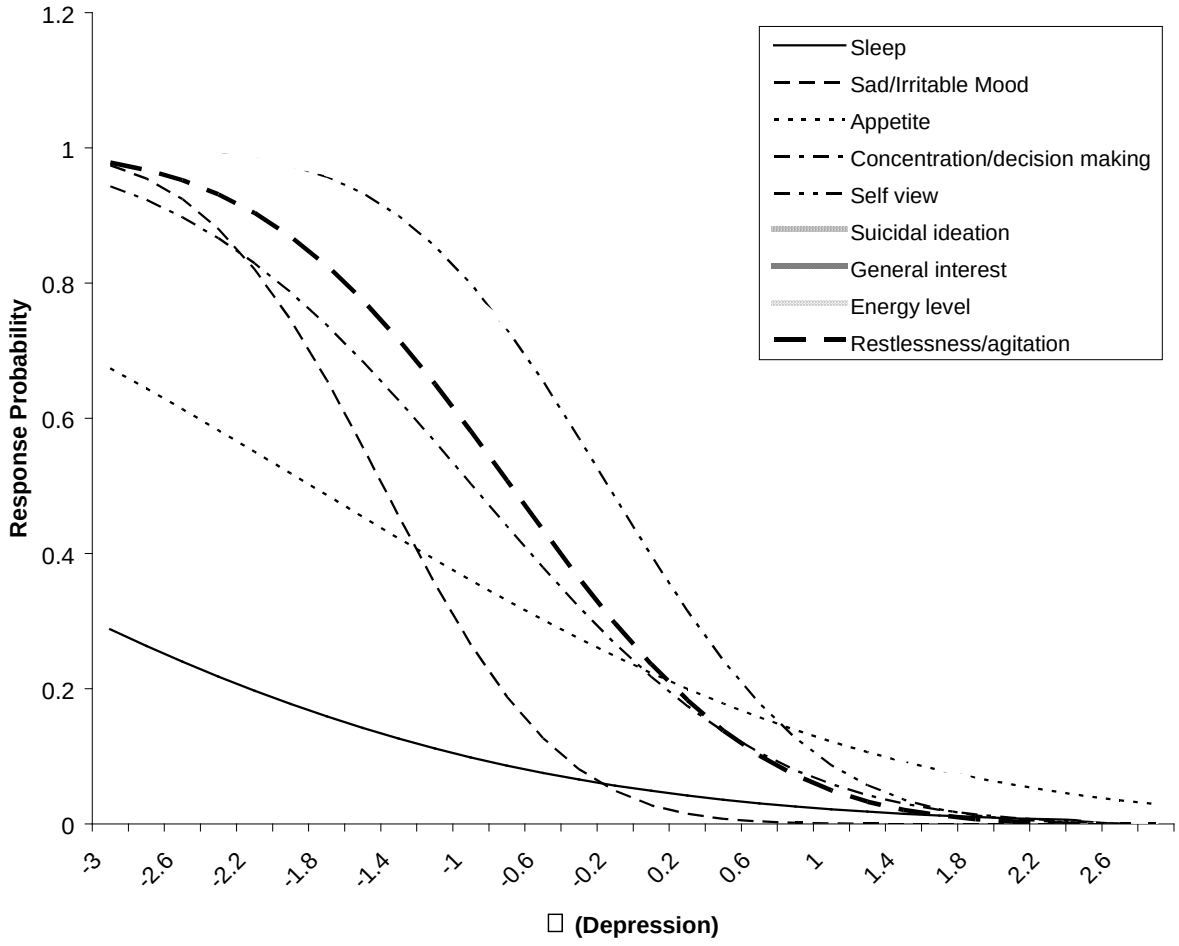


Figure E10.

Category Response Functions: Self-Report QIDS-A-SR (Response Category 1)

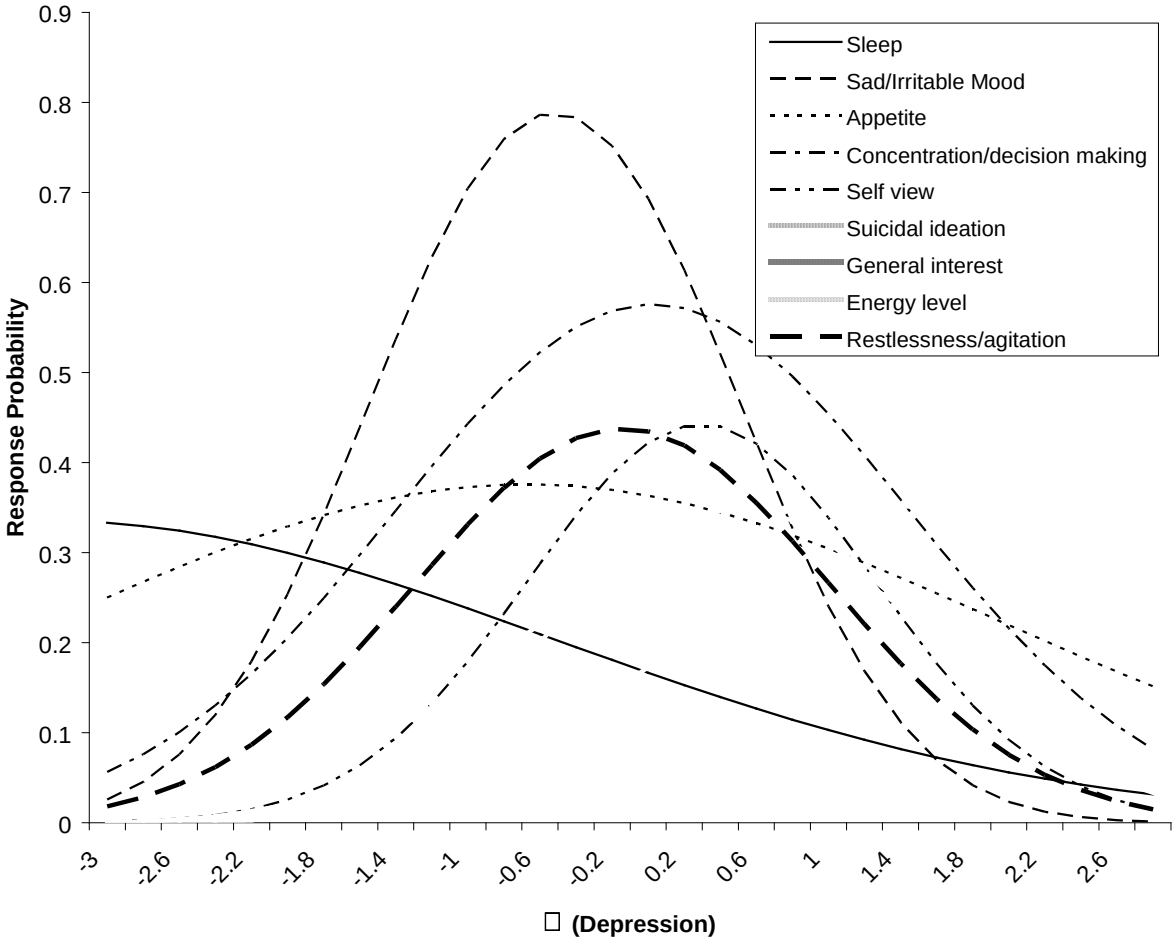


Figure E11.

