The Use of Psychological Testing for Treatment Planning and Outcomes Assessment

THIRD EDITION
Volume 2

Instruments for Children and Adolescents

Edited by
Mark E. Maruish

The Use of Psychological Testing for Treatment Planning and Outcomes Assessment

Third Edition

Volume 2 Instruments for Children and Adolescents

The Use of Psychological Testing for Treatment Planning and Outcomes Assessment

Third Edition

Volume 2 Instruments for Children and Adolescents

Edited by

Mark E.Maruish

Southcross Consulting

LEA 2004

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS

Mahwah, New Jersey London

This edition published in the Taylor & Francis e-Library, 2011.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.

Senior Consulting Editor:	Susan Milmoe
Editorial Assistant:	Kristen Depken
Cover Design:	Kathryn Houghtaling Lacey
Textbook Production Manager:	Paul Smolenski
Full-Service Compositor:	Tech Books
Text and Cover Printer:	Hamilton Printing Company

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430
www.erlbaum.com

Library of Congress Cataloging-in-Publication Data

The use of psychological testing for treatment planning and outcomes assessment/edited by Mark E.Maruish.—3rd ed.
p. cm.
Includes bibliographical references and index.
Volume 1: ISBN 0-8058-4329-9 (casebound: alk. paper)—Volume 2: ISBN 0-8058-4330-2 (casebound: alk paper)—Volume 3: ISBN 0-8058-4331-0 (casebound: alk. paper)
1. Psychological tests. 2. Mental illness—Diagnosis. 3. Mental illness—Treatment—Evaluation. 4. Psychiatric rating scales. 5. Outcome assessment (Medical care) I. Maruish, Mark E. (Mark Edward)

RC473.P79U83 2004
616.89′075–dc22 2003025432

ISBN 1-4106-1062-4 Master e-book ISBN

For my family

Contents

Preface

Like other medical and behavioral health care services, the practice of test-based psychological assessment has not entered the era of managed care unscathed. Limitations placed on total moneys allotted for psychological services have had an impact on the practice of psychological testing. However, for those skilled in its use, psychological testing's ability to help quickly identify psychological problems, plan and monitor treatment, and document treatment effectiveness presents many potentially rewarding opportunities during a time when health care organizations must (a) provide problem-focused, time-limited treatment; (b) demonstrate the effectiveness of treatment to payers and patients; and (c) implement quality improvement initiatives.

With the opportunity at hand, it is now up to those with skill and training in psychological assessment to make the most of this opportunity to contribute to (and benefit from) efforts to control health care costs. However, this may not be as simple a task as it would appear. Many trained professionals are likely to have only limited knowledge of how to use test results for planning, monitoring, and assessing the outcomes of psychological interventions, Consequently, although the basic skills are there, many well-trained clinicians—and graduate students as well—need to develop or expand their testing knowledge and skills so as to be better able to apply them for such purposes. This need served as the impetus for the development of the first two editions of this book, and the development of this third edition of the work attests to its continued presence.

In developing the contents of this and the previous editions of this work, it was decided that the most informative and useful approach would be one in which aspects of broad topical areas are addressed separately. The first area has to do with general issues and recommendations to be considered in the use of psychological testing for treatment planning and outcomes assessment in today's behavioral health care environment. The second and third areas address issues related to the use of specific psychological tests and scales for these same purposes, one dealing with child and adolescent instruments, the other dealing with adult instruments. The fourth area concerns the future of psychological testing, including future developments in this area. For the current edition, issues related to future developments have been incorporated into the general considerations section. Because of increased content and a desire to better meet the needs of individual practitioners, each of the three sections is now printed in a separate volume.

Volume 1 of this third edition represents an update and extension of the first and fourth parts of the second edition. It is devoted to general considerations that pertain to the need for and use of psychological testing for treatment planning and outcome assessment. The introductory chapter provides an overview of the status of the health care delivery system today and the ways in which testing can contribute to making the system more cost-effective. Three chapters are devoted to issues related to treatment

planning, whereas five chapters focus on issues related to outcomes assessment. The first of the planning chapters deals with the use of psychological tests for screening purposes in various clinical settings. Screening can serve as the first step in the treatment planning process; for this reason, it is a topic that warrants the reader's attention. The second of these chapters presents a discussion of the research suggesting how testing may be used as a predictor of differential response to treatment and its outcome. Each of these chapters represents updated versions of the original work. The next chapter deals with treatment planning within Prochaska's Transtheoretical Model—a widely accepted and researched approach that takes the patient's stage of readiness to change into consideration in developing and revising treatment plans.

The five chapters on the use of testing for outcomes assessment are complementary. The first provides an overview of the use of testing for outcomes assessment purposes, discussing some of the history of outcomes assessment, its current status, its measures and methods, individualizing outcome assessment, the distinction between clinically and statistically significant differences in outcomes assessment, and some outcomes-related issues that merit further research. The next four chapters expand on the groundwork laid in this chapter. The first of these four presents an updated discussion of a set of specific guidelines that can be valuable to clinicians in their selection of psychological measures for assessing treatment outcomes. These same criteria also are generally applicable to the selection of instruments for treatment planning purposes. Two chapters provide a discussion of statistical procedures and research design issues related to the measurement of treatment progress and outcomes with psychological tests. One chapter specifically addresses the analysis of individual patient data; the other deals with the analysis of group data. As noted in the previous editions of this work, knowledge and skills in these areas are particularly important and needed by clinicians wishing to establish and maintain an effective treatment evaluation process within their particular setting. The other outcomes-related chapter presents a discussion of considerations relevant to the design, implementation, and maintenance of outcomes management programs in behavioral health care settings.

Volume 1 also includes a chapter addressing a frequently neglected topic in discussions of outcomes assessment, that is, ethical considerations related to outcomes assessment. The volume concludes with a future-oriented chapter, written to discuss predictions and recommendations related to the use of psychological assessment for treatment planning and outcomes assessment.

Volumes 2 and 3 address the use of specific psychological instruments for treatment planning and outcome assessment purposes. Volume 2 deals with child and adolescent instruments, with one chapter devoted to a review of the research related to the conceptualization of quality of life (QOL) as is applies to children and how it has evolved over the years. The purpose of this chapter is to present a foundation for the future development of useful measures of child QOL—something that currently appears to be in short supply. Volume 3 focuses on instruments that are exclusively or primarily intended for use with adult populations.

Instruments considered as potential chapter topics for Volumes 1 and 3 were evaluated against several selection criteria, including the popularity of the instrument among clinicians; recognition of its psychometric integrity in professional publications; in the case of recently released instruments, the potential for the instrument to become widely accepted and used; the perceived usefulness of the instrument for treatment planning and outcomes assessment purposes; and the availability of a recognized expert on the instrument (preferably its author) to contribute a chapter to this book. In the end, the

instrument-specific chapters selected for inclusion were those judged most likely to be of the greatest interest and utility to the majority of the book's intended audience.

Each of the chapters in the second edition had previously met these selection criteria; thus, Volumes 2 and 3 consist of updated or completely revised versions of the instrumentation chapters that appeared in the first edition. Both volumes also contain several new chapters discussing instruments that were not included in the second edition for one reason or another (e.g., was not developed at the time, has only recently gained wide acceptance for outcomes assessment purposes). Indeed, recognition of the potential utility of each of these instruments for treatment planning or evaluation served as one impetus for revising the second edition of this work.

A decision regarding the specific content of each of the chapters in Volumes 2 and 3 was not easy to arrive at. However, in the end, the contributors were asked to address those issues and questions that are of the greatest concern or relevancy for practicing clinicians. Generally, these fall into three important areas: (1) What the instrument does and how it was developed; (2) how one should use this instrument for treatment planning and monitoring; and (3) how it should be used to assess treatment outcomes. Guidelines were provided to assist the contributors in addressing each of these areas. Many of the contributors adhered strictly to these guidelines; others modified the contents of their chapter to reflect and emphasize what they judged to be important to the reader to know about the instrument when using the it for planning, monitoring, or outcome assessment purposes. Some may consider the chapters in Volumes 2 and 3 to be the "meat" of this revised work, because they provide "how to" instructions for tools that are commonly found in the clinician's armamentarium of assessment instruments. In fact, these chapters are no more or less important than those found in Volume 1. They are only extensions and are of limited value outside of the context of the chapters in Volume 1.

As was the case with the previous two editions, the third edition of *The Use of Psychological Testing for Treatment Planning and Evaluation* is not intended to be a definitive work on the topic. However, it is hoped that the reader will find its chapters useful in better understanding general and test-specific considerations and approaches related to treatment planning and outcomes assessment, and in effectively applying them in his or her daily practice. It also is hoped that it will stimulate further endeavors in investigating the application of psychological testing for these purposes.

—Mark E.Maruish
Minneapolis, MN

List of Contributors

Brian V.Abbott
Texas A&M University
College Station, TX

Thomas M.Achenbach
University of Vermont
Burlington, VT

Ross B.Andelman
Contra Costa Children's Mental Health Services
Concord, CA

Robert P.Archer
Eastern Virginia Medical School
Norfolk, VA

C.Clifford Attkisson
University of California
San Francisco, CA

Steven E.Bailley
University of Texas-Houston Health Sciences Center
Houston, TX

Thomas Beers
Kaiser Permanente San Diego Chemical Dependency Program
San Diego, CA

Albert J.Belanger
Harvard Medical School
Boston, MA

Larry E.Beutler
University of California
Santa Barbara, CA

Phillip J.Brantley
Pennington Biomedical Research Center
Baton Rouge, LA

Gary M.Burlingame
Brigham Young University
Provo, UT

James N.Butcher
University of Minnesota
Minneapolis, MN

David L.Carlston
Ohio University
Athens, OH

Antonio Cepeda-Benito
Texas A&M University
College Station, TX

Dianne L.Chambless
University of Pennsylvania
Philadelphia, PA

James A.Ciarlo
University of Denver
Denver, CO

Paul D.Cleary
Harvard Medical School
Boston, MA

James R.Clopton
Texas Tech University
Lubbock, TX

John D.Cone
Alliant International University
San Diego, CA

C.Keith Conners
Duke University School of Medicine
Durham, NC

Jonathan C.Cox
Brigham Young University
Provo, UT

William J.Culpepper
University of Maryland
Baltimore, MD

Constance J.Dahlberg
Alliant International University
San Diego, CA

Allen S.Daniels
Alliance Behavioral Care, University of Cincinnati
Cincinnati, OH

Edwin de Beurs
Leiden University Medical Center Leiden, The Netherlands

Leonard R.Derogatis
Johns Hopkins University School of Medicine
Baltimore, MD

Kathy Dowell
Ohio University
Athens, OH

Gareth R.Dutton
Louisiana State University
Baton Rouge, LA

William W.Eaton
Johns Hopkins University, Bloomberg School of Public Health
Baltimore, MD

Susan V.Eisen
Center for Health Quality, Outcomes, and Economic Research, Edith Nourse Rogers Veterans Hospital
Boston, MA

Jeffery N.Epstein
Duke University School of Medicine
Durham, NC

Alex Espadas
University of Texas-Houston Health Sciences Center
Houston, TX

Laura E.Evison
Johns Hopkins University School of Medicine
Baltimore, MD

Kya Fawley
Northwestern University
Evanston, IL

Maureen Fitzpatrick
Johns Hopkins University School of Medicine
Baltimore, MD

Jenny Fleming
University of California
Santa Barbara, CA

Michael B.Frisch
Baylor University
Waco, TX

Anthony B.Gerard
Western Psychological Services
Los Angeles, CA

Sona Gevorkian
Massachusetts General Hospital
Boston, MA

David H.Cleaves
Texas A&M University
College Station, TX

Pamela Greenberg
American Managed Behavioral Healthcare Association
Washington, DC

Roger L.Greene
Pacific Graduate School of Psychology
Palo Alto, CA

Thomas K.Greenfield
University of California and Public Health Institute Berkeley
San Francisco, CA

Ann T.Gregersen
Brigham Young University
Provo, UT

Grant R.Grissom
Polaris Health Directions
Langhorne, PA

Seth D.Grossman
Institute for Advanced Studies in Personology
Coral Gables, FL

Kurt Hahlweg
Technical University of Braunschweig
Braunschweig, Germany

Steven R.Hahn
Albert Einstein College of Medicine
New York, NY

Ashley E.Hanson
University of Alabama
Tuscaloosa, AL

Nancy M.Hatcher
University of Georgia
Athens, GA

Derek Hatfield
Ohio University
Athens, OH

Eric J.Hawkins
Brigham Young University
Provo, UT

Jena Helgerson
Northwestern University
Evanston, IL

Kay Hodges
Eastern Michigan University
Ann Arbor, MI

Elizabeth A.Irvin
Services Research Group, Inc. and Simmons College, Graduate School of Social Work
Boston, MA

Gary Jeager
Kaiser Permanente Harbor City Chemical Dependency Program
Harbor City, CA

R.W.Kamphaus
University of Georgia
Athens, GA

Jennifer M.Karpe
University of Alabama
Tuscaloosa, AL

Sangwon Kim
University of Georgia
Athens, GA

Kenneth A.Kobak
Dean Foundation for Health Research and Education
Madison, WI

Scott H.Kollins
Duke University School of Medicine
Durham, NC

Teresa L.Kramer
University of Arkansas for Medical Sciences
Little Rock, AR

Kurt Kroenke
Regenstrief Institute for Health Care, Indiana University School of Medicine
Indianapolis, IN

Samuel E.Krug
MetriTech, Inc.