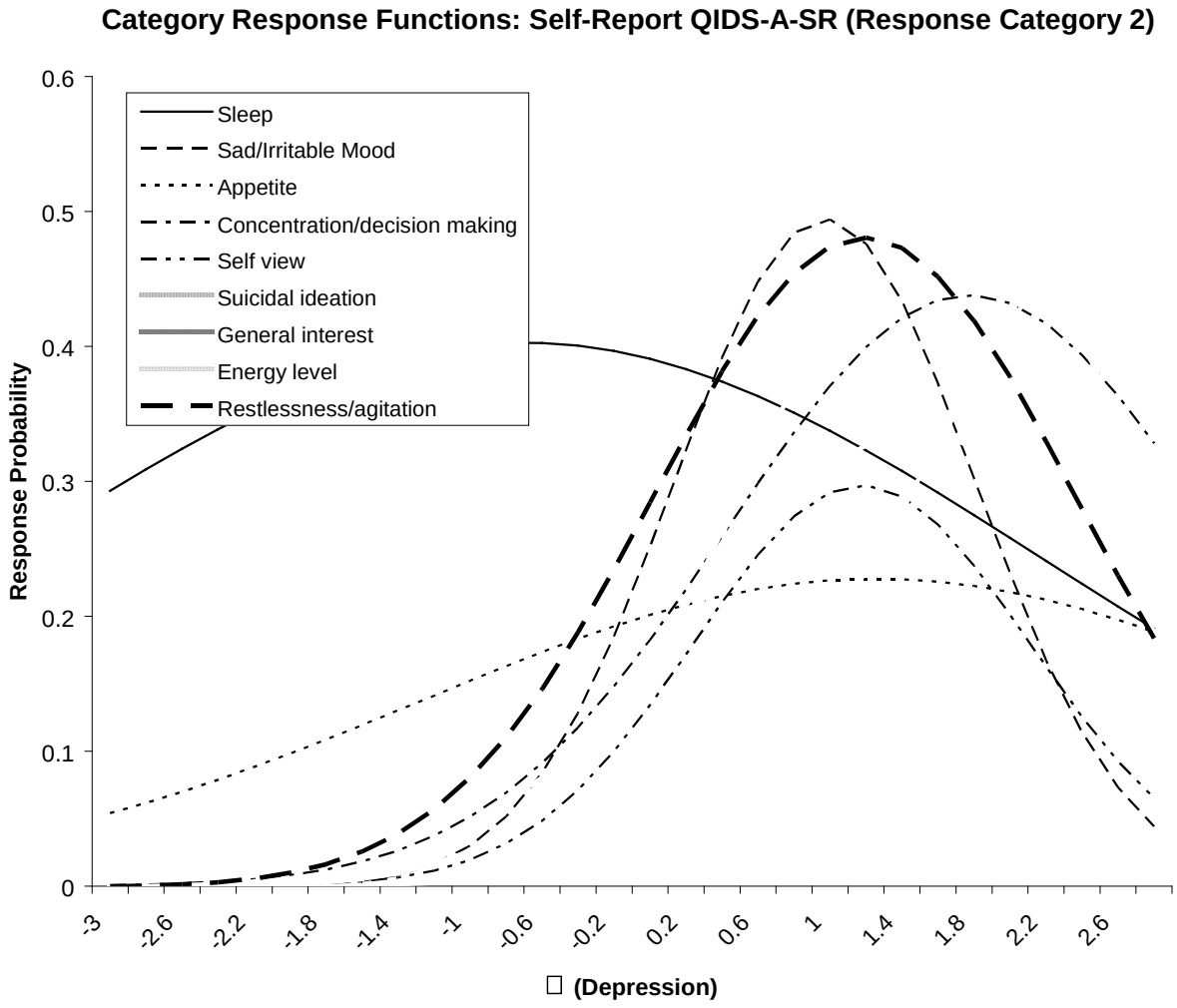

Category Response Functions: Self-Report QIDS-A-SR (Response Category 2)
Response Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
Sleep
Sad/Irritable Mood
Appetite
Concentration/decision making
Self view
Suicidal ideation
General interest
Energy level
Restlessness/agitation
-3
-2.6
-2.2
-1.8
-1.4
-1
-0.6
-0.2
0.2
0.6
1
1.4
1.8
2.2
2.6
(Depression)

Figure E12.

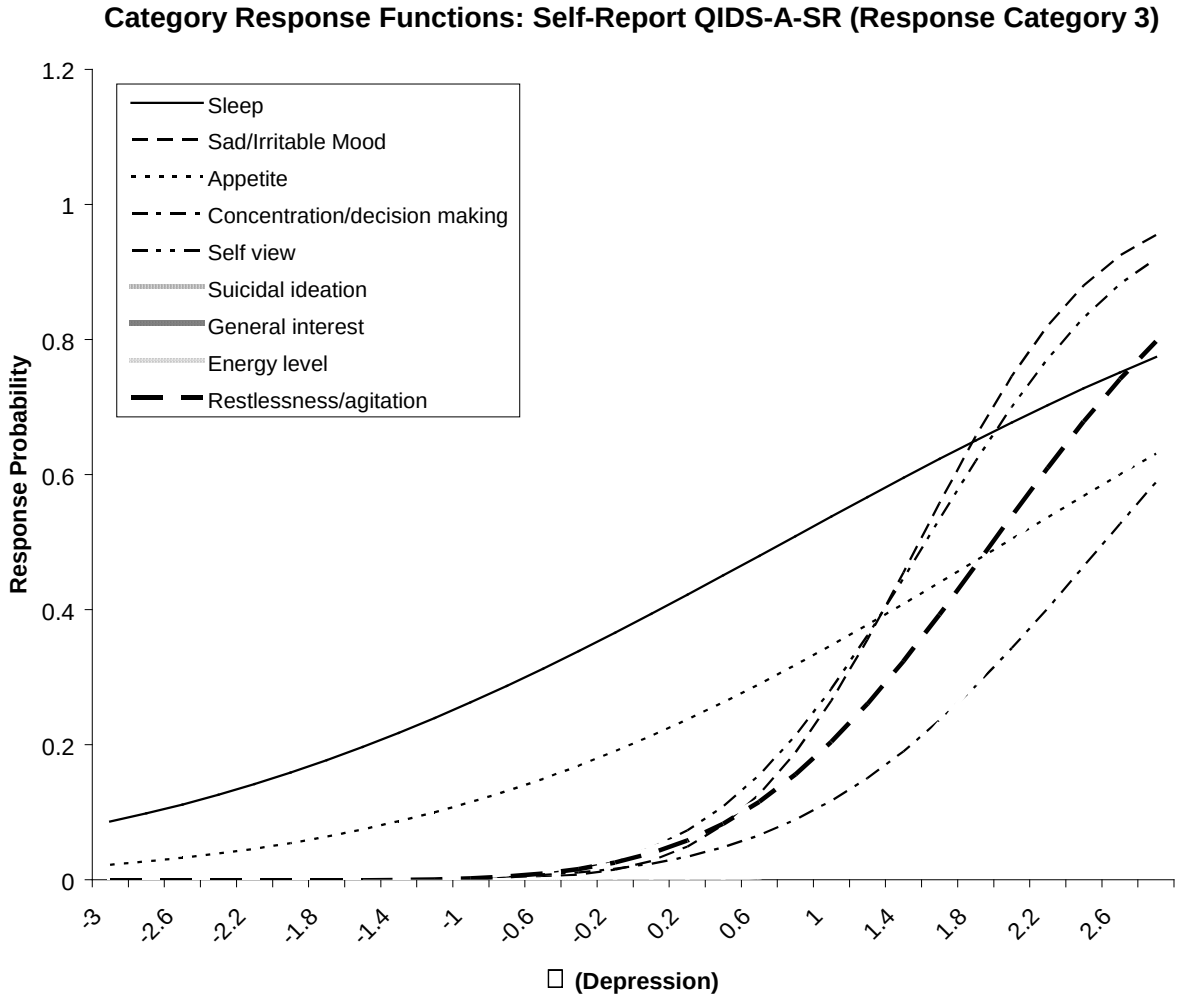

Category Response Functions: Self-Report QIDS-A-SR (Response Category 3)
Response Probability
1.2
1
0.8
0.6
0.4
0.2
0
Sleep
Sad/Irritable Mood
Appetite
Concentration/decision making
Self view
Suicidal ideation
General interest
Energy level
Restlessness/agitation
-3
-2.6
-2.2
-1.8
-1.4
-1
-0.6
-0.2
0.2
0.6
1
1.4
1.8
2.2
2.6
(Depression)

Figure E13.

Category Response Functions: CDRS (Response Category 0)