Champaign, IL
David Lachar
University of Texas-Houston Health Sciences Center
Houston, TX

Michael J.Lambert
Brigham Young University
Provo, UT

Jeanne M.Landgraf
HealthAct
Boston, MA

William W.Latimer
Johns Hopkins University
Baltimore, MD

Jean-Philippe Laurenceau
University of Miami
Miami, FL

John S.Lyons
Northwestern University
Evanston, IL

Melanie Buddin Lyons
Buddin Praed Foundation
Winnetka, IL

Mary Malik
University of California
Santa Barbara, CA

John S.March
Duke University Medical Center
Durham, NC

Mark E.Maruish
Southcross Consulting
Burnsville, MN

Sarah E.Meagher
University of Miami
Miami, FL

Gregorio Melendez
Ohio University
Athens, OH

Theodore Millon
Institute for Advanced Studies in Personology and Psychopathology
Coral Gables, FL

Carla Moleiro
University of California
Santa Barbara, CA

Leslie C.Morey
Texas A&M University
College Station, TX

Carles Muntaner
University of Maryland School of Nursing
College Park, MD

Jack A.Naglieri
George Mason University
Fairfax, VA

Charles Negy
University of Central Florida
Orlando, FL

Frederick L.Newman
Florida International University
Miami, FL

Sharon-Lise T.Normand
Harvard Medical School and Harvard School of Public Health
Boston, MA

Benjamin M.Ogles
Ohio University
Athens, OH

Ashley E.Owen
University of South Florida
Tampa, FL

James D.A.Parker
Trent University
Peterborough, ON, Canada

Julia N.Perry
Veteran's Administration Hospital

Minneapolis, MN
Steven I.Pfeiffer
Duke University
Durham, NC

James O.Prochaska
Cancer Prevention Research Center
Kingston, RI

Janice M.Prochaska
Pro-Change Behavior Systems, Inc.
Kingston, RI

Eric C.Reheiser
University of South Florida
Tampa, FL

Leslie A.Rescorla
Bryn Mawr College
Bryn Mawr, PA

Cecil R.Reynolds
Texas A&M University
College Station, TX

William M.Reynolds
Humboldt State University
Arcata, CA

James M.Robbins
University of Arkansas for Medical Sciences
Little Rock, AR

Abram B.Rosenblatt
University of California
San Francisco, CA

Douglas Rugh
Florida International University
Miami, FL

Scott Sangsland
Kaiser Permanente
Southern California Permanente Medical Group
Pasadena, CA

Forrest R.Scogin
University of Alabama
Tuscaloosa, AL
James A.Shaul
Harvard Medical School
Boston, MA

Gill Sitarenios
Multi-Health Systems, Inc.
Toronto, ON, Canada

Corey Smith
Johns Hopkins University, Bloomberg School of Public Health
Baltimore, MD

G.Richard Smith
University of Arkansas for Medical Sciences
Little Rock, AR

Douglas K.Snyder
Texas A&M University
College Station, TX

Charles D.Spielberger
University of South Florida
Tampa, FL

Robert L.Spitzer
New York State Psychiatric Institute
New York, NY

Steven Stein
Multi-Health Systems, Inc.
Toronto, ON, Canada

Randy Stinchfield
University of Minnesota
Minneapolis, MN

Sumner J. Sydeman
Northern Arizona University
Flagstaff, AZ

Elana Sydney
Albert Einstein College of Medicine
New York, NY

Hani Talebi
University of California
Santa Barbara, CA

Manuel J.Tejeda
Barry University
Miami Shores, FL

Allen Tien
MDLogix, Inc.
Towson, MD

John E.Ware, Jr.
QualityMetric Inc. and Tufts University Medical School
Lincoln, RI

Dana Aron Weiner
Northwestern University
Evanston, IL

Irving B.Weiner
University of South Florida
Tampa, FL

M.Gawain Wells
Brigham Young University
Provo, UT

Douglas L.Welsh
University of Alabama
Tuscaloosa, AL

Janet B.W.Williams
New York State Psychiatric Institute
New York, NY

Kimberly A.Wilson
Stanford University Medical School
Palo Alto, CA

Ken C.Winters
University of Minnesota
Minneapolis, MN

Stephen E.Wong
Florida International University
Miami, FL

Karen B.Wood
Louisiana State University
Baton Rouge, LA

Michele Ybarra
Johns Hopkins University, Bloomberg School of Public Health
Baltimore, MD

The Use of Psychological Testing for Treatment Planning and Outcomes Assessment

Third Edition

Volume 2 Instruments for Children and Adolescents

1

Use of the Children's Depression Inventory

Gill Sitarenios and Steven Stein
Multi-Health Systems, Inc.

CHILDHOOD DEPRESSION

From a clinical perspective, a syndrome is a characteristic constellation of psychopathologic symptoms and signs. A depressive syndrome typically encompasses a negative dysphoric mood and complaints such as a sense of worthlessness or hopelessness, preoccupation with death or suicide, difficulties in concentration or making decisions, disturbance in patterns of sleep and food intake, and reduced energy. A disorder is a particular syndrome that has been shown to have the characteristics of a diagnosable condition. That is, it has a recognizable pattern of onset and course, clear negative consequences with respect to the individual's functioning, distinct biologic or related correlates, an association with known etiologic or risk factors, and a course that may be altered in predictable ways by various treatments.

Major depressive disorder and dysthymic disorder are two forms of depressive disorder that affect children as well as adults. Episodes of major depression in childhood last about 10 months on average and may have psychotic or melancholic features associated with them (Kovacs, Obrosky, Gatsonis, & Richards, 1997). Major depression often is comorbid with other disorders, most commonly with disorders of anxiety and conduct (Kovacs, Gatsonis, Paulauskas, & Richards, 1989; Puig-Antich, 1982; Strober & Carlson, 1982). Major depression in childhood is associated with a high rate of recovery; there is, however, a very high risk of episode recurrence, and an increased risk for the development of other related disorders (Kovacs, 1996a, 1996b; Kovacs et al,, 1989; Strober & Carlson, 1982). Compared with major depression, dysthymic disorder is milder and possibly less impairing. However, dysthymia usually lasts longer than major depression, with an average duration of about 3 and a half years or longer (Kovacs et al., 1997). Like major depression, dysthymia has a high rate of eventual recovery. Dysthymia is associated with a high rate of comorbid psychiatric disorders and increases the risk for major depression and other related conditions (Kovacs, Akiskal, Gatsonis, & Parrone, 1994; Kovacs et al., 1997).

Weiss et al. (1991) noted that depression in childhood, which was once thought to be rare or nonexistent, is now the subject of much clinical and research activity and is currently recognized by almost all authoritative sources (e.g., *The Diagnostic and Statistical Manual of Mental Disorders,* American Psychiatric Association, 1994). In

fact, estimates of prevalence rates of depressive disorders in children have been found to be quite high (e.g., see Kashani et al., 1981), and some clinicians have diagnosed them as early as preschool age (e.g., Kashani & Carlson, 1985). The pattern of symptoms seen in childhood depression is similar to that seen in adults with similar affective, cognitive, behavioral, and somatic complaints (Kaslow, Rehm, & Siegel, 1984), and there appears to be little variability in the associated features of the disorder across the life span (Kovacs, 1996a). Depressive disorders can disrupt the functioning of children and adolescents in a number of areas—most notably in school— and cause significant developmental delays. Moreover, children who have depressive disorders may have trouble "catching up" in development (Kovacs & Goldston, 1991, p. 189).

ASSESSMENT OF DEPRESSION USING SELF-REPORT

Assessment of depression can focus on (a) the early identification of the extent and severity of depressive symptoms, (b) the diagnosis of depression and associated disorders, and (c) the monitoring the effectiveness of interventions.

Self-rated inventories have long been a part of the assessment of depressive symptoms in adults (e.g., Beck Depression Inventory; Beck, 1967). Such inventories typically are easy to administer, inexpensive, and readily analyzable. Because they quantify the severity of the depressive syndrome, they have been used for descriptive purposes, to assess treatment outcomes, to test research hypotheses, and to select research subjects. However, because self-rated inventories do not assess the temporal features, the onset, the course, or the contributing factors of the syndrome being examined, they cannot yield diagnostic information.

For children, self-report inventories nonetheless provide especially useful information in that many features of depression are internal and are not easily identified by informants such as parents or teachers. Moreover, according to psychological models, children's self-perceptions are of predictive value in their own right (Kovacs, 1992; Saylor, Finch, Baskin, Furey, & Kelly, 1984).

The Children's Depression Inventory (GDI) has been one of the most widely used and cited inventories of depression. According to a recent report by Fristad, Emery, and Beck (1997), the GDI was used in over 75% of the studies with children in which self-report depression inventories were employed.

The initial version of the GDI was developed in 1977. Formal publication of the instrument in 1992 increased its accessibility. This chapter provides a timely opportunity to summarize the research history and usage of the GDI since its inception 25 years ago and since its publication about a decade ago. The GDI, as well as its various versions, associated manuals, and scoring forms are described in the first part of this chapter. Current research and theory related to the GDI are also highlighted. The GDI manual (Kovacs, 1992) includes an annotated bibliography of about 150 related research studies up to the end of 1991. At least 200 additional articles pertaining to the GDI had been published by 1997 (Fristad et al, 1997).

Other goals of this chapter are to examine current use of the GDI, distinguish proper from improper use of the instrument, and address questions frequently asked by practitioners. The GDI can be helpful in the early identification of symptoms and in the monitoring of treatment effectiveness. The GDI also can play a role in the diagnostic process, but, as already noted, *it should not be used alone* to diagnose a depressive disorder. Finally, this chapter describes the ongoing development of the GDI, including anticipated accessories, future research directions, and extended applications.

SUMMARY OF THE DEVELOPMENT OF THE GDI

The Beck Depression Inventory (Beck, 1967), a clinically based, 21-item, self-rated symptom scale for adults, was the starting point for the development of a paper-andpencil tool that would be appropriate for children. The research literature supported the decision to use an "adult" scale as the model, given that there appeared to be much overlap between the salient manifestations of depressive disorders in juveniles and in adults (Kovacs & Beck, 1977). Scale construction proceeded in four phases.

Phase I

The first version of the children's inventory (dated March 1975) was derived with the help of a group of 10- to 15-year-old "normal" youths and similar-aged children from an urban inpatient and partial hospitalization program. After the purpose of the scale revision project was explained individually to each child, he or she was asked for advice on how the items could be worded to make them "clear to kids." In this phase of scale construction, the Beck item pertaining to sexual interest was replaced by an item on loneliness, but the content and format of 20 items of the adult scale were essentially retained. However, five "Appendix" items, adapted from Albert and Beck (1975), were added; these concerned school and peer functioning. Piloting yielded further semantic changes.

Phase II

Data from normal youths and children who were under psychiatric-psychological care were used along with a semantic and conceptual item analysis to produce a second major revision (dated February 1976) that also included a new item on selfblame. This version of the inventory was administered to thirty-nine 8- to 13-year-old children who were consecutively admitted to a child guidance center's hospitalization units, twenty "normal" 8- to 13-year-olds with no history of psychiatric contacts, and one hundred and twenty-seven 10- to 13-year-old fifth- and sixth-grade students in the Toronto public school system.

The resultant data were analyzed according to standard psychometric principles, and the findings were used to derive a completely new version of the scale. Two of the original 21 items (shame and weight loss) and two of the appendix items (family fights and self-blame) were replaced by four new items that had face validity and appeared age appropriate (e.g., feeling unloved).

The GDI item-choice distributions in these samples also revealed that the items could be recast into a three-choice format: one choice reflects "normalcy," the middle choice pertains to definite although not disabling symptom severity, and the other response option reflects a clinically significant complaint. In order to prevent response bias, approximately 50% the items (randomly selected) were worded so that the first response choice suggested the most pathology, and the response choice order was reversed for the remaining items.

Phase III

The newly modified version of the GDI (dated May 1977) was again pilot-tested and sent to colleagues for a critique, A cover page was added with revised instructions and a sample item. Based on the results of pilot-testing, the items were further refined and reworded in order to improve face validity and comprehensibility.

TABLE 1.1
Authorized GDI Translations

Afrikaans	Japanese
Dutch	Lithuanian
French (European)	Norwegian
French (Canadian)	Polish
German	Russian
Greek	Spanish
Hebrew	Swedish
Icelandic	Turkish
Italian	Ukrainian

Phase IV

One minor change preceded preparation of the final version of the GDI (dated August 1979). The score values were eliminated from the inventory, and scoring templates were developed.

Current Work

Since the initial development of the GDI, additional psychometric analyses have been conducted. Based on these analyses, five factors have been identified and are fully described in the GDI manual (Kovacs, 1992). A short form of the GDI has been derived as well, and software has been developed for online administration, scoring, and reporting. The instrument is now available in several foreign languages. A listing of available translations appears in Table 1.1.

OVERVIEW OF THE GDI

The GDI is appropriate for children and adolescents aged 7 to 17 years. The instrument quantifies a range of depressive symptoms, including disturbed mood, problems in hedonic capacity and vegetative functions, low self-evaluation, hopelessness, and difficulties in interpersonal behaviors. Several items pertain to the consequences of depression with respect to contexts that are specifically relevant to children (e.g., school). Each of the 27 GDI items consists of three choices, keyed 0 (absence of a symptom), 1 (mild symptom), or 2 (definite symptom), with higher scores indicating increasing severity. The total scale score can range from 0 to54.

In addition to the total score, the GDI also yields scores for five factors or subscales: Negative Mood, Interpersonal Problems, Ineffectiveness, Anhedonia, and Negative Self-Esteem. Although author-approved definitions of these subscales have been available to users for some time, the definitions have not been widely published (although they are given in the recent *Software User's Manual;* Kovacs, 1995). Therefore, these definitions are provided in Table 1.2.

Reliability

Psychometric information on reliability is directly related to the proper use and interpretation of an instrument. The reliability of the GDI has been examined in terms of internal consistency, test-retest reliability, and standard error.

TABLE 1.2
Definitions of the Subscales of the GDI

Scale	*Definition*
Negative Mood	This subscale reflects feeling sad, feeling like crying, worrying about "bad things," being bothered or upset by things, and being unable to make up one's mind
Interpersonal Problems	This subscale reflects problems and difficulties in interactions with people, including trouble getting along with people, social avoidance, and social isolation
Ineffectiveness	This subscale reflects negative evaluation of one's ability and school performance
Anhedonia	This subscale reflects "endogenous depression," including impaired ability to experience pleasure, loss of energy, problems with sleeping and appetite, and a sense of isolation
Negative Self-Esteem	This subscale reflects low self-esteem, self-dislike, feelings of being unloved, and a tendency to have thoughts of suicide

TABLE 1.3
Estimates of Internal Consistency of the CDI and the Five CDI Factors

Scale	*Internal Consistency (Cronbach's Alpha)*
Total CDI	Alphas ranging from .71 to .89 (Kovacs, 1992)
Negative Mood	Normative sample: .62; Canadian sample: .65
Interpersonal Problems	Normative sample: .59; Canadian sample: .60
Ineffectiveness	Normative sample: .63; Canadian sample: .59
Anhedonia	Normative sample: .66; Canadian sample: .64
Negative Self-Esteem	Normative sample: .68; Canadian sample: .66

Internal Consistency. Internal consistency refers to the fact that all items on the given instrument consistently measure the same dimension. Kovacs (1992) summarized several research studies that reported alpha reliability statistics for the CDI. Alpha coefficients from .60 to .70 are usually taken to indicate satisfactory reliability (DeVellis, 1991), .70 to .80 indicate good reliability, and .80 to .95 indicate excellent reliability. The majority of the studies reported total score alpha values over .80, and all of the values were greater than .70. For instance, Kovacs (1985) found the total score coefficient alpha to be .86 for a heterogeneous, psychiatric referred sample of children, .71 for a pediatric-medical outpatient group, and .87 for a large sample of public school students (N=860).

Although the internal consistency of the CDI total score has often been reported, data on alpha coefficients for the five factor scores have been less available. Therefore, the internal consistency of the five subscales was assessed using two large data sets: the CDI normative sample of 1,266 children and an independent sample of 894 Canadian children. The reliability values obtained are shown in Table 1.3, along with a summary of alpha values previously reported for the CDI total score.

Although the reliability for the five subscales is not as high as for the CDI total score, the findings for the subscales are satisfactory. Furthermore, the alpha values obtained from the two samples are very similar.

Test-Retest Reliability. The GDI is completed based on the respondent's feelings, moods, and functioning during the 2-week period just prior to the test administration. Thus, rather than measuring traits, which are less changeable over time, the inventory measures state symptoms. Because the GDI measures a state rather than a trait, the retest interval for assessing reliability should be short (2 to 4 weeks). In the research reviewed by Kovacs (1992), studies done with normal youths and psychiatric inpatients using such short intervals (Finch, Saylor, Edwards, & Mclntosh, 1987; Kaslow et al, 1984; Meyer, Dyck, & Petrinack, 1989; Nelson & Politano, 1990; Saylor, Finch, Spirito, & Bennett, 1984; Wierzbicki, 1987) found testretest correlations between .56 to .87 (an outlier of .38 was obtained in one study), and the median test-retest correlation was .75. Thus, the GDI has acceptable short-term stability.

Standard Error. Two types of standard error (Lord & Novick, 1968) are most relevant to the GDI: standard error of measurement *(SEM_1)* and standard error of prediction *(SEM_2)*. SEM_1 is calculated using Cronbach's alpha and represents the standard deviation of *observed* scores if the true score is held constant. This means that, if parallel forms are used to assess the same individual at the same time, about 68% of the scores would fall within a 1 SEM_1 unit of the score obtained on the GDI scale and about 95% of the scores would fall within 1.96 SEM_1 units.

SEM_2 has particular relevance because it has an intimate connection to outcomes assessment. SEM_2 is calculated using the test-retest coefficient and represents the standard deviation of *predicted* scores if the obtained score is held constant. That is, if 100 individuals were reassessed on the GDI, about 68% of the retest scores would fall within 1 SEM_2 unit of the predicted scores and about 95% of the retest scores would fall within 1.96 SEM_2 units of the predicted scores. Thus, the SEM_2 value is one way of assessing how much GDI scores can be expected to change due to random fluctuation. Any change in GDI scores that substantially exceeds the expected random fluctuation is most likely attributable to a significant change in the status of the individual's symptoms.

The absolute value for SEM_1 or SEM_2 varies according to both the estimate of reliability and the estimate of the population standard deviation used in the calculation. The above noted SEM_1 value was calculated based on the median Cronbach alpha for the GDI total score, shown in Table 1.3, and SEM_2 values were derived using the median 2- to 4-week test-retest reliability estimate for the GDI total score. The resultant values for standard error of measurement are presented in Table 1.4.

TABLE 1.4
Standard Error Values for the GDI Total Score

Gender (Age Group)	*Standard Error of Measurement* (SEM_1)	*Standard Error of Prediction* (SEM_2)
Boys (overall)	2.9	3.8
Boys (7–12)	2.8	3.7
Boys (13–17)	3.1	4.2
Girls (overall)	2.6	3.5
Girls (7–12)	2.7	3.6
Girls (13–17)	2.4	3.2
Overall	2.7	3.7

Validity

The validity of an instrument is evaluated by estimating the extent to which it correctly measures the construct or constructs that it purports to assess. Constructs cannot be directly observed, so validity is assessed through empirical means. Specifically, construct validity is assessed through its correlation with other scales purported to measure the same construct, by its correlation with scales purported to measure related constructs, or by its correlation with independent ratings of behavior. Other aspects of validation include factor analyses examining the scale's subscale structure (factorial validity) and its ability to predict appropriate behaviors (predictive validity). Thus, the validity of a test rests on accumulated evidence from a number of studies using various methodologies (Campbell & Fiske, 1959).

The GDI has been utilized in hundreds of clinical and experimental research studies, and its validity has been well established using a variety of techniques. Overall, the weight of the evidence indicates that the inventory assesses important constructs that have strong explanatory and predictive utility in the characterization of depressive symptoms in children and adolescents. Table 1.5 lists some of the research related to different aspects of validity. Also, see Barreto (1994) for a brief review of validity information and Saylor, Finch, Baskin, et al. (1984) and Saylor, Finch, Spirito, et al. (1984), who used the multitrait, multimethod approach to assess the construct validity of the CDI. Further validation data pertinent to specific uses of the CDI are presented later in this chapter (see the section entitled "Use of the CDI for Clinical Purposes").

META-ANALYSIS OF THE CDI

Twenge and Nolen-Hoeksema (2002) conducted a within-scale meta-analysis using the CDI to examine children and adolescents with depressive symptoms. The studies included were examined in terms of age, gender, birth cohort, race, and class differences. Whereas a traditional meta-analysis computes an effect size for each study, a within-scale meta-analysis utilizes the sample means. A within-scale meta-analysis was used because it allows for generalization over many domains, gathering data that were collected at many different locations and times. The authors argued that this form of analysis is the best method for examining individual differences in CDI scores. They recognized that the chosen analytic method is limited to examining only one measure but asserted that the focus on the CDI was well justified because it is the most frequently used scale measuring depressive symptoms of children. Research studies were located using the Web of Science Citation Index, the Science Citation Index, and the Arts and Humanities Citation Index. Several criteria were used to select studies for inclusion. First, samples had to be from the United States or Canada. Second, each study had to include at least 15 subjects. Third, retained samples could not consist of psychiatric patients, delinquents, hospital patients, people diagnosed with any particular disease, or any other group singled out for maladjustment. Fourth, the samples had to be unselected groups (e.g., not specifically high or low depression groups and not groups that would be extremely high or low on any measure that might be correlated with the CDI). Fifth, the CDI mean scores had to be included in the research report.

In total, 310 data sets were included in the meta-analysis, representing 61,424 children (29,637 boys and 31,787 girls) between the ages of 8 and 16.

TABLE 1.5
Studies Containing Information Relevant to the Validity of the GDI

Reference	*Salient Measures or Methodology*
Construct Validity	
CDI compared with other measures of childhood depression	
Bodiford, Eisenstadt, Johnson, & Bradlyn, 1988	CBCL
Hammen et al., 1987	"
Hepperlin, Stewart, & Rey, 1990	"
Lam, 2000	"
Weiss & Weisz, 1988	"
Wolfe et al., 1987	"
Worchel et al., 1990	"
Nieminen & Matson, 1989	RADS
Shain, Naylor, & Alesi, 1990	RADS, Hamilton
Faulstich, Carey, Ruggiero, Enyart, & Gresham, 1986	CESD
Felner, Rowlison, Raley, & Evans, 1988	"
Weissman, Orvaschel, & Padian, 1980	CESD and SAS
Bartell & Reynolds, 1986	CDS
Haley, Fine, Marriage, Moretti, & Freeman, 1985	CDS and others
Rotundo & Hensley, 1985	CDS
Seligman et al., 1984	BDI
Lipovsky, Finch, & Belter, 1989	MMPI–D
Asarnow & Carlson, 1985	DSRS
CDI compared with measures of related constructs	
Eason, Finch, Brasted, & Saylor, 1985	Anxiety (RCMAS)
Felner, Rowlison, Raley, & Evans, 1988	"
Kovacs, 1985	"
Norvell, Brophy, & Finch, 1985	"
Ollendick & Yule, 1990	"
Blumberg & Izard, 1986	Anxiety (STAI)
Wolfe et al., 1987	
Allen & Tarnowski, 1989	Self-concept (Piers-Harris)
Elliott & Tarnowski, 1990	"
Knight, Hensley, & Waters, 1988	"
Kovacs, 1985	"
McCauley, Mitchell, Burke, & Moss, 1988	"
Rotundo & Hensley, 1985	"
Saylor, Finch, Baskin, Furey, & Kelly, 1984	"
Saylor, Finch, Spirito, & Bennett, 1984	"
Kaslow, Rehm, & Siegel, 1984	Self-esteem (Coopersmith)
Kovacs, 1985	"
Reynolds, Anderson, & Bartell, 1985	"
Kazdin, French, Unis, & Esveldt-Dawson, 1983	Self-esteem (Self-Esteem Inventory)
Bodiford, Eisenstadt, Johnson, & Bradlyn, 1988	Attributional style (CASQ)
Curry & Craighead, 1990	"
Gladstone & Kaslow, 1995	"
Hammen, Adrian, & Hiroto, 1988	"
Kuttner, Delamater, & Santiago, 1989	"
McCauley, Mitchell, Burke, & Moss, 1988	"
Nolen-Hoeksema, Girgus, & Seligman, 1986	"
Elliott & Tarnowski, 1990	Hopelessness (Hopelessness Scale)
Kazdin, French, Unis, & Esveldt-Dawson, 1983	"
Kazdin, French, Unis, Esveldt-Dawson, & Sherick, 1983	"
McCauley, Mitchell, Burke, & Moss, 1988	"
Spirito, Overholser, & Hart, 1991	"
Fauber, Forehand, Long, Burke, & Faust, 1987	Perceived Competence Scale
Weissman, Orvaschel, & Padian, 1980	Social Adjustment Scale