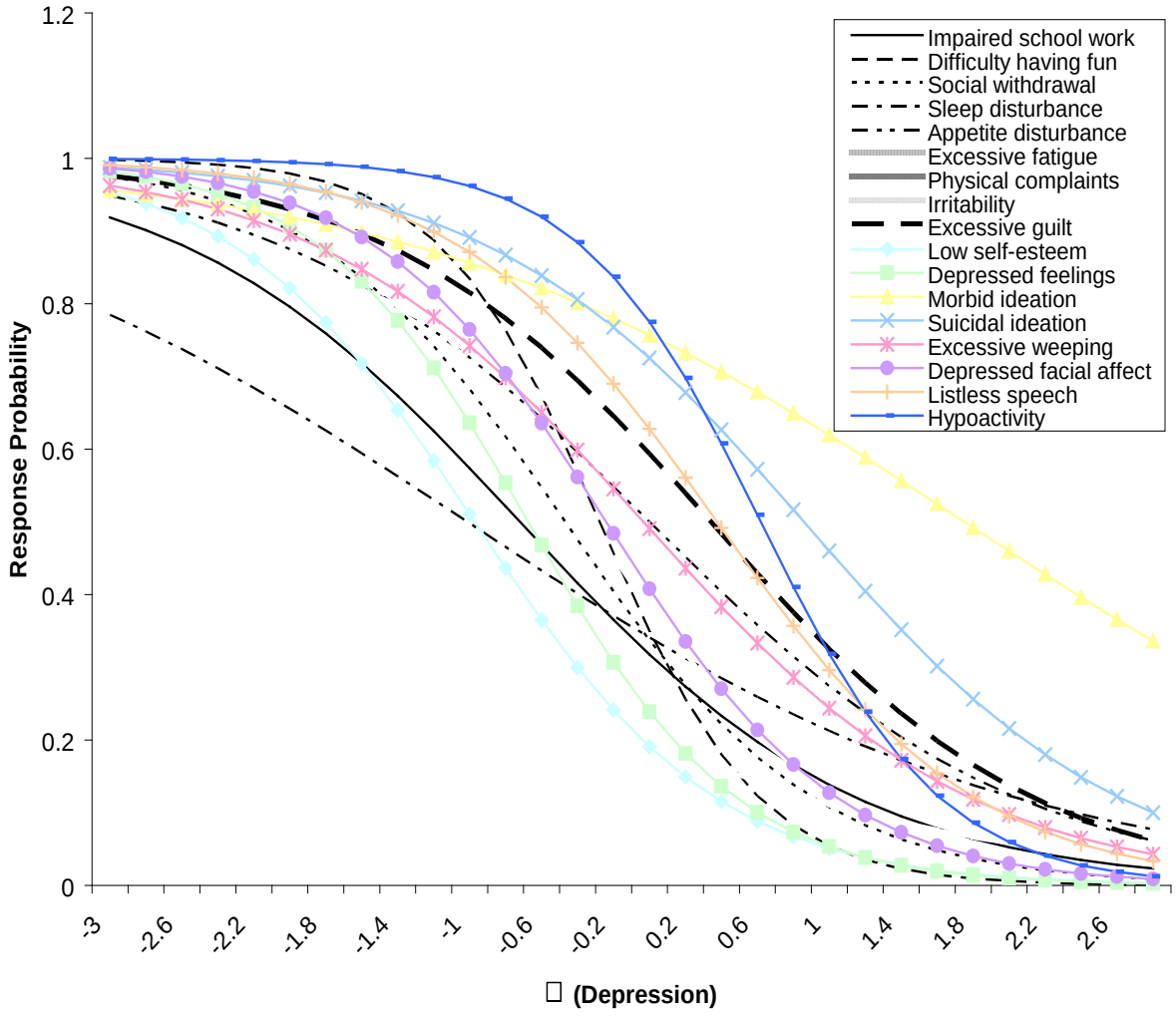


Figure E14.

Category Response Functions: CDRS (Response Category 1)

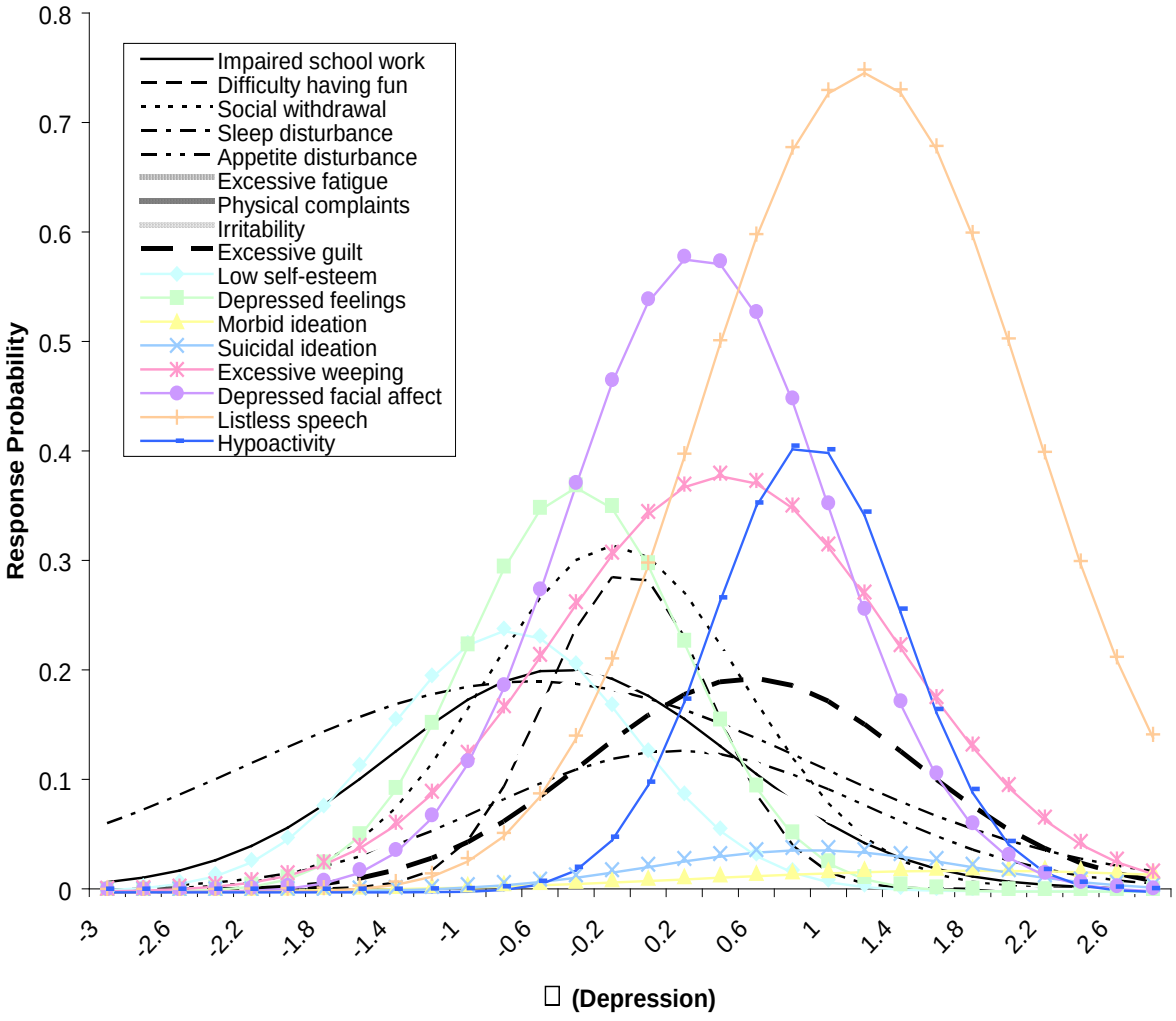


Figure E15.

Category Response Functions: CDRS (Response Category 2)