(*Continued*)

TABLE 1.5
(*Continued*)

Reference	*Salient Measures or Methodology*
CDI compared with behavioral measures/observations of depressive behavior/symptoms	
Blumberg & Izard, 1986	Parent/teacher rating/observation
Huddleston & Rust, 1994	"
Ines & Sacco, 1992	"
Renouf & Kovacs, 1994	"
Reynolds, Anderson, & Bartell, 1985	"
Sacco & Graves, 1985	"
Shah & Morgan, 1996	"
Slotkin, Forehand, Fauber, McCombs, & Long, 1988	"
Breen & Weinberger, 1995	Therapist/staff ratings
Stocker, 1994	Perceptions of relationships/adjustment
Hodges, 1990	Interview findings
Saylor, Finch, Baskin, Furey, & Kelly, 1984	Peer reports
Factorial Validity	
Carey, Faulstich, Gresham, Ruggiero, & Enyart, 1987	
Helsel & Matson, 1984	
Kovacs, 1992	
Lam, 2000	
Saylor, Finch, Spirito, & Bennett, 1984	
Weiss & Weisz, 1988	
Weiss et al., 1991	
Predictive Validity	
Devine, Kempton, & Forehand, 1994	Longitudinal procedure used
DuBois, Felner, Bartels, & Silverman, 1995	"
Mattison, Handford, Kales, Goodman, & McLaughlin, 1990	"
Reinherz, Frost, & Pakiz, 1991	"
Marciano & Kazdin, 1994	Statistical prediction procedure used
Slotkin, Forehand, Fauber, McCombs, & Long, 1988	"

Means and Standard Deviations Relative to the Existing GDI Norms

The norms used in the current version of the GDI are based on a sample of 1,266 children that are described in detail in the GDI manual (Kovacs, 1992) and in a report by Finch, Saylor, and Edwards (1985). Although the means and standard deviations provided in Twenge and Nolen-Hoeksema's (2002) meta-analysis do not constitute GDI norms, the large samples based on unselected, nonclinical groups makes for an intriguing comparison. The meta-analysis mean values and GDI normative values are shown comparatively in Table 1.6. For girls, the means and standard deviations from the existing GDI norms match up extremely well with the values from the metaanalysis. For boys, however, the GDI norms are notably higher than the values obtained in the meta-analysis. The upcoming GDI restandardization will provide the information needed to determine if these differences require changes in the male GDI norms.

Age and Gender Differences

For boys, there was no relationship between age and depression scores, although the mean for 12-year-old boys was considerably higher than the mean observed for boys of other ages. It is possible that this "spike" in the data might reflect the difficulties in

TABLE 1.6
Boys' and Girls' Scores and Standard Deviations by Age on the Children's Depression Inventory

Source	*Age/Sex*	M	SD
Meta-analysis	8–12/boys	8.5–9.9	7.2–7.9
CDI existing norms	7–12/boys	10.8	7.4
Meta-analysis	13–16/boys	8.7–9.1	6.4–7.1
CDI existing norms	13–17/boys	11.4	8.3
Meta-analysis	8–12/girls	8.4–9.4	7.0–7.7
CDI existing norms	7–12/girls	9.0	7.1
Meta-analysis	13–16/girls	9.1–10.5	6.7–7.3
CDI existing norms	13–17/girls	9.7	6.3

coping with the onset of puberty occurring at about that age. For girls between 8 to 13 years of age, GDI scores and age, again, were unrelated. Also, as with the boys, 12-year-olds yielded the highest score in the 8–13 age bracket. Unlike boys, however, 14- to 16-year-old girls scored considerably higher (range; 10.1–10.5) than younger girls (range: 8.4–9.4).

In terms of gender differences, for children up to 12 years of age, Twenge and NolenHoeksema (2002) observed no significant differences between boys and girls. For 13-to 16-year-olds, however, the scores for girls were significantly higher. The *DSM-IV* (1994, p. 341) notes that Major Depressive Disorder is twice as common in adolescent females as in adolescent males. Although the *DSM-IV* notation pertains to those clinically diagnosed, the meta-analytic finding of greater depressive symptoms in unselected, nonclinical females is certainly consistent with the *DSM-IV* in this regard.

Socioeconomic Status (SES)

All samples included in the meta-analysis were coded as either lower class, lower to middle class, middle class, or middle to upper class. There were no significant correlations with values ranging from r=.03 to r=.06. This result indicates that depression is unrelated to SES in unselected, nonclinical samples.

Race/Ethnicity

Only studies in which 90% or more of the sample were from one racial/ethnic background were used for comparison. Sufficient data were available to perform mean-ingful comparisons between Whites, Blacks, and Hispanics. In total, 109 mixed-sex samples were used. Although there were no significant differences between Whites and Blacks, Hispanics scored significantly higher than both these groups, produc-ing substantial effect sizes (d=0.62 in relation to Whites and d=1.31 in relation to Blacks). The authors noted that the high scores for Hispanics are consistent with some other research findings but indicated that further research is required to fully explain and interpret the results.

CDI Short Form

The 10-item CDI Short Form was developed to enable more rapid and economical assessment of depressive symptoms than the long form. The CDI Short Form can be used

when a quick screening measure is desired or when the examiner's time with the child is limited. The short form takes 5 to 10 minutes to administer, about half the time it takes to administer the long version. However, the long and short forms generally provide comparable results. That is, the correlation between the GDI total score and the GDI Short Form total score was r=.89 (Kovacs, 1992).

ADMINISTRATION OF THE CDI

Reading Level

Past computations of the reading level for the CDI have produced different grade readability estimates (Berndt, Schwartz, & Kaiser, 1983; Kazdin & Petti, 1982). A firstgrade reading level for the CDI is most frequently cited (e.g., Kovacs, 1992). Variable assessments of the instrument's reading level probably reflect the use of different reading level formulas. The Dale-Chall formula (Dale & Chall, 1948) has been found to be the most valid and accurate of the nine commonly utilized readability formulas (e.g., Harrison, 1980). It is based on semantic (word) difficulty and syntactic (sentence) difficulty.

Usually, two 100-word samples are taken to calculate the reading level using the Dale-Chall formula (Chall & Dale, 1995). However, to provide greater accuracy, the computation reported here used all of the CDI items. In accordance with the Dale-Chall standard procedure for determining reading level, the number of complete sentences were counted and divided into the number of words to determine average sentence length (WDS/SEN). Next, the "unfamiliar" words (UFMWDS) were counted. A word is considered unfamiliar if it does not appear on a list of 3,000 "familiar" words compiled by Edgar Dale (revised in 1983). Familiar words are known by at least 80% of children in the fourth grade. Consideration of the number of familiar and unfamiliar words in a sample of text increases the accuracy of the reading level assessment. The grade level was determined using the following formula:

$$\text{Grade}=(0.1579\times\text{PERCENT UFMWDS})+(0.0496\times\text{WDS/SEN})+3.6365$$

The Dale-Chall procedure produced a Grade 3 reading level for the CDI, suggesting that the often cited Grade 1 reading level for the CDI is not definitive. Administrators and practitioners should not assume that all younger children will be able to understand the language on the inventory. For 7- and 8-year-olds and children with reading difficulties, it is recommended (Kovacs, 1992) that the administrator read aloud the instructions and the CDI items while the child reads along on his or her own form.

Administration Methods

One way to administer the CDI is to allow children to indicate their responses on a special Quikscore form (Kovacs, 1992). The Quikscore form is self-contained and includes all materials needed to score and profile the CDI. Conversion to *T*-scores is automatically made in the Quikscore form. The CDI also can be computer administered and scored using an IBM-compatible microcomputer (Kovacs, 1995).

Regardless of which option or format is chosen, the administrator should make sure that the child carefully reads the instructions and fully understands the inventory. As already noted, for younger children or those with reading difficulties, it may be necessary

to read the instructions and the items aloud while the child reads along on his or her own form or the computer screen, After reading each item, the child selects one of the three response options provided. A child may say that none of the choices in a given item really applies to him or her. In such a case, the child should be instructed to select the item choice that fits him or her *best*.

Although the GDI is most often administered on an individual basis, group administration is permitted (e.g., Friedman & Butler, 1979; Saylor, Finch, Baskin, Saylor, et al., 1984). Additionally, with nonclinical populations, some test administrators have considered inclusion of the suicide item to be inappropriate; in such instances, it may be preferable to use the GDI Short Form, which does not include this item.

APPLICABLE POPULATIONS

In interpreting clinically significant patterns of total scale and factor scores on the GDI, it is important to consider the background of the child, including his or her socioeconomic status, country of origin, and ethnicity. The norms presented in the main manual for the GDI (Kovacs, 1992) are based on a select sample of North American children. The validity of the instrument for other groups of children is suggested by research studies with different populations. In general, this body of research, cited in Tables 1.7 and 1.8, shows very widespread applicability of the GDI.

Table 1.7 lists research citations in connection with the use of the GDI with children from different cultures and from different countries. The GDI research includes data on children who were African American, Mexican American, North American, Irish, Italian, Spanish, Chinese (from Hong Kong), Dutch, German, American Indian, Australian, Egyptian, Japanese, Brazilian, Icelandic, Croatian, and French. These references should be consulted to aid in the interpretation of GDI results regarding those populations. Tables 1.1 and 1.7 cite some of the translated versions of the GDI that have been developed or used in research.

Table 1.8 lists some of the research on the GDI with children in special circumstances. Data have been obtained from samples of children from families of low socioeconomic status, urban and rural children, children in public housing situations, and children with mental retardation or learning/intellectual disabilities. A large amount of data was also collected from samples of children who have experienced emotional problems in some form. This would include children who have experienced trauma related to a familial suicide or cancer and children who have witnessed alcohol and substance abuse (e.g., marijuana use) or have been affected by it prenatally. More invasive experiences include sexual or physical abuse of boys and girls and war. The GDI has been also used with children going through the tribulations of parental divorce and children who have insulin-dependent diabetes mellitus.

APPROACHES TO CDI INTERPRETATION

The manner in which CDI results are used or interpreted is generally a function of the setting in which the instrument is administered and the ostensible reason for the administration. Consequently, the interpretative focus can be on the specific responses of a given child to each individual item on the total CDI *T*-score or individual CDI factor *T*-scores, each of which "rank" the child in comparison to "normal" age- and gender-matched peers.

TABLE 1.7
Research Reports on the Use of GDI with Children of Different Ethnic and National Backgrounds

Reference	*Notes*
Abdel-Khalek, 1993	$N = 2{,}558^a$, Arabic version
Abdel-Khalek, 1996	$N = 1{,}981$, Arabic version, Kuwaiti students
Arnarson, Smari, Einarsdottir, & Jonasdottir, 1994	$N = 436$, Icelandic version
Bahls, 2002	$N = 463$, Brazilian sample
Canals, Henneberg, Fernandez-Ballart, & Domenech, 1995	$N = 534$, Spanish sample
Chan, 1997	$N = 621$, Hong Kong
Chartier & Lassen, 1994	$N = 792^a$, North American sample
M. Donnelly, 1995	$N = 887$, Northern Ireland sample
DuRant, Getts, Cadenhead, Emans, & Woods, 1995	$N = 225$, African American sample
Dyer, 1995	$N = 33$, American Indian sample
Fitzpatrick, 1993	$N = 221$, African American sample
Frias, Mestre, del Barrio, & Garcia-Ros, 1992	$N = 1{,}286$, Spanish sample
Frigerio, Pesenti, Molteni, Snider, & Battaglia, 2001	$N = 284$, Italian sample
Ghareeb & Beshai, 1989	$N = 2{,}029^a$, Arabic version
Goldstein, Paul, & Sanfilippo-Cohn, 1985	$N = 85$, African American sample
Gouveia, Barbosa, de Almeida, & de Andrade-Gaiao, 1995	$N = 305$, Brazilian version
Houghton, O'Connell, & O'Flaherty, 1998	$N = 1090^a$, Irish sample
Koizumi, 1991	$N = 1{,}090^a$, Japanese version
Lobert, 1989, 1990	$N = 128$, German version
Mestre, Frias, & Garcia-Ros, 1992	$N = 952^a$, Spanish sample
Oy, 1991	$N = 432$, Turkish sample
Reicher & Rossman, 1991	$N = 658$, German version
Reinhard, Bowi, & Rulcovius, 1990	$N = 84$, German version
Rybolt, 1995	$N = 91$, Mexican American and Caucasian
Saint-Laurent, 1990	$N = 470$, French version
Sakurai, 1991	$N = 237$, Japanese version
Spence & Milne, 1987	$N = 386^a$, Australian sample
Steinsmeier-Pelster, Schurmann, & Duda, 1991	$N = 918$, German version
Steinsmeier-Pelster, Schurmann, & Urhahne, 1991	$N = 319$, German sample
Timbremont & Braet, 2001	$N = 663$, Dutch version
Worchel et al., 1990	$N = 135$, Hispanic sample
Yu & Li, 2000	$N = 1645^a$, Chinese sample
Zivcic, 1993	$N = 480$, Croatian version

[a] Sample sufficient to be considered normative data for this group.

Determining the Validity of the Results

Regardless of the interpretive focus, GDI results need to be examined in the context of potential threats to validity. One approach is to determine the *quality* of the completed inventory Another approach is to examine the inconsistency index.

Procedural Issues. The following issues should be kept in mind in assessing the quality of the completed GDI:

1. Has the inventory been filled in properly? Missing items will invalidate the total score. Although the administrator may prorate a missing item (e.g., by taking the average score on all remaining items and assigning that value to the missing item), subsequent interpretation must take any missing items into account.

TABLE 1.8
Research Reports on the Use of GDI with Special Groups

Reference	*Notes*
Benavidez & Matson, 1993	$N = 25$, mentally retarded children
Davis, 1996	$N = 120$, gifted children
T. F. Donnelly, 1995	$N = 61$, sexually abused children
Drucker & Greco-Vigorito, 2002	$N = 202$, children of substance abusers
DuRant, Getts, Cadenhead, Emans, & Woods, 1995	$N = 225$, public housing
Finkelstein, 1996	$N = 111$, learning disabled population
Gillick, 1997	$N = 20$, intrafamilial child abuse
Goldstein, Paul, & Sanfilippo, 1985	$N = 85$, learning disabled children
Gray, 1999	$N = 626$, prenatal substance exposure
Kovacs, Iyengar, Stewart, Obrosky, & Marsh, 1990	$N = 95$, diabetes mellitus
Lanktree, & Briere, 1995	$N = 105$, sexually abused children
Linna et al., 1999	$N = 6{,}000$, intellectual disability
Llabre & Hadi, 1997	$N = 151$, children assessed after war
Meins, 1993	$N = 798$, mentally retarded adults
Mestre, Frias, & Garcia-Ros, 1992	$N = 25$, mentally retarded children
Nelson, Politano, Finch, Wendel, & Mayhall, 1987	$N = 535$, emotionally disturbed children
Oy, 1991	$N = 432$, different socioeconomic status
Pfeffer, Karus, Siegel, & Jiang, 2000	$N = 80$, parental death from cancer/suicide
Polaino-Lorente & del-Pozo-Armentia, 1992	$N = 30$, familial cancer
Politano, Nelson, Evans, Sorenson, & Zeman, 1985	$N = 551$, emotionally disturbed children
Pons-Salvador & del Barrio, 1993	$N = 193$, parental divorce
Preiss, 1998	$N = 307$, children assessed after war
Rick, 1999	$N = 25$, sexually abused boys
Saylor, Finch, Spirito, & Bennett, 1984	$N = 154$, emotional-behavioral problems
Siegel, Karus, & Raveis, 1996	$N = 97$, familial cancer

2. Is there an apparent response bias? Response bias may be operating if a child *consistently* checks the first option on each item, the middle option, or the last option. Random checking of options, which may be inferred by the detection of apparently contradictory answers to similar items, may represent biased responding as well. Such patterns invalidate the GDI total score.

3. Are there any suggestions of lack of truthfulness? In a clinical setting that involves testing a child who has been referred, this possibility is indicated if the child "denies" every symptom or endorses the most severe option of every, or almost every, item. In such instances, inquiring into the child's expectations regarding the evaluation may be more informative than focusing on the GDI score itself.

4. Is the testing environment appropriate for psychological examination? As with all forms of psychological assessment, the GDI should be completed in a setting that is free from distraction, affords the child the requisite privacy, and is reasonably comfortable. An unsuitable testing environment is likely to threaten the validity of the child's responses and must be considered in score interpretation.

The Inconsistency Index, Children may exaggerate or misrepresent symptoms in some circumstances. As a result, some self-rated instruments include special items or scales to identify distorted responses (e.g., Beitchman, 1996; Reynolds & Richmond, 1985). Alternatively, for some instruments (e.g., MMPI-2 VRIN, and TRIN scales [Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989]; MASC Inconsistency Index [March, 1997]), an inconsistency index has been developed that does not usually

require special items. Inconsistency indexes are based on the premise that the most similar items, or the most highly correlated items, on a measure elicit similar (although not necessarily identical) responses. As determined by statistical procedures, if there is a large discrepancy in the responses for several correlated item pairs, then inconsistent and possibly invalid responding must be considered.

An inconsistency index exists for the CDL Each of the five scales on the GDI (i.e., Negative Mood, Interpersonal Problems, Ineffectiveness, Anhedonia, and Negative Self-Esteem) contains sets of items that are highly correlated with one another. If a pair of items is highly correlated, then a child whose response is indicative of a symptom for one item of the pair should give a response indicative of a symptom for the other item of the pair. Although such consistency is generally expected, some inconsistency can and will occur to a limited extent, the magnitude of which can be assessed through the GDI Inconsistency Index (Kovacs, 1995). This index is generated based on a computer algorithm taking into account the factor loadings of items. For the Negative Mood Scale, the highly correlated item set used to measure consistency comprises Items 1,8,10, and 11; for Interpresonal Problems, the set consists of Items 5, 26, and 27; for Ineffectiveness, the set consists of Items 15, 23, and 24; for Anhedonia, the set consists of Items 16,19,20, and 22; and for Negative Self Esteem, the set consists of Items 7, 9,14, and 25.

In the normative sample for the GDI, only 89 children out of 1,266 (6.9%) scored greater than or equal to 7 on the inconsistency index. And only 36 out of 1,266 (2.8%) scored greater than or equal to 9. Based on these data, the results from the inconsistency index are assessed as follows: If the index is less than 7, then the responses are considered sufficiently consistent. If the index is greater than or equal to 7 but less than 9, then the responses are considered somewhat inconsistent. If the index is greater than or equal to 9, then the responses are considered very inconsistent.

A high inconsistency index score should not be interpreted to mean that the GDI results should be disregarded. Inconsistent responding can occur for a variety of reasons, including an inability on the part of the child to concentrate on the task or understand the instructions. Such considerations must be part of interpreting the inconsistency index for a respondent.

Interpretive Steps

Interpretation of GDI results in the context of community-based or epidemiological studies are straightforward in so far as they usually employ clinically validated cutoff scores or normative *T*-scores to define "caseness." Thus, such cases will not be discussed in this chapter. Likewise, when the GDI is used as a screening instrument, a priori defined raw cutoff scores (or *T*-scores) are generally employed, with no need for specific interpretation. Because most questions regarding GDI score interpretation arise in the context of *clinical assessment* and for clinical purposes such as planning interventions or evaluations, pertinent information on these aspects of GDI use are now described in detail.

Interpretation of Total Scores and Factor Scores as T-Scores. Normative data tables are incorporated into the Profile Form for the CDL The normative data tables utilize *T*-scores, which are standardized to have a mean or average of 50 and a standard deviation of 10. The normative tables automatically compare the child being assessed to children in the normative sample of the same gender and age and allow each component in the profile to be compared to every other. *T*-scores above 65 are generally considered clinically significant when the child being studied is from a "high baserate" group, such

TABLE 1.9
Interpretive Guidelines for CDI T-Scores

T-Score	*Interpretation of Overall Symptoms/Complaints*[a]
Above 70	Very much above average
66 to 70	Much above average
61 to 65	Above average
56 to 60	Slightly above average
45 to 55	Average
40 to 44	Slightly below average
35 to 39	Below average
30 to 34	Much below average
Below 30	Very much below average

[a] Compared to children of similar age and gender in the normative sample.

as children in a clinical setting. When the child is believed to be from a"low base-rate" group, such as children without identified behavioral problems, a much higher cutoff, for example, a *T*-score of 70 or 75, should be used for inferring clinical problems. High scores suggest a problem and low scores indicate the absence of the problem.

It should be noted that the *T*-scores used with the CDI are linear *T*-scores. Linear *T*-scores do not transform the actual distributions of the variables, and hence, though each variable has been transformed to have a mean of 50 and a standard deviation of 10, the distributions of the scale scores do not change. Variables that are not normally distributed in the raw data will continue to be nonnormally distributed after the transformation.

As a rule of thumb, *T*-scores for the CDI can be interpreted using the guidelines in Table 1.9. These interpretations reflect how an individual child's score compares to those of children of the same age range and gender from the normative sample. Note, however, that the suggested adjectives are guidelines and that there is no reason to believe that a perceptible psychological difference is associated with the difference, for instance, between a *T*-score of 55 and a *T*-score of 56. *Therefore, these guidelines should not be used as absolute rules*.

For many clinical tests, it is common practice to interpret the overall profile based on the most elevated test scores. In such a case, a clinically elevated test score (in the metric of *T*-scores) would be defined as above 65. If, for a given set of scores, no test scores are above a *T*-score of 65, the profile is usually considered to be "normal." A profile in which a single *T*-score is elevated above 65 is usually considered to have a "one-point" code and is referred to by the single elevated scale. In general, given the high correlations of the factors of the CDI, such profiles should be relatively rare and, when encountered, may be viewed as only moderate evidence of a problem. When two or more subscale scores are clinically elevated, the profile is usually categorized by the two factors that are the highest and is called a "two-point code." Although two-point codes have not usually been employed with the CDI, some clinical practitioners may find it useful to use them. Experience with inventories such as the MMPI and the Personality Inventory for Children (PIC) indicates that two-point codes tend to be useful and robust ways of categorizing clinically meaningful patterns of behavior (Lachar & Gdowski, 1979).

In general, therefore, thoughtful examination of the GDI subscale profile should be more informative than consideration of only the total score. The GDI subscale *T*-score profile can be used to indicate specific areas of vulnerability as well as areas of strength. For example, from a clinical perspective, elevated *T*-scores on the Anhedonia factor or the Ineffectiveness factor may be particularly important. Because the Anhedonia factor contains items traditionally associated with "endogenous" depression, a child with a high *T*-score on this factor may be at particular risk for a serious depressive episode. A high score on the Ineffectiveness factor may indicate notable functional impairment, which may warrant additional interventions for a particular child. Concomitantly, in interpreting the GDI profile, a child who has elevated *T*-scores on both of these scales may be of greater clinical concern than a child who has an elevated score on the Anhedonia factor but an average score on the Ineffectiveness factor. In the former case, the child may be evidencing both functional impairment and troublesome depressive symptoms, whereas in the latter case, the troublesome depressive symptoms (area of vulnerability) are somewhat counteracted by child's having maintained reasonable functioning (area of strength).

Examination of the Total Raw Score and Item Response Pattern. A practitioner conducting a clinical assessment may decide to focus on the raw GDI score and individual item responses. For example, a total GDI score of 20 may result if a child endorses only 10 items but each to its most severe degree. Alternatively, a child may receive a score of 20 by endorsing up to 20 items but each to a mild degree. Examination of the number of items and the options for the items that contributed to the total GDI score can provide useful information about the extent and severity of the child's complaints and symptoms,

The examiner also may find it helpful to group the items endorsed by a child into phenomenologically meaningful categories. This approach can provide an additional perspective on the nature of the child's complaints. For example, if most or all endorsed GDI items pertain to physical and neurovegetative symptoms (somatic complaints, problems with sleep, appetite, and energy), a pediatric examination may be warranted. If all items with symptomatic responses relate to school or peer problems, a closer examination of those aspects of the child's life may be in order.

Examination of Individual Item Responses. By studying the individual responses of a child to the GDI items, the examiner may form hypotheses about the range and type of the child's difficulties. Furthermore, in conjunction with other information, item analysis can help to determine if the child is at particular risk for serious depression, even in the absence of a highly elevated total score.

For example, endorsements of the most severe options on Item 1 (sadness), Item 4 (anhedonia), and Item 10 (crying) are indicative of pervasive despondent mood. In so far as the presence of such a mood state has been shown to represent an early phase of depression, a child with such responses may warrant ongoing monitoring. Similarly, research evidence has suggested that children who are isolated may be at risk for a variety of adjustment problems. Thus, even if the total GDI score is low, a child who endorses both Item 20 (loneliness) and Item 22 (lack of friends) may be at risk for subsequent difficulties and could benefit from monitoring.