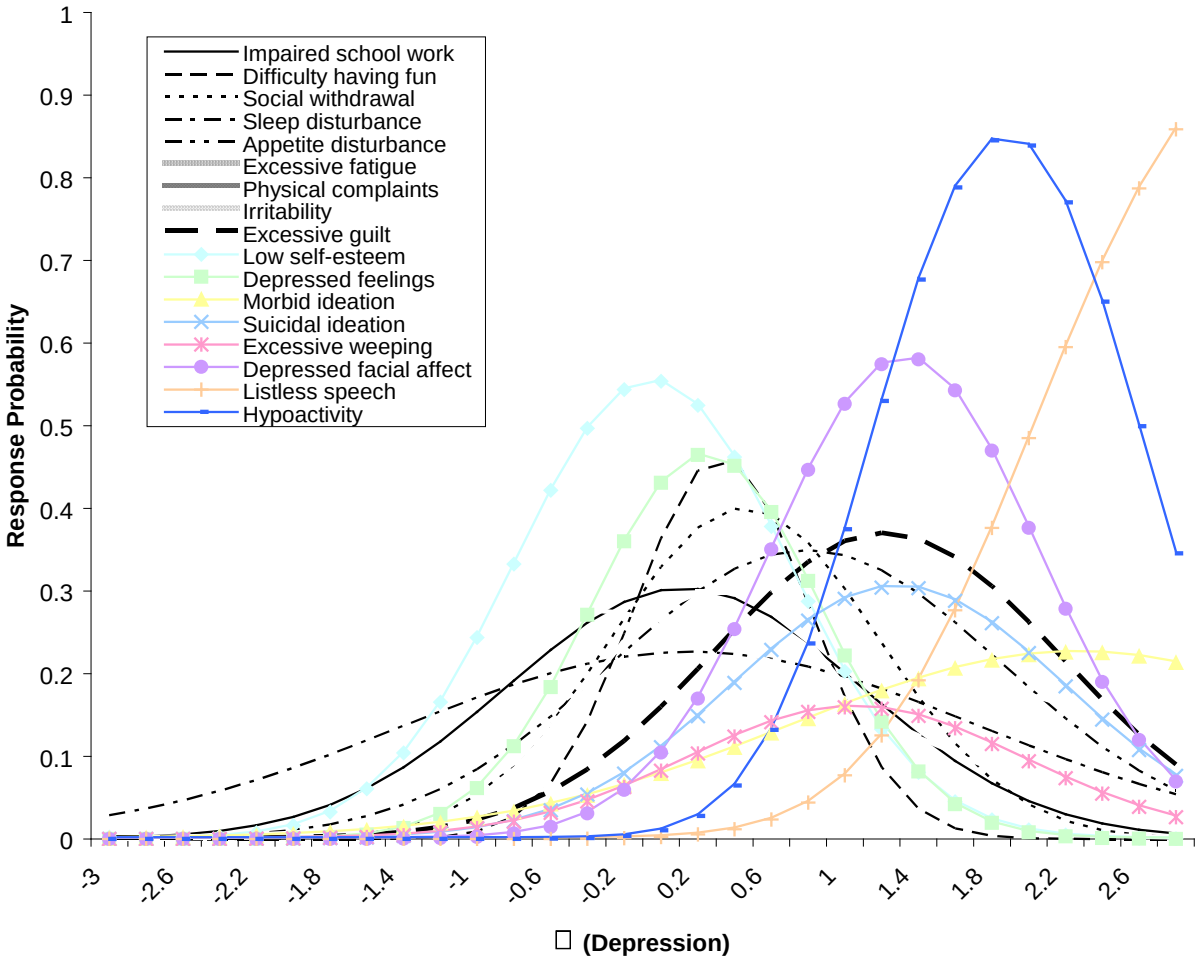
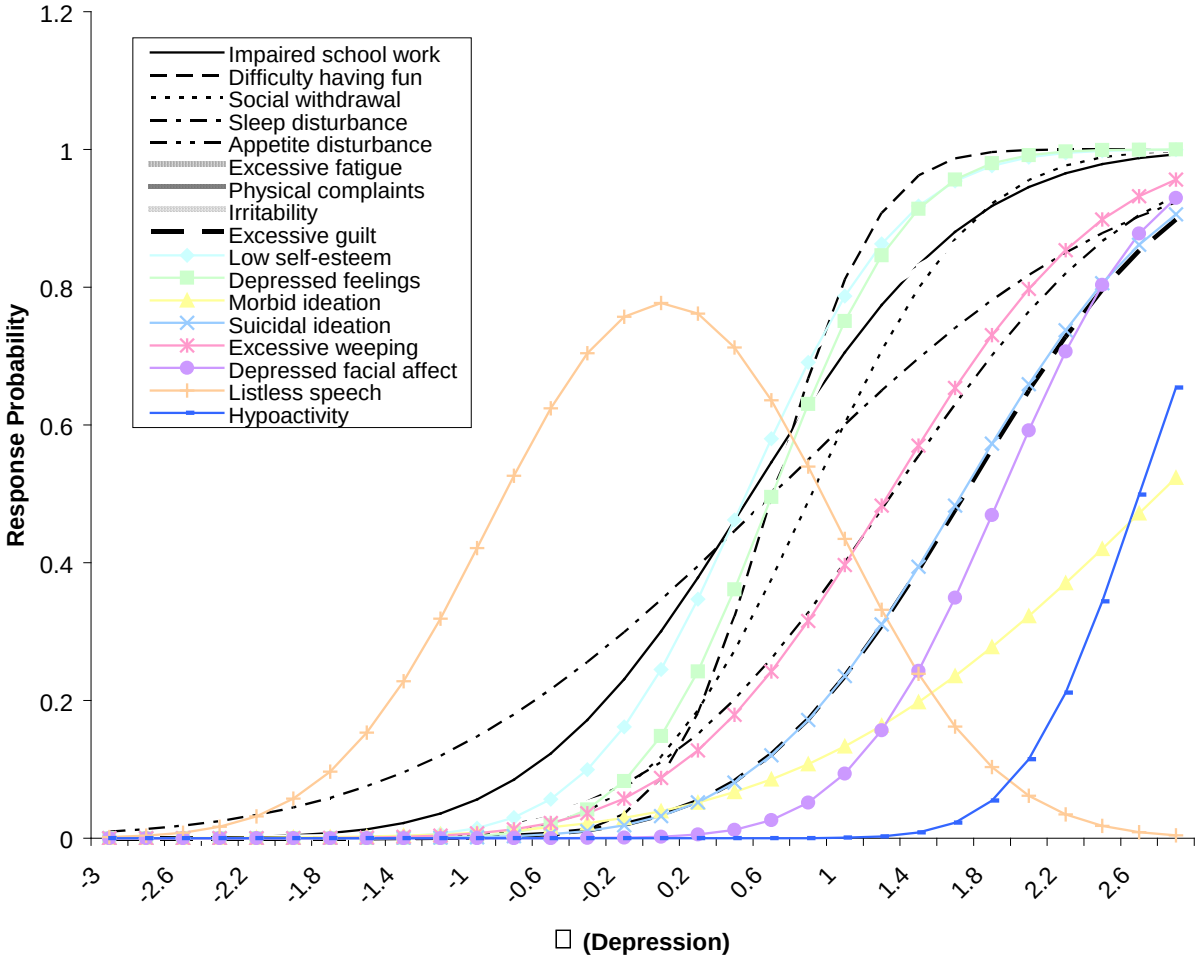


Figure E16.

Category Response Functions: CDRS (Response Category 3)



APPENDIX F

QIDS-A_{17}-C(Parent) Results

Internal consistency was equal for the QIDS-A_{17}-C(Adolescent) and the QIDS-A_{17}-C(Composite) ($\alpha = 0.84$), with a minimal difference for the QIDS-A_{17}-C(Parent) at $\alpha = 0.83$. The QIDS-A-C(Parent) correlated highly with the CDRS-R ($r = .84$). The QIDS-A-SR correlated least with the QIDS-A-C(Parent) (.58). Correlations between the QIDS-A-C(Adolescent) and the QIDS-A-C(Parent) were strong (.79). QIDS-A-C(Parent) was unidimensional for this sample.

The Samejima IRT parameter estimates for all QIDS-A measures were examined for the pattern of influence of depression on each domain response (i.e., sad mood). The self view domain strongly characterized depression on the versions that included adolescent input (i.e., QIDS-A-SR, QIDS-A-C(Adolescent), and QIDS-A-C(Composite), but not as much on the parent report (QIDS-A-C[Parent]). Similarly, while sad mood was most influenced by depression (i.e., highest slope) on the adolescent self report, and was also strongly influenced on the adolescent QIDS-A-C and the composite QIDS-A-C, it was not among the three highest slopes (i.e., most influenced by depression) on the parent report. Specifically, general interest, energy level, and concentration/decision making were more characterized by depression on the parent report than sad mood. This difference in parent reported symptoms (i.e., higher for observable symptoms) likely reflects limitations of parent report in measuring the internalized symptoms of depression (sad mood, self view).

For diagnostic validity, the QIDS-A-C(Parent) was similarly discriminating between depressed and nondepressed subjects as the other QIDS-A-C measures, using

univariate and multivariate analyses. The area under the ROC curve for QIDS-A-C(Parent) was .902, which was similar to the other measures. In sum, although the parent input increases reliability slightly, the adolescent's input is essential and enough by itself for purposes of screening.

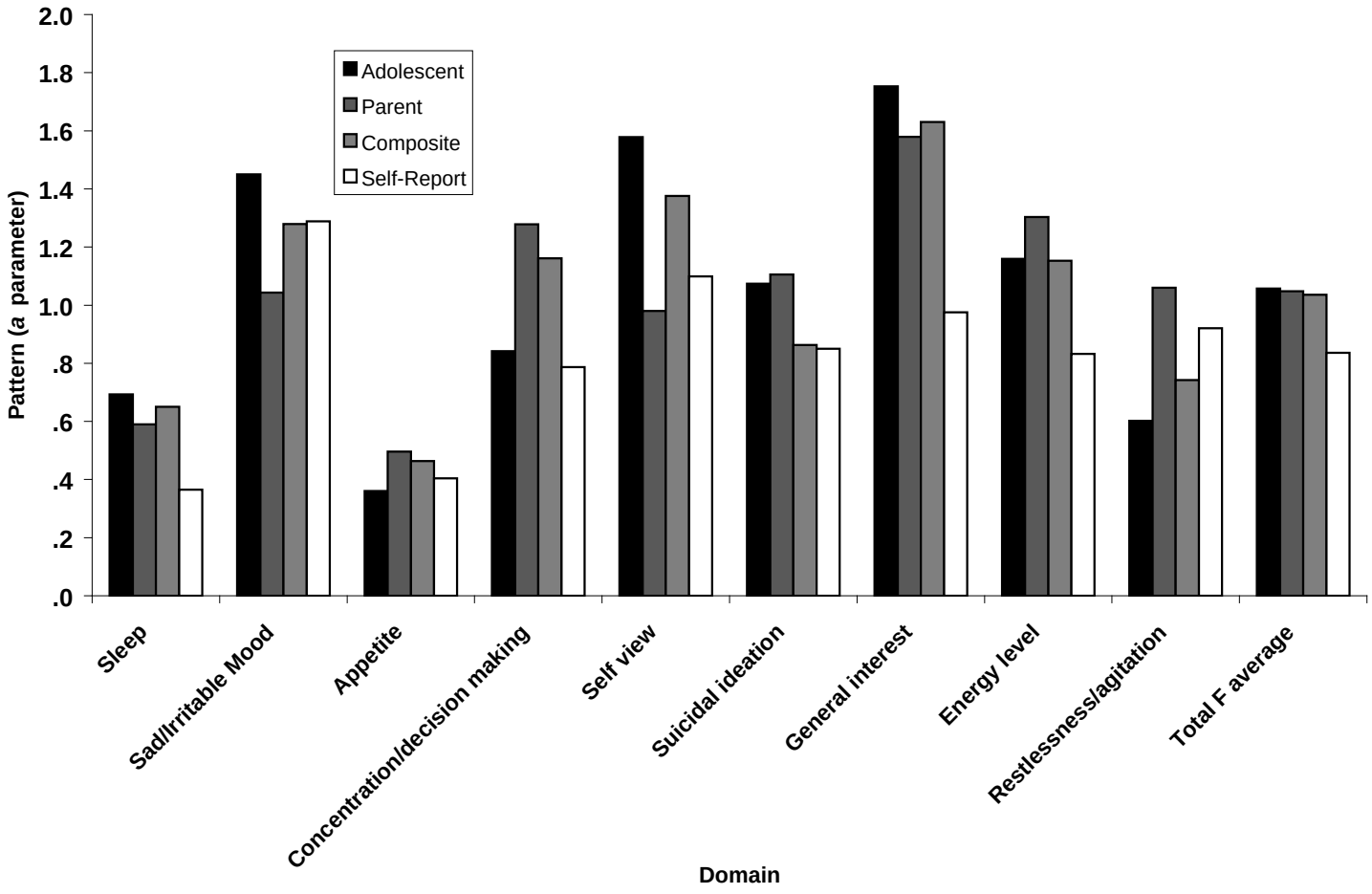
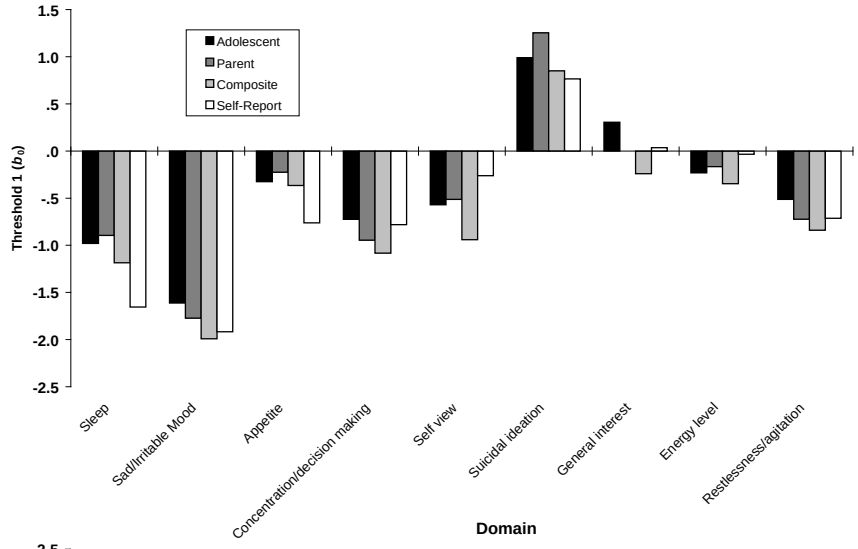
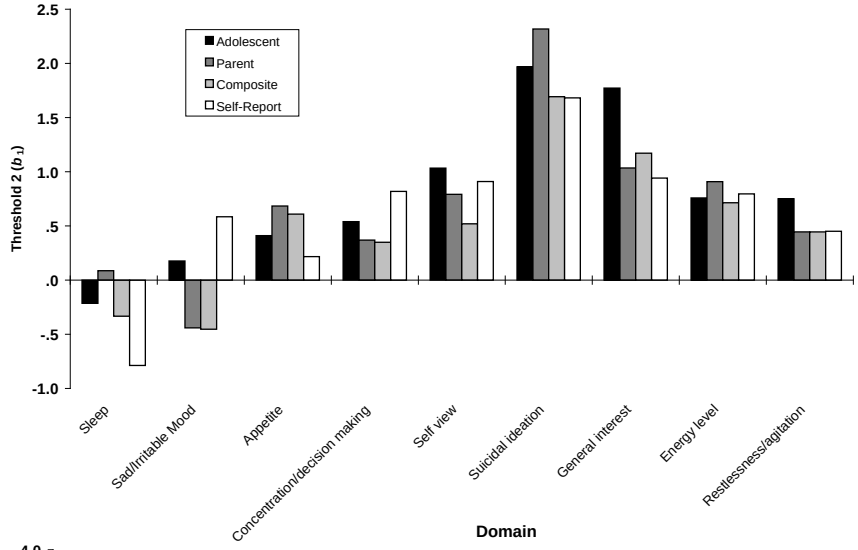


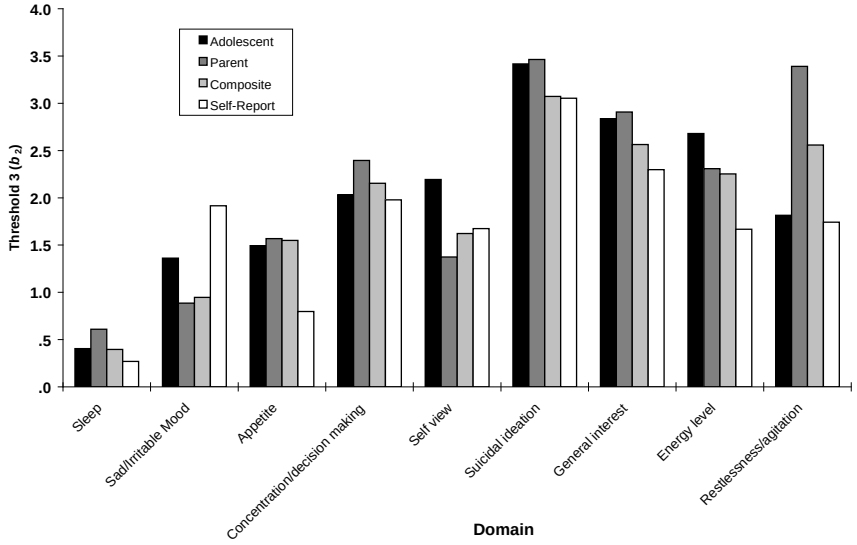
Figure F1. Influence of Depression on Domain Response



(2a)



(2b)



(2c)