Unlike many other inventories, specific items on the GDI have not been designated as "critical," *All* of the items have been preselected by the author and validated by numerous investigators. *All* of the items are pertinent to the syndrome of depression in the juvenile years. However, the question pertaining to suicidal thoughts (Item 9) may

be particularly important for screening children in clinical settings or identifying those at risk. Endorsement of this item should prompt the examiner to conduct a detailed clinical assessment to determine the frequency and severity of suicidal ideation, whether it involves a specific contemplated method, and whether the child has ever attempted suicide. The information obtained should facilitate the planning of strategies for management or treatment.

Integrate the GDI Scores With All Other Information About the Child. The examiner should observe the child directly and the GDI results should be integrated with other test scores and with information about the child's background, family history, and school adjustment. Interviews with the child, parent, and perhaps teachers should be conducted. Consideration of such diverse information sources should result in a more valid conclusion regarding the child's problems and strengths and the extent to which depression may be undermining his or her functioning.

Determination of Appropriate Intervention Strategy for the Child. Based on all sources of information, the examiner should decide what kinds of feedback are appropriate and ethical for the parents and how to make that information available, how and when a report should be filed, and who should have access to the information. A treatment plan should be developed jointly with the parents or an appropriate referral should be made.

The results of the GDI can be particularly useful in determining suitable interventions for the child and in selecting treatment targets. As already noted, GDI factor scores and responses to items can identify problems or areas of concern. For example, a child with an elevated score on the Interpersonal Problems factor may benefit from social-skills training, modeling, or targeted group intervention as a way to treat his or her depression. A child with an elevated score on the Ineffectiveness factor may benefit from remedial help as well as behavior modification. A very high score on the Negative Mood factor may indicate consideration of referral for antidepressant pharmacotherapy If a child has a particularly high score on the Negative Self-Esteem factor, the intervention may focus on improving self-image and building confidence. In a similar vein, endorsement of items such as "I never have fun at school" and "I have to push myself all the time to do my schoolwork" would suggest that the treatment have a school-based component.

USE OF THE GDI FOR CLINICAL PURPOSES

Standards for Educational and Psychological Testing, developed through the collaboration of the American Psychological Association (1985) and the Association of Test Publishers, emphasizes the need to validate a measure with respect to each of its proposed purposes or uses. Therefore, in the following sections, validation information is integrated with descriptions of the main uses of the CDI.

Screening for Depression

The CDI is recommended as a screening tool and has been widely used for this purpose (e.g., Aronen & Soininen, 2000; Bahls, 2002; Canals, Henneberg, FernandezBallart, & Domenech, 1995; Congleton, 1996; Fristad, Weller, Weller, Teare, & Preskorn, 1988; Garvin, Leber, & Kalter, 1991; Jacobs, 1990; Kazdin, Colbus, & Rodgers, 1986; Krane, 1996; Lobovits & Handal 1985; Polaino-Lorente & Domenech, 1993; Rybolt, 1995;

TABLE 1.10
Research Showing Differences on the GDI Between Depressed and Nondepressed Children

Armsden, McCauley, Greenberg, Burke, & Mitchell, 1990
Carey, Faulstich, Gresham, Ruggiero, & Enyart, 1987
Craighead, Curry, & Ilardi, 1995
Fine, Moretti, Haley, & Marriage, 1985
Fristad, Weller, Weller, Teare, & Preskorn, 1988
Hodges, 1990
Hodges & Craighead, 1990
Jensen, Bloedau, Degroot, Ussery, & Davis, 1990
Kazdin, Esveldt-Dawson, Unis, & Rancurello, 1983
Kazdin, Rodgers, & Colbus, 1986
Knight, Hensley, & Waters, 1988
Kovacs, 1985
Lipovsky, Finch, & Belter, 1989
Liss, Phares, & Liljequist, 2001
Lobovits & Handal, 1985
Marriage, Fine, Moretti, & Haley, 1986
McCauley, Mitchell, Burke, & Moss, 1988
Moretti, Fine, Haley, & Marriage, 1985
Rotundo & Hensley, 1985
Saylor, Finch, Spirito, & Bennett, 1984
Spirito, Overholser, & Hart, 1991
Stark, Kaslow, & Laurent, 1993
Worchel, Nolan, & Willson, 1987

Stavrakaki, Williams, Walker, Roberts, & Kotsopoulos, 1991; Timbremont & Braet, 2001). As a screening tool, the GDI can serve to identify children who are "at risk" for a depressive disorder and may require further assessment with a more complex test battery (including behavioral observations, interviews, other psychological testing, etc.). The validity of the use of the GDI for this purpose largely depends on the ability of the inventory to differentiate children identified with depressive disorders from those who have not been identified with a depressive disorder. Many research studies have shown that the GDI effectively differentiates between depressed and nondepressed children. Some of this supporting literature is listed in Table 1.10.

The validity of the GDI as a screening tool also has been examined in terms of sensitivity and specificity. *Sensitivity* refers to the percentage of diagnosable depressed children who are correctly classified by the test, *specificity* to the percentage of nondepressed children who are correctly classified. For example, Craighead, Curry, and Ilardi (1995) reported that the five GDI factor scores classified participants as depressed versus not depressed with a high degree of accuracy. Using the GDI total score cutoff of 17 as the classification criterion, these investigators also found a sensitivity of 80% and a specificity of 84%.

When the GDI is used for screening purposes, a specific cutoff is usually selected, and children scoring above the cutoff are identified as those at risk. Different cutoff values may be used depending on the relative importance of sensitivity and specificity in a particular screening situation (Kovacs, 1992). In general, raising the cutoff value decreases sensitivity while it increases specificity, Lowering the cutoff value has the opposite effect: It increases sensitivity and decreases specificity.

High cutoff scores are more appropriate than low ones when it is important to minimize false positives, that is, nondepressed children falsely identified as at risk for depression. As noted, however, with high cut-off scores, the false-negative rate is increased; that is, many individuals who fall below the cutoff but are actually depressed will not be identified as at risk. Low cutoff scores are preferred when it is important to minimize false negatives, that is, depressed children wrongly identified as not at risk. However, the use of a low cutoff score will result in a higher false-positive rate; that is, more nondepressed individuals will be identified as at risk.

When the GDI is used as a general population-based screen, Kovacs (1992) recommended the raw score of 20 as a cutoff.[1] An example of a situation where the GDI can be used as a general screen with this cutoff score is in a school system in which routine testing is conducted on a large segment of the student population. On the other hand, for screening in clinical settings, a lower cutoff is appropriate because the base rate of depression can be expected to be higher. In the research literature (e.g., Garvin et al., 1991; Kazdin et al., 1986; Lobovits & Handal, 1985), cutoff scores as low as 12 or 13 have been proposed for clinical contexts.

Use as an Aid in the Diagnostic Process

Although the GDI can serve as an aid in the diagnostic process, it cannot by itself yield a diagnosis. As already noted, a psychiatric diagnosis of major depression or dysthymia requires that certain inclusion and exclusion diagnostic criteria be met, that the constellation of symptoms and signs be present for a particular duration, and that they should be associated with distress or functional impairment (American Psychiatric Association, 1980, 1985, 1994). The necessary information can only be obtained through a detailed clinical diagnostic interview. Regrettably, current usage of the GDI has not been satisfactory in this regard. An assessment by Fristad et al. (1997) found that 44% of the studies that used the GDI alone referred to high GDI scorers as "depressed" without providing a clear cautionary statement.

After a referred child has been administered the GDI, the results can be used in various ways to facilitate the process of diagnosis. If the clinical interview has confirmed the presence of a depressive disorder, the child's GDI score can serve as an indicator of the overall *severity* of his or her current symptoms. For example, a youngster whose GDI score is 28 is clearly more severely depressed than a comparably aged child whose GDI score is 16.

The GDI results also can be useful in reaching a diagnosis in cases where, subsequent to having interviewed the parents about the child, the clinician is unable to conduct a full face-to-face clinical assessment of the referred child. In such a case, information from the GDI may clarify aspects of the data provided by the parents because the test items and the DSM criteria for depression overlap. Ponterotto, Pace, and Kavan (1989), who reviewed the most commonly used depression measures, noted that the GDI was the only measure that had items pertaining to each of the *DSM-III-R* symptom criteria for major

[1] Matthey and Petrovski (2002) rightly critique the text and tables presented in the GDI manual, for these do poorly explicate the measure's value as a screening tool. Further to their credit, these authors (as is done here) identify myriad articles in support of the GDI as a screening tool. Inexplicably, however, Matthey and Petovski then ignore all of this research in their conclusions, which, as they state, are made "on the basis of the data reported in the manual" (p. 148). The conclusions presented here and in the GDI manual reflect a more appropriate appraisal based on all of the existing literature on the GDI.

TABLE 1.11
Correspondence of CDI Items to DSM-IV Symptom Criteria for Major Depression

DSM–IV *Criterion*	*Related CDI Item and the Most Symptomatic Response*
1. Depressed mood	Item 1: "I am sad all the time." Item 2: "Nothing will ever work out for me." Item 10: "I feel like crying every day." Item 20: "I feel alone all of the time."
2. Markedly diminished interest or pleasure	Item 4: "Nothing is fun at all."
3. Significant weight loss or decreased appetite nearly every day	Item 18: "Most days I do not feel like eating."
4. Insomnia or hypersomnia	Item 16: "I have trouble sleeping every night."
5. Psychomotor agitation or retardation	Item 15: "I have to push myself all the time to do my schoolwork."
6. Fatigue or loss of energy nearly every day	Item 17: "I am tired all the time."
7. Feelings of worthlessness or excessive guilt nearly every day	Item 3: "I do everything wrong." Item 7: "I hate myself." Item 8: "All bad things are my fault." Item 25: "Nobody really loves me."
8. Diminished ability to think or concentrate or indecisiveness	Item 13: "I cannot make up my mind about things."
9. Recurrent thoughts of death, suicidal ideation, or suicide attempt	Item 9: "I want to kill myself."

depression. The criteria for major depression essentially have remained the same in the *DSM-III, DSM-III-R,* and *DSM-IV*. Table 1.11 shows the correspondence between the nine criterion symptoms and specific CDI items.

Alternatively, the child's responses on the CDI can be used as starting points for probes in the clinical interview. The evaluator may note which particular CDI items were endorsed, then, citing to the child his or her item responses, the evaluator can ask the child during the interview to provide further information or elaborate on those responses.

USE OF THE CDI FOR TREATMENT MONITORING AND OUTCOMES ASSESSMENT

Because the CDI yields a quantified rating, the instrument is appropriate for monitoring levels of depressive symptoms during and at the end of treatment. For example, the CDI has been used to assess the effects of group therapy (e.g., Congleton, 1996; Garvin et al., 1991; Simmer-Dvonch, 1999), social training (e.g., Milne & Spence, 1987), pharmacotherapy (e.g., Preskorn, Weller, Hughes, Weller, & Bolte, 1987), cognitivebehavioral family therapy (e.g., Asarnow, Scott, & Mintz, 2002), and preventive intervention (e.g., Garvin et al., 1991). The application of the CDI in clinical practice or treatment monitoring entails several issues or considerations; these are described in the following sections.

Evaluation Against NIMH Criteria

The National Institute of Mental Health (NIMH) has specified 11 criteria for evaluating outcome measures (Ciarlo, Brown, Edwards, Kiresuk, & Newman, 1986; Newman, Ciarlo, & Carpenter, 1999). The GDI rates favorably with respect to these criteria, each of which is indicated below by means of italics.

The CDI has been highly useful with various populations and in different settings. As described earlier, it has been validated with the key *target groups* of nonreferred children as well as clinically depressed children. It has also been used with various other populations. As emphasized throughout this chapter, proper use of the CDI involves its integration with information from *multiple informants* and sources in order to make diagnostic and treament decisions. An amendment to the CDI, currently in progress, includes the development of parallel forms that can be completed by parents and teachers. Preliminary versions of the GDI-Parent version (CDI-P: Kovacs, 1997a) and CDI-Teacher version (CDI-T: Kovacs, 1997b) are being pilot-tested and standardized. The complementary Emotional Regulation Scales (Kovacs, in press) are also being developed and directly link to treatment planning.

It has been demonstrated that the CDI has a high degree of *utility in the area of clinical services and is compatible with a variety of clinical theories and practices*. Its results can easily be translated so as to be appropriate and useful in clinical treatment strategies. The CDI also can be used to evaluate the effectiveness of such treatment strategies. The CDI adheres to the NIMH criterion that an outcome measure be useful in *identifying relevant changes in the client during the process of treatment*—changes that can act as "behavioral markers of progress or risk level" (Newman, Ciarlo, & Carpenter, 1999, p. 160). Several strategies for assessing the significance of changes in CDI scores during treatment are described in this chapter.