Figure F2a-c. IRT Domain Response: QIDS-A_{17} Threshold 1-3

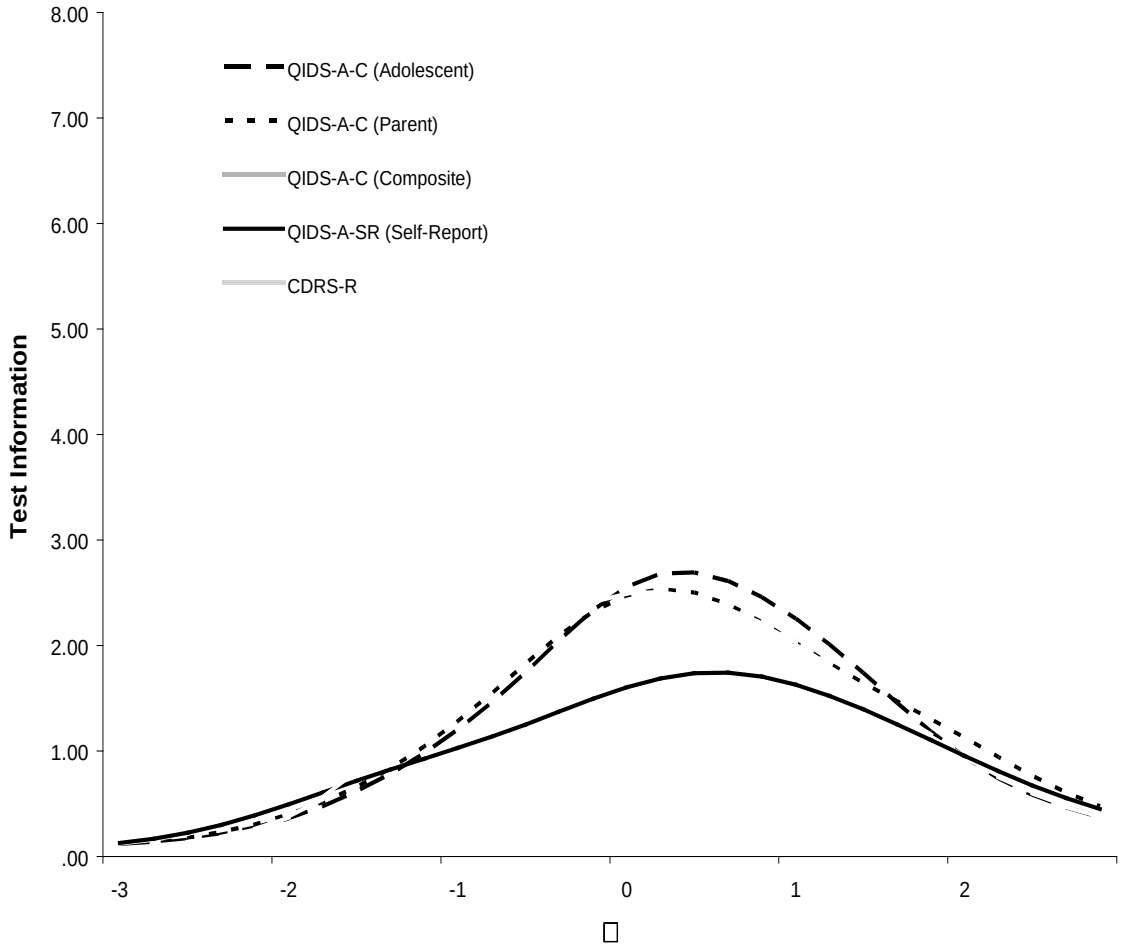


Figure F3. Test Information Functions

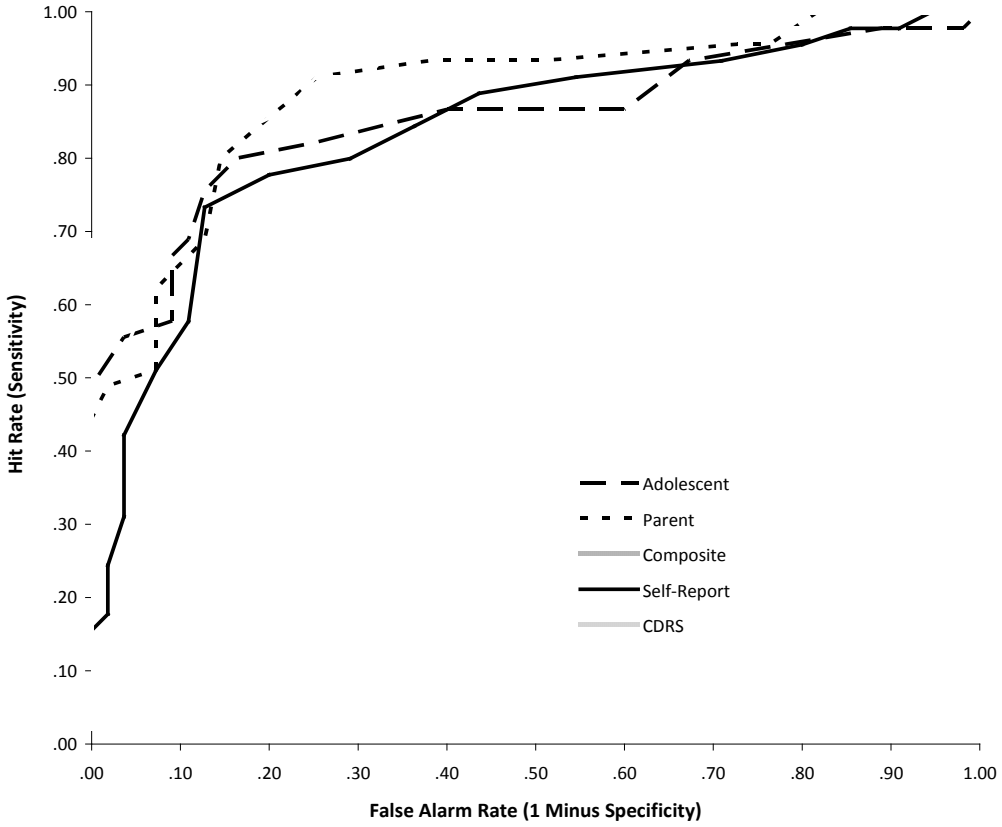


Figure F4. ROC Curves

REFERENCES

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

American Psychiatric Association. (2000b). Practice guideline for the treatment of patients with major depressive disorder (revision). *American Journal of Psychiatry, 157,* 1-45.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213-232.

Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (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., & Steer, R. A. (1993). *Beck Depression Inventory (BDI) Manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Bernstein, I. H., Rush, A. J., Carmody, T. J., Woo, A., & Trivedi, M. H. (2007). Clinical vs. self-report versions of the quick inventory of depressive symptomatology in a public sector sample. *Journal of Psychiatric Research, 41,* 239-246.

Bernstein, I. H., Rush, A. J., Trivedi, M. H., Hughes, C. W., Macleod, L., Witte, B. P., Jain, S., Emslie, G. E. (2008). *Psychometric Properties of the Quick Inventory of Depressive Symptomatology in Adolescents*. Manuscript in preparation.

Bernstein, I. B., Wendt, B., Nasr, S. J., Rush, A. J. (in press). Screening for Major Depression in Private Practice. *Journal of Psychiatric Practice*.

Birmaher, B., Brent, D., Bernet, W., Bukstein, O., Walter, H., Benson, R. S. et al. (2007). Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 1503-1526.

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E. et al. (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.

Braaten, E. B., Biederman, J., DiMauro, A., Mick, E., Monuteaux, M. C., Muehl, K. et al. (2001). Methodological complexities in the diagnosis of major depression in youth: an analysis of mother and youth self-reports. *Journal of Child and Adolescent Psychopharmacology, 11,* 395-407.

Brent, D. A. (1987). Correlates of the medical lethality of suicide attempts in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 87-91.

Brent, D. A., Perper, J. A., Goldstein, C. E., Kolko, D. J., Allan, M. J., Allman, C. J. et al. (1988). Risk factors for adolescent suicide. A comparison of adolescent suicide victims with suicidal inpatients. *Archives of General Psychiatry, 45,* 581-588.

Brent, D. A., Perper, J. A., Moritz, G., Allman, C., Friend, A., Roth, C. et al. (1993). Psychiatric risk factors for adolescent suicide: a case-control study. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 521-529.

Bridge, J. A., Iyengar, S., Salary, C. B., Barbe, R. P., Birmaher, B., Pincus, H. A. et al. (2007). Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *Journal of the American Medical Association, 297,* 1683-1696.

Brooks, S. J. & Kutcher, S. (2001). Diagnosis and measurement of adolescent depression: a review of commonly utilized instruments. *Journal of Child and Adolescent Psychopharmacology, 11,* 341-376.

Brown, E. S., Murray, M., Carmody, T. J., Kennard, B. D., Hughes, C. W., Khan, D. A. et al. (2008). The Quick Inventory of Depressive Symptomatology-Self-report: a psychometric evaluation in patients with asthma and major depressive disorder. *Annals of Allergy, Asthma, and Immunology, 100,* 433-438.

Burke, K. C., Burke, J. D., Jr., Rae, D. S., & Regier, D. A. (1991). Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five US community populations. *Archives of General Psychiatry, 48,* 789-795.

Burke, K. C., Burke, J. D., Jr., Regier, D. A., & Rae, D. S. (1990). Age at onset of selected mental disorders in five community populations. *Archives of General Psychiatry, 47,* 511-518.

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.

Carmody, T. J., Rush, A. J., Bernstein, I. H., Brannan, S., Husain, M. M., & Trivedi, M. H. (2006). Making clinicians lives easier: guidance on use of the QIDS self-report in place of the MADRS. *Journal of Affective Disorders, 95,* 115-118.

Cheung, A. H., Emslie, G. J., & Mayes, T. L. (2005). Review of the efficacy and safety of antidepressants in youth depression. *Journal of Child Psychology and Psychiatry, 46,* 735-754.

Cicchetti, D. & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology, 70,* 6-20.