The *psychometric strengths* of the CDI are well established and documented by an abundance of research publications. *'Normative data,* described in the CDI manual (Kovacs, 1992), provide clinicians with benchmarks that act as *objective referents* to be used in interpreting test results. The norms in the manual are based on a North American sample, but data from many other countries are also available. Furthermore, in accordance with stipulations of the American Psychological Association and the Association of Test Publishers, the CDI has been validated for each of its proposed uses.

From a pragmatic perspective, the CDI is *simple and easy to use;* manuals and materials are available to facilitate proper administration, scoring, and interpretation. In addition, the CDI is extremely *cost-efficient,* and its results are both easy to relay and *readily comprehensible* by nonprofessional audiences,

For all of the above reasons, the CDI is well deserving of the worldwide attention it has received as both a research and a clinical tool in a wide range of contexts. And its adherence to NIMH standards for assessment instruments also supports its suitability for monitoring treatment and assessing outcomes.

Establish Baseline Severity of Symptoms

If feasible and appropriate, the CDI should be administered twice at baseline. The resultant two scores can be averaged to yield an index of initial symptom severity. This procedure, also known as multiple baseline assessment, has been recommended by Milich, Roberts, Loney, and Caputo (1980), Conners (1997), and Nelson and Politano (1990), particularly for studies designed to evaluate treatment outcomes. Repeated administration of a scale can produce a declines in scores influenced by methodological artifacts such as statistical

regression to the mean, placebo response to the initial assessment, or spontaneous improvement (Finch et al., 1987; Kaslow et al., 1984; Meyer et al., 1989). Therefore, a multiple baseline (rather than a single baseline) assessment is usually considered to yield a more valid index of symptom severity at the beginning of treatment.

Determine a Treatment Goal

The goals of treatment can include an a priori defined decrement in overall symptom severity, the absence of depressive symptoms, and improvements in specific areas of the child's functioning. Changes in the total CDI score can be interpreted as reflecting changes in the severity of the child's depressive symptoms. If CDI item responses scored "2" are initially selected as treatment targets, the clinician's goal may include the lessening or elimination of these particular complaints. Additionally, change (or lack of change) in factor scores may help pinpoint areas of functioning in which therapy has had the most (or least) impact.

Determine Frequency of CDI Administration During Treatment

Practical considerations are likely to affect how often the CDI can be readministered during treatment. Such considerations may include the time interval between sessions with the child as well as the burden of other assessments to which the child may be subjected. In general, a 2-week test-retest interval may be most appropriate (Kovacs, 1992), and the time required for any given test battery (including the CDI) should not exceed 20 minutes or so, particularly with younger patients. If possible, the instrument should be administered at about the same time of day each time and in the same location in order to control extraneous variables that might impact the responses.

Assess the Statistical/Clinical Significance of Changes in CDI Scores

CDI scores for the same respondent are likely to vary with repeated administration owing to random fluctuation in responses. Therefore, it is important to define the magnitude of change in CDI scores that is to be considered significant.

On a purely descriptive level, significant improvement can be defined in terms of a desired change in responses to selected CDI items. For example, if one treatment target is to improve the child's sleep, then a change on Item 16 from "I have trouble sleeping every night" to "I have trouble sleeping many nights" or "I sleep pretty well" maybe considered clinically meaningful. As Conners (1994) noted:

> Clinically…it is always useful in assessing change to…circle three to five items that… are the most crucial problem areas. Then, regardless of changes in factor scores, it is possible to examine particular target symptoms or behaviors for evidence of a treatment effect. Obviously, one must be mindful of the possibility of interpreting random fluctuations as real change, but this is precisely the reason for not relying on a single outcome measure, (p. 569)

From a clinical perspective, *T*-score changes of five or more points on the CDI subscales also may be considered to be indicative of significant change (e.g., Conners, 1994). This approach, which is suggested as a rough guideline, has the advantage of ease of application, and it is useful in most instances.

Other methods, including the procedure described in Jacobsen and Truax (1991), address "significant change" with reference to statistical criteria (for a review, see Speer

& Greenbaum, 1995). The Jacobson-Truax method involves obtaining the difference between the baseline raw score and the raw score obtained during or after treatment, which is then divided by the standard error of the differences. This formula utilizes an appropriate reliability value for the test instrument; this value can be a test-retest, Cronbach's alpha, or split-half reliability value.

A repeated measures *t*-test represents an alternative statistical method of estimating significant change in scores. The responses from the baseline GDI administration are paired with the responses from the administration during or after treatment. The repeated measures *t*-test procedure is produced automatically by the GDI software program (Kovacs, 1995), and thus information regarding the significance of change in GDI scores is readily accessible.

Decide on the Effects of Treatment

In general, downward trends in GDI scores are likely to indicate that treatment is progressing in a proper direction. If GDI scores rise or fluctuate unpredictably from one administration to the next, a full clinical reassessment is warranted to verify the child's psychiatric status and reevaluate the appropriateness of the intervention. Treatment studies of adults have shown that most of the improvement in symptom status occurs by the eighth treatment session (Howard, Kopta, Krause, & Orlinsky, 1986). Thus, after one or two months of treatment, there should be an observable reduction in the child's depressive symptoms, although full remission would not yet be evident.

Decisions about the effects of treatment with a depressed child should not depend solely on the GDI. For example, one research study found a tendency among children to deny symptoms and to respond defensively (Joiner, Schmidt, & Schmidt, 1996). Such findings reinforce the need to corroborate self-report information prior to making decisions about the effects of treatment.

A HYPOTHETICAL CASE STUDY

A hypothetical case study is now provided (using elements of actual clinical cases) to illustrate some of the aforementioned principles in the use of the GDI. This case study includes screening, treatment planning, treatment monitoring, and outcomes assessment components.

Tamara is a 9-year-old girl who has been living with her mother. Tamara's mother had contacted the clinic because of concern regarding her daughter's behavior. The mother described Tamara as being overly sensitive and emotionally labile and prone to extreme emotional outbursts. During some of these outbursts, Tamara screamed, cried, and voiced concern that her mother would leave her. The GDI was first administered to Tamara after the initial contacts with the mother. The first administration yielded a GDI total raw score of 34, which is well above established cutoff points for identifying children who are at risk.

In the 3-year period before the initial assessment, Tamara experienced several major negative life events, including a fire in the family home that resulted in the death of Tamara's older brother and the destruction of all of the family's personal belongings and the subsequent disappearance of her natural father.

A psychiatric interview with the mother revealed symptoms for Tamara that dated back to the disappearance of her natural father. At the time of his disappearance, Tamara had developed considerable sadness, crying, negative self-esteem, and guilt. She also

had difficulty sleeping. After the fire, she additionally developed nightmares. Tamara started to experience occasional thoughts of wanting to die as well as difficulty with concentration, The latter symptom was verified by her school records and declining school grades.

In a psychiatric interview with Tamara, it became clear that she was aware of what was upsetting her and talked about her fear of being apart from her mother She spoke of her long-standing sadness, difficulty with concentration, difficulty in sleeping, and feeling like a burden to others. She also believed that nothing would change in her life. She admitted to not wanting to go to school because of how the other children were treating her.

Based on the information obtained during these detailed psychiatric interviews, it was determined that Tamara met psychiatric diagnostic criteria for dysthymic disorder (American Psychiatric Association, 1994). She also had a diagnosable anxiety disorder. By examining her GDI factor scores, it became apparent that negative affect, ineffectiveness, and anhedonia were more problematic for her than behavior problems or low self-esteem. The Ineffectiveness score was relatively elevated and was consistent with her recent school problems.

The combining of information from the GDI with the developmental history and clinical information resulted in the development of an intervention plan. Before treatment began, a second administration of the GDI was conducted in order to strengthen the accuracy of the baseline and corroborate other clinical observations. Recommendations for individual and concomitant parent-child therapy sessions were made, and treatment began approximately 1 month after the initial evaluation.

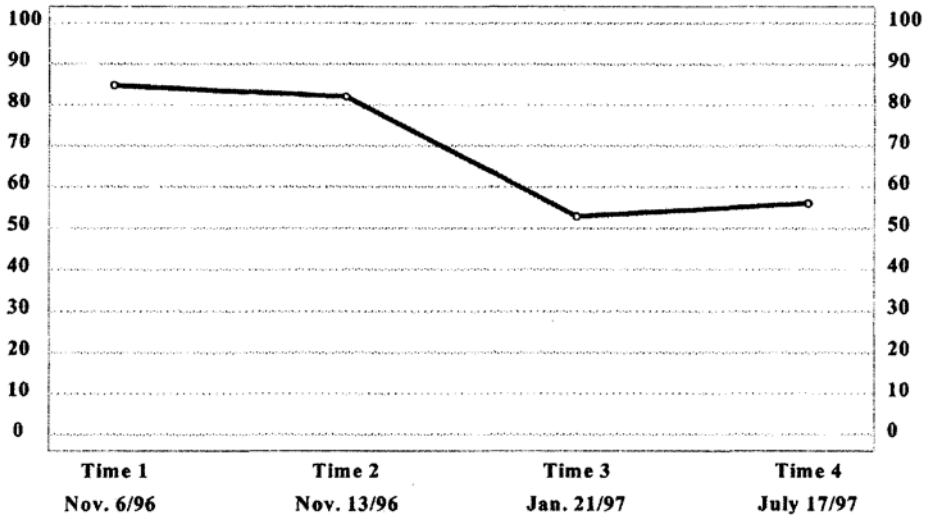
Over the next few months of the intervention program, important improvements were noted. A third administration of the GDI was done, and it appeared that the symptoms had been reduced to an acceptable level. On the third administration, Tamara's total GDI raw score had dropped to 11. The GDI software program was used to generate a comparison between the posttreatment administration and the baseline scores, and the large change was determined to be statistically significant. A full clinical evaluation at that point suggested that Tamara had recovered from her depression and anxiety. Periodic follow-up checks were done to make sure that Tamara had maintained the gains from the therapeutic intervention. Six months after discontinuing intervention, a follow-up (fourth) administration of the GDI was given, and although the scores had increased slightly compared with the third administration, Tamara continued to show reasonably benign levels of depressive symptoms.

Figure 1.1 shows portions of the report produced by the GDI software, which includes a graph of the four GDI administrations and a statistical assessment of the magnitude of the change that occurred over administrations. There was no significant difference between the two baseline administrations, but after treatment Tamara's scores were significantly lower than both of the baseline results. These findings strongly suggest that the treatment was effective in dealing with Tamara's depression.

NEW DEVELOPMENTS

Parent and Teacher Versions of the CDI

Youth self-report provides a valuable means of gathering information about depressive symptoms, Ideally, however, assessments from apporpriate observers should supplement



Time 1 (Nov. 6/96) vs. Time 2 (Nov. 13/96)
Total CDI score at Time 1 = 85, total CDI score at Time 2 = 82.
There was a drop in the CDI total score. Although this drop may reflect improvement between the two administrations, the statistics shown below indicate that the change was small and may reflect random fluctuation as opposed to significant change.
Statistical analysis: $t = 1.00$, $df = 26$, not statistically significant.

Time 2 (Nov. 13/96) vs. Time 3 (Jan. 21/97)
Total CDI score at Time 2 = 82, total CDI score at Time 3 = 53.
There was a substantial decline in the CDI total score, indicating improvement between the two administrations. The statistics indicate that this improvement was statistically significant.
Statistical analysis: $t = 5.05$, $df = 26$, $p < .05$.

Time 3 (Jan. 21/97) vs. Time 4 (July 17/97)
Total CDI score at Time 3 = 53, total CDI score at Time 4 = 56.
There was an increase in the CDI total score. Although this change may reflect a worsening condition between the two administrations, the statistics indicate that the change was small and may reflect random fluctuation as opposed to significant change.
Statistical analysis: $t = -1.44$, $df = 26$, not statistically significant.

Raw Scores

	Time 1 11/6/96	Time 2 11/13/96	Time 3 1/21/97	Time 4 7/17/97
Q#1	2	2	0	0
Q#2	1	1	1	1
Q#3	1	1	1	1
Q#4	2	2	1	1
Q#5	0	0	0	0
Q#6	2	1	0	1
Q#7	0	0	0	0
Q#8	1	1	0	1
Q#9	2	2	0	0
Q#10	1	2	0	0
Q#11	2	2	1	1
Q#12	1	1	0	0
Q#13	1	1	1	1
Q#14	1	1	1	1
Q#15	2	2	1	1
Q#16	2	2	0	0
Q#17	2	2	0	0
Q#18	1	0	0	0
Q#19	0	0	0	0
Q#20	2	1	1	1
Q#21	2	2	1	1
Q#22	1	1	1	1
Q#23	2	2	0	0
Q#24	1	1	1	1
Q#25	1	1	0	0
Q#26	1	1	0	0
Q#27	0	0	0	0

the self-assessment. Specifically, parent and teacher versions of the GDI would be of great value. The *DSM-IV* emphasizes the importance of "multirater" assessments, and other measures have effectively created child, parent, and teacher versions (e.g., Conners Rating Scales Revised; Conners, 1997).