Clark, L. A., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Coyle, J. T., Pine, D. S., Charney, D. S., Lewis, L., Nemeroff, C. B., Carlson, G. A. et al. (2003). Depression and bipolar support alliance consensus statement on the unmet needs in diagnosis and treatment of mood disorders in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1494-1503.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Dorz, S., Borgherini, G., Conforti, D., Scarso, C., & Magni, G. (2004). Comparison of self-rated and clinician-rated measures of depressive symptoms: a naturalistic study. *Psychological Psychotherapy, 77,* 353-361.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Emslie, G. J., Heiligenstein, J. H., Wagner, K. D., Hoog, S. L., Ernest, D. E., Brown, E. et al. (2002). Fluoxetine for acute treatment of depression in children and adolescents: a placebo-controlled, randomized clinical trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 1205-1215.

Emslie, G. J., Rush, A. J., Weinberg, W. A., Kowatch, R. A., Hughes, C. W., Carmody, T. et al. (1997). A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Archives of General Psychiatry, 54,* 1031-1037.

Fergusson, D. M. & Woodward, L. J. (2002). Mental health, educational, and social role outcomes of adolescents with depression. *Archives of General Psychiatry, 59,* 225-231.

Flanery, R.C. (1990). Methodological and psychometric considerations in child reports. In A.M. La Greca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 57-82). Boston, MA: Allyn & Bacon.

Fleming, J. E. & Offord, D. R. (1990). Epidemiology of childhood depressive disorders: a critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 571-580.

Garber, J., Keiley, M. K., & Martin, C. (2002). Developmental trajectories of adolescents' depressive symptoms: predictors of change. *Journal of Consulting and Clinical Psychology, 70,* 79-95.

Garrison, C. Z., Jackson, K. L., Marsteller, F., McKeown, R., & Addy, C. (1990). A longitudinal study of depressive symptomatology in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 581-585.

Gibbons, R. D., Clark, D. C., & Kupfer, D. J. (1993). Exactly what does the Hamilton Depression Rating Scale measure? *Journal of Psychiatric Research, 27,* 259-273.

Gould, M. S., King, R., Greenwald, S., Fisher, P., Schwab-Stone, M., Kramer, R. et al. (1998). Psychopathology associated with suicidal ideation and attempts among children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 37,* 915-923.

Gullion, C. M. & Rush, A. J. (1998). Toward a generalizable model of symptoms in major depressive disorder. *Biological Psychiatry, 44,* 959-972.

Guo, Y., Nilsson, M. E., Heiligenstein, J., Wilson, M. G., & Emslie, G. (2006). An exploratory factor analysis of the children's depression rating scale-revised. *Journal of Child and Adolescent Psychopharmacology, 16,* 482-491.

Guy, W. (1976). Clinical Global Impressions scale. In *ECDEU assessment manual for psychopharmacology, revised* (pp. 218-222). Rockville, MD: National Institute of Mental Health.

Hamilton, M. (1960). A rating scale for depression. *Journal of neurology, neurosurgery and psychiatry, 23,* 56-62.

Harrington, R., Fudge, H., Rutter, M., Pickles, A., & Hill, J. (1990). Adult outcomes of childhood and adolescent depression. I: Psychiatric status. *Archives of General Psychiatry, 47,* 465-473.

Hatcher-Kay, C. & King, C. A. (2003). Depression and suicide. *Pediatrics in Review, 24,* 363-371.

Herjanic, B. & Reich, W. (1982). Development of a structured psychiatric interview for children: agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology., 10,* 307-324.

Heron, M. P. (2007). Deaths: Leading causes for 2004 (National Vital Statistics Report, Vol. 56, No. 5). Hyattsville, MD: National Center for Health Statistics.

Hibbs, E. D., & Jensen, P. S. (Eds.). (2004). Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice (2nd ed.). Washington DC: American Psychological Association

Horn, J. L. (1965). An empirical comparison of various methods for estimating common factor scores. *Educational and Psychological Measurement, 25*, 313-322.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55

Hughes, C. W., Emslie, G. J., Crismon, M. L., Wagner, K. D., Birmaher, B., Geller, B. et al. (1999). The Texas Children's Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Childhood Major Depressive Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1442-1454.

Hughes, C. W., & Melson, A. G. (2008). Child and adolescent measures for diagnosis and screening. In A. J. Rush, H. A. Pincus, & M. B. First (Eds.) *Handbook of Psychiatric Measures* (pp. 251-308). Washington, DC: American Psychiatric Publishing, Inc.

Humphreys, L. G., & Ilgen, D. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measures, 29,* 571-578.

Humphreys, L. G.,& Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10,* 193-206.

Jain, S., Carmody, T. J., Trivedi, M. H., Hughes, C., Bernstein, I. H., Morris, D. W. et al. (2007). A psychometric evaluation of the CDRS and MADRS in assessing depressive symptoms in children. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 1204-1212.

Johnson, J. G., Harris, E. S., Spitzer, R. L., & Williams, J. B. (2002). The patient health questionnaire for adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *Journal of Adolescent Health, 30*(3), 196-204.

Kandel, D. B. & Davies, M. (1986). Adult sequelae of adolescent depressive symptoms. *Archives of General Psychiatry, 43,* 255-262.

Kashani, J. H., Beck, N. C., Hoeper, E. W., Fallahi, C., Corcoran, C. M., McAllister, J. A. et al. (1987). Psychiatric disorders in a community sample of adolescents. *American Journal of Psychiatry, 144,* 584-589.

Kashani, J. H., Carlson, G. A., Beck, N. C., Hoeper, E. W., Corcoran, C. M., McAllister, J. A. et al. (1987). Depression, depressive symptoms, and depressed mood among a community sample of adolescents. *American Journal of Psychiatry, 144,* 931-934.

Kazdin, A. E. (1989). Identifying depression in children: a comparison of alternative selection criteria. *Journal of Abnormal Child Psychology, 17,* 437-454.

Keller, M. B. (2004). Remission versus response: the new gold standard of antidepressant care. *Journal of Clinical Psychiatry, 65*(Suppl. 4), 53-59.

Kennard, B., Silva, S., Vitiello, B., Curry, J., Kratochvil, C., Simons, A. et al. (2006). Remission and residual symptoms after short-term treatment in the Treatment of Adolescents with Depression Study (TADS). *Journal of the American Academy of Child and Adolescent Psychiatry, 45,* 1404-1411.

Kennard, B. D., Emslie, G. J., Mayes, T. L., & Hughes, J. L. (2006). Relapse and recurrence in pediatric depression. *Child and Adolescent Psychiatry Clinics of North America, 15,* 1057-79, xi.

Kennard, B. D., Emslie, G. J., Mayes, T. L., Nightingale-Teresi, J., Nakonezny, P. A., Hughes, J. L. et al. (2008). Cognitive-behavioral therapy to prevent relapse in pediatric responders to pharmacotherapy for major depressive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 47,* 1395-1404.

Kennard, B. D., Silva, S. G., Mayes, T. L., Rohde, P., Hughes, J. L., Vitiello, B. et al. (2009). Assessment of safety and long-term outcomes of initial treatment with placebo in TADS. *American Journal of Psychiatry, 166,* 337-344.

Kessler, R. C., Avenevoli, S., & Ries, M. K. (2001). Mood disorders in children and adolescents: an epidemiologic perspective. *Biological Psychiatry, 49,* 1002-1014.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S. et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry, 51,* 8-19.

Kobak, K. A., Greist, J. H., Jefferson, J. W., Mundt, J. C., & Katzelnick, D. J. (1999). Computerized assessment of depression and anxiety over the telephone using interactive voice response. *MD Computing, 16,* 63-68.

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

Kovacs, M., Feinberg, T. L., Crouse-Novak, M. A., Paulauskas, S. L., & Finkelstein, R. (1984). Depressive disorders in childhood. I. A longitudinal prospective study of characteristics and recovery. *Archives of General Psychiatry, 41,* 229-237.

Kovacs, M., Feinberg, T. L., Crouse-Novak, M., Paulauskas, S. L., Pollock, M., & Finkelstein, R. (1984). Depressive disorders in childhood. II. A longitudinal study of the risk for a subsequent major depression. *Archives of General Psychiatry, 41,* 643-649.

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

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: age at onset, episode duration, and time to recurrence. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 809-818.

Lewinsohn, P. M., Fenn, D. S., Stanton, A. K., & Franklin, J. (1986). Relation of age at onset to duration of episode in unipolar depression. *Psychological Aging, 1,* 63-68.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology, 102,* 133-144.

Lewinsohn, P. M., Roberts, R. E., Seeley, J. R., Rohde, P., Gotlib, I. H., & Hops, H. (1994). Adolescent psychopathology: II. Psychosocial risk factors for depression. *Journal of Abnormal Psychology, 103,* 302-315.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1993). Psychosocial characteristics of adolescents with a history of suicide attempt. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 60-68.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1994). Psychosocial risk factors for future adolescent suicide attempts. *Journal of Consulting and Clinical Psychology, 62,* 297-305.

March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J. et al. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *Journal of the American Medical Association, 292,* 807-820.

McConaughy, S. H., & Achenbach, T. M. (1989). Empirically based assessment of serious emotional disturbance. *Journal of School Psychology, 27,* 91-117.

Mintz, J., Mintz, L. I., Arruda, M. J., & Hwang, S. S. (1992). Treatments of depression and the functional capacity to work. *Archives of General Psychiatry, 49,* 761-768.

Mokros, H. B., Poznanski, E., Grossman, J. A., & Freeman, L. N. (1987). A comparison of child and parent ratings of depression for normal and clinically referred children. *Journal of Child Psychology and Psychiatry, 28,* 613-624.

Montanelli, R. G., & Humphreys, L G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: a Monte Carlo study. *Psychometrika, 41,* 341-348.

Montgomery, S. A., & Asberg, M. (1979). A new depression rating scale designed to be sensitive to change. *British Journal of Psychiatry, 134,* 382-389.

Moore, H. K., Hughes, C. W., Mundt, J. C., Rush, A. J., Macleod, L., Emslie, G. J. et al. (2007). A pilot study of an electronic, adolescent version of the quick inventory of depressive symptomatology. *Journal of Clinical Psychiatry, 68,* 1436-1440.

Mundt, J. C. (1997). Interactive voice response systems in clinical research and treatment. *Psychiatric Services, 48,* 611-612.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus User's Guide*. (5th ed.). Los Angeles, CA: Muthén & Muthén.

Myers, K. & Winters, N. C. (2002a). Ten-year review of rating scales. I: Overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 114-122.

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

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

Olfson, M., Shaffer, D., Marcus, S. C., & Greenberg, T. (2003). Relationship between antidepressant medication treatment and suicide in adolescents. *Archives of General Psychiatry, 60,* 978-982.

Orlando, M., Sherbourne, C. D., & Thissen, D. (2000). Summed-score linking using item response theory: application to depression measurement. *Psychological Assessment, 12,* 354-359.

Pavuluri, M., & Birmaher, B. (2004). A practical guide to using ratings of depression and anxiety in child psychiatric practice. *Current Psychiatry Reports, 6*, 108-116.

Pfeffer, C. R., Klerman, G. L., Hurt, S. W., Lesser, M., Peskin, J. R., & Siefker, C. A. (1991). Suicidal children grow up: demographic and clinical risk factors for adolescent suicide attempts. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 609-616.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Poznanski, E. O., Cook, S. C., & Carroll, B. J. (1979). A depression rating scale for children. *Pediatrics, 64,* 442-450.

Poznanski, E. O., Freeman, L., & Mokros, H. (1985). Children's depression rating scale-revised. *Psychopharmacology Bulletin, 21,* 979-989.

Poznanski, E. O., & Mokros, H. B. (1996). *Children's depression rating scale, revised (CDRS-R) manual*. Los Angeles, CA: Western Psychological Services.

Radloff, L. (1977). The CED-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19*(4), 353-368.

Rao, U., Ryan, N. D., Birmaher, B., Dahl, R. E., Williamson, D. E., Kaufman, J. et al. (1995). Unipolar depression in adolescents: clinical outcome in adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 566-578.

Reinherz, H. Z., Giaconia, R. M., Silverman, A. B., Friedman, A., Pakiz, B., Frost, A. K. et al. (1995). Early psychosocial risks for adolescent suicidal ideation and attempts. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 599-611.

Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale: Professional Manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M., & Mazza, J. J. (1998). Reliability and validity of the Reynolds adolescent depression scale with young adolescents. *Journal of School Psychology, 36,* 295-312.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1994). Are adolescents changed by an episode of major depression? *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 1289-1298.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S., Mundt, J. C. et al. (2006). An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry, 59,* 493-501.

Rush, A. J., Carmody, T. J., Ibrahim, H. M., Trivedi, M. H., Biggs, M. M., Shores-Wilson, K. et al. (2006). Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatric Services, 57,* 829-837.

Rush, A. J., Carmody, T. J., & Reimitz, P. E. (2000). The Inventory of Depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research, 9,* 45-59.

Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J., & Burns, C. (1986). The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Research, 18,* 65-87.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine, 26,* 477-486.

Rush, A. J., Kraemer, H. C., Sackeim, H. A., Fava, M., Trivedi, M. H., Frank, E. et al. (2006). Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology, 31,* 1841-1853.

Rush, A. J., & Trivedi, M. H. (1995). Treating depression to remission. *Psychiatric Annals, 25,* 704-709.

Rush, A. J., Trivedi, M. H., Carmody, T. J., Ibrahim, H. M., Markowitz, J. C., Keitner, G. I. et al. (2005). Self-reported depressive symptom measures: sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients. *Neuropsychopharmacology, 30,* 405-416.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N. et al. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54,* 573-583.

Samejima, F. (1997) Graded response model. In W. van Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). New York: Springer-Verlag.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Shaffer, D., Gould, M. S., Brasic, J., Amrosini, P., Fisher, P., Bird, H., Aluwahlia, S. (1985). A children's global assessment scale (CGAS) (for children 4 to 16 years of age). *Psychopharmacological Bulletin, 21(4),* 747-748.

Shaffer, D., Gould, M. S., Fisher, P., Trautman, P., Moreau, D., Kleinman, M. et al. (1996). Psychiatric diagnosis in child and adolescent suicide. *Archives of General Psychiatry, 53*, 339-348.

Shain, B. N., Naylor, M., & Alessi, N. (1990). Comparison of self-rated and clinician-rated measures of depression in adolescents. *American Journal of Psychiatry, 147*, 793-795.

Shemesh, E., Annunziato, R. A., Shneider, B. L., Newcorn, J. H., Warshaw, J. K., Dugan, C. A. et al. (2005). Parents and clinicians underestimate distress and depression in children who had a transplant. *Pediatric Transplantation, 9,* 673-679.

Stark, K. D., Reynolds, W. M., & Kaslow, N. J. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *Journal of Abnormal Child Psychology, 15,* 91-113.

Tanner, W. P., Jr. & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401-409.

Thissen, D. (2003). *Multilog for Windows*. Lincolnwood, IL: Scientific Software International.

Treatment for Adolescents With Depression Study (TADS) Team. (2003). Treatment for Adolescents With Depression Study (TADS): rationale, design, and methods. *Journal of the American Academy of Child and Adolescent Psychiatry, 42,* 531-542.

Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T. et al. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychological Medicine, 34,* 73-82.

Wagner, K. D., Robb, A. S., Findling, R. L., Jin, J., Gutierrez, M. M., & Heydorn, W. E. (2004). A randomized, placebo-controlled trial of citalopram for the treatment of major depression in children and adolescents. *American Journal of Psychiatry, 161,* 1079-1083.

Warren, W. L. (1997). *Revised Hamilton Rating Scale for Depression (HRSD): Manual.* Los Angeles: Western Psychological Services.

Weissman, M. M., Bruce, M. L., Leaf, P. J., Florio, L. P., & Holzer, C. (1991). Affective Disorders. In L. N. Robins & D. A. Regier (Eds.), *Psychiatric disorders in America: The epidemiologic catchment area study* (pp. 1351-1387). New York: The Free Press.

Weissman, M. M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning self-report scales: Comparison of mothers' and children's reports. *Journal of Nervous and Mental Disease, 168,* 736-740.

Weissman, M. M., Wolk, S., Goldstein, R. B., Moreau, D., Adams, P., Greenwald, S. et al. (1999). Depressed adolescents grown up. *Journal of the American Medical Association, 281,* 1707-1713.

Weisz, J. R. & Hawley, K. M. (2002). Developmental factors in the treatment of adolescents. *Journal of Consulting and Clinical Psychology, 70,* 21-43.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychological Bulletin, 132,* 132-149.

Welner, Z., Reich, W., Herjanic, B., Jung, K. G., & Amado, H. (1987). Reliability, validity, and parent-child agreement studies of the Diagnostic Interview for Children and Adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 649-653.

Wu, P., Hoven, C. W., Bird, H. R., Moore, R. E., Cohen, P., Alegria, M. et al. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1081-1090.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68,* 1038-1050.

Yule, W. (1993). Developmental considerations in child assessment. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child and adolescent assessment* (Vol. 167, pp. 15-25). Boston: Allyn & Bacon.