Parent and teacher versions of the GDI have, in fact, appeared sporadically in the literature (e.g., Cole, Hoffman, Tram, & Maxwell, 2000; Cole, Martin, Peeke, Truglio, & Seroczynski, 1998; Fristad, Weller, Weller, Teare, & Preskorn, 1991; Hoffman, Cole, Martin, Tram, & Seroczynski, 2000; Slotkin, Forehand, Fauber, McCombs, & Long, 1988; Wierzbicki, 1987). Use of these versions has been problematic since they are idiosyncratic, lack proper norms, and often have insufficient reliability and validity. To correct this problem, standard parent and teacher versions of the GDI have been created by Kovacs (1997a, 1997b). The CDI-P consists of 17 items, and the CDI-T consists of 12 items, The items were selected to correspond to items on the self-report version but were rephrased for administration to parents and teachers. Only items that maximize validity when answered by parents and teachers as respondents were retained.

A significant amount of research has been conducted with these standardized versions of the parent and teacher forms (Kovacs, 1997a, 1997b), and some of the preliminary results are provided here. For the parent form, 467 (205 women and 262 men) completed forms have been compiled from nonclinical sites, with 167 (49 women and 118 men) clinical cases also collected. For the teacher form, 583 (266 women and 317 men) completed sets of responses were compiled from nonclinical sites, and 114 (32 women and 82 men) clinical cases were obtained. The ethnic breakdown of the samples was approximately 80% white, 7% Hispanic, 4% Asian, 6% black, and 3% other.

In terms of reliability, total scores for both the parent and teacher forms were evaluated using Cronbach's alpha statistic. For the parent form, the overall alpha was .90, with alphas of .90 and .87 for the nonclinical and clinical groups, respectively. For the teacher form, the alpha was .89 for the overall combined sample as well as for the nonclinical and clinical samples treated separately, The values obtained suggest excellent internal reliability for the total scores for the parent and teacher forms of the GDI.

In another set of analyses, ANCOVAs were conducted to see if the CDI-P and CDI-T could differentiate between nonclinical and clinical cases. Gender and age (covariate) were controlled in the analysis. The CDI-P total score significantly differentiated nonclinical from clinical cases ($F_{1,529}$=31.6, p<.001), and the CDI-T was also successful in this regard ($F_{1,692}$=44.2, p<.001). These analyses provide evidence of the validity of the teacher and parent versions of the GDI and show that they successfully discriminate between nonclinical and clinical cases.

Finally, further analyses were done comparing the parent, teacher, and self-report versions. The CDI-P and CDI-T correlated at r=.55 (n=193, p<.001), the CDI-P and GDI-self correlated at r=.45 (n=188, p<.001), and the CDI-T and GDI-self correlated at r=.52 (n=140, p<.001). This range of correlations suggests comparability among the measures and some overlap in the observers. At the same time, the correlations indicate sufficient variation among parents, teachers, and youths to highlight the importance of capturing the ratings of all three sources. By examining and comparing the information provided from the three informants, clinicians can explore discrepancies for more accurate assessments.[2] For example, if the parent and teacher disagree, then the clinician should

[2] Preliminary data comparing mothers' and fathers' ratings on the CDI-P indicate that fathers reported more depressive symptoms than mothers (Total: M_F=13.89, M_m=12.32, p<.05; Emotional Problems: M_F=4.94, M_m=4.51, not significant; Behavioral Problems: M_F=4.10, M_m=3.62, p<.10).

explore both perspectives to resolve the difference. If the child and teacher indicate that there is depressed mood but the parent does not, he or she might be denying the problem or underestimating its importance. The clinician may have to work through the parent's mindset to facilitate the appropriate intervention. If, on the other hand, the results from different informants are comparable, showing that everyone agrees on the assessment, the clinician will likely have greater confidence in his or her conclusions and actions, and the intervention could become easier to carry out.

Emotional Regulation Scales (ERS)

The GDI can play a valuable role in identifying depressive symptoms and offer insights into the nature of the symptoms via its subscales and items. The ERS scales (Kovacs, in press) are linked to the GDI but generate clinical information about the strategies individuals use to contend with emotions and emotional situations. Thus, they provide a mechanism that relates directly to clinical understanding and treatment of depressed pateints.

The scales were specifically designed to assess frequency of utilization (rated on a 3-point scale: "not true of me," "sometimes true of me," "many times true of me") of various emotion-regulatory strategies in response to situations that evoke sadness, fear, anger, or happiness. The items sample strategies from four emotion-regulatory domains; physical/biologic, behavioral, cognitive, and social-interpersonal. The items are classified into four sets: behavioral (25 items), cognitive (21 items), socialinterpersonal (15 items), and physical (5 items), each set with both positive and negative items. Two additional items that reflect overall competence at regulating emotion were classified as "not domain specific." For each of the four sets of items, two scores are computed. "Frequency" scores for each domain reflect the frequency with which strategies in the given domain are used regardless of whether they are positive/adaptive or negative/maladaptive. "Skill" scores reflect the skill with which the individual uses strategies in the given domain, that is, the degree to which the individual uses positive/adaptive strategies and avoids negative/maladaptive strategies. A Frequency subscale indicates how typically the respondent uses the given strategy (regardless of whether it is adaptive or not). Skill subscale items are scored in the direction of increasing adaptive strategy. There are three versions of the ERS: for youth self-report, for parent report, and for adult self-report.

CONCLUSION

Given the high prevalence of depressive disorders in children and adolescents and their likely disruption of functioning in a number of areas, the development of assessment tools designed for this population is of utmost important. The Children's Depression Inventory (GDI) was developed to address this need, and it has since become one of the most widely used and cited inventories of depression. This chapter described the various versions of the GDI and current research and theory related to the GDI. It examined the current use of this instrument, distinguished proper use from improper use, and presented answers to questions frequently asked by practitioners. It also addressed the research history, administration, psychometric properties, and interpretation of the GDI.

The GDI, which is appropriate for children and adolescents aged 7 to 17, quantifies a range of depressive symptoms, including disturbed mood, problems in hedonic capacity and vegetative functions, low self-esteem, hopelessness, and difficulties in interpersonal behaviors. It is useful for the early identification of symptoms and for monitoring

treatment effectiveness. It can also play a role in the diagnostic process as part of a larger assessment battery, Psychometric strengths of the GDI have been well established. Reliability, examined in terms of internal consistency, test-retest reliability, and standard error, has been found to range from satisfactory to excellent. The GDI has been used in many clinical studies and experimental research studies and has proved capable of assessing important constructs that have strong explanatory and predictive utility in characterizing depressive symptoms in children and adolescents.

The GDI has also been found to be useful with various populations and in different settings. Amendments to the GDI are currently being developed; these include parallel versions to be completed by teachers and parents. The complementary Emotional Regulation Scales, which are currently being developed, are directly linked to treatment planning. For all of the above reasons, as well as for its simplicity and ease of use and its adherence to NIMH standards for assessement instruments, the GDI is well deserving of its worldwide use in research and clinical settings.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Maria Kovacs, Ph.D., Lila Elkhadem, B.A., Jagruti Parmar, M.A., Joanne Morrison, B.A., and Karen Hirscheimer, M.A. for their contributions to this chapter.

REFERENCES

Abdel-Khalek, A.M. (1993). The construction and validation of the Arabic Children's Depression Inventory. *European Journal of Psychological Assessment, 9,* 41–50.

Abdel-Khalek, A.M. (1996). Factorial structure of the Arabic Children's Depression Inventory among Kuwaiti subjects. *Psychological Reports, 78,* 963–967.

Albert, N., & Beck, A.T. (1975), Incidence of depression in early adolescence: A preliminary study. *Journal of Youth and Adolescence, 4,* 301–307.

Allen, D.M., & Tarnowski, K.J. (1989). Depressive characteristics of physically abused children. *Journal of Abnormal Child Psychology, 17,* 1–11.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1985). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychological Association. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

Armsden, G. C.., McCauley, E., Greenberg, M.T., Burke, P.M., & Mitchell, J.R. (1990). Parent and peer attachment in early adolescent depression. *Journal of Abnormal Child Psychology, 18,* 683–697.

Arnarson, E.O., Smari, J., Einarsdottir, H,, & lonasdottir, E. (1994). The prevalence of depressive symptoms in pre-adolescent school children in Iceland. *Scandinavian Journal of Behaviour Therapy, 23,* 121–130.

Aronen, E.T., & Soininen, M. (2000). Childhood depressive symptoms predict psychiatric problems in young adults. *Canadian Journal of Psychiatry, 45,* 465–470.

Asarnow, J.R., & Carlson, G.A. (1985). Depression Self-Rating Scale: Utility with child psychiatric inpatients. *Journal of Consulting and Clinical Psychology, 53,* 491–499.

Asarnow, J.R., Scott, C.V., & Mintz, J. (2002). A combined cognitive-behavioral family education intervention for depression in children: A treatment development study. *Cognitive Therapy and Research, 26,* 221–229.

Bahls, S. (2002). Epidemiology of depressive symptoms in adolescents of a public school in Curitiba, Brazil. *Revista Brasileira de Psiquiatria, 24,* 63–67.

Barreto, S.J. (1994). Understanding the Children's Depression Inventory (GDI): A critical review. *Child Assessment News, 3,* 3–5.

Bartell, N.P., & Reynolds, W.M. (1986). Depression and self-esteem in academically gifted and nongifted children: A comparison study. *Journal of School Psychology, 24*, 55–61.

Beck, A.T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York, NY: Harper & Row.

Beitchman, J.H. (1996). *Feelings, Attitudes, and Behaviors Scale for Children (FAB-C)*. Toronto: Multi-Health Systems Inc.

Benavidez, D.A., & Matson, J.L. (1993), Assessment of depression in mentally retarded adolescents. *Research in Developmental Disabilities, 14*, 179–188.

Berndt, D.J., Schwartz, S., & Kaiser, C.F. (1983). Readability of self-report depression inventories. *Journal of Consulting and Clinical Psychology, 51*, 627–628.

Blumberg, S.H., & Izard, C.E. (1986). Discriminating patterns of emotions in 10- and 11-year-old children's anxiety and depression. *Journal of Personality and Social Psychology, 51,* 852–857.

Bodiford, C.A., Eisenstadt, T.H., Johnson, J.H., & Bradlyn, A.S. (1988). Comparison of learned helpless cognitions and behavior in children with high and low scores on the Children's Depression Inventory. *Journal of Clinical Child Psychology, 17,* 152–158.

Breen, M.P., & Weinberger, D.A. (1995). Regulation of depressive affect and interpersonal behavior among children requiring residential or day treatment. *Development and Psychopathology,* 7, 529–541.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A.M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Canals, J., Henneberg, C., Fernandez-Ballart, J., & Domenech, E. (1995). A longitudinal study of depression in an urban Spanish pubertal population. *European Child and Adolescent Psychiatry,* 4,102–111.

Carey, M.P., Faulstich, M.E., Gresham, F.M., Ruggiero, L., & Enyart, P. (1987). Children's Depression Inventory. Construct and discriminant validity across clinical and nonreferred (control) populations. *Journal of Consulting and Clinical Psychology, 55,* 755–761.

Chall, J.S., and Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

Chan, D.W. (1997). Depressive symptoms and perceived competence among Chinese secondary school students in Hong Kong. *Journal of Youth and Adolescence, 26, 303–319.*

Chartier, G.M., & Lassen, M.K. (1994). Adolescent depression: Children's Depression Inventory norms, suicidal ideation, and (weak) gender effects. *Adolescence, 29,* 859–864.

Ciarlo, J.A., Brown, T.R., Edwards, D.W., Kiresuk, T.J., & Newman, F.L. (1986). *Assessing mental health treatment outcome measurement techniques* (DHHS Pub. No. [ADM] 86–1301). Washington, DC: U.S. Government Printing Office.

Cole, D.A., Hoffman, K.B., Tram, J.M., & Maxwell, S.E. (2000). Structural differences in parent and child reports of children's symptoms of depression and anxiety. *Psychological Assessment,* 12, 174–185.

Cole, D.A., Martin, J.M., Peeke, J., Truglio, R., & Seroczynski, A.D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Counseling and Clinical Psychology, 66*, 451–460.

Congleton, A.B. (1996). The effect of a cognitive-behavioral group intervention on the locus of control, attributional style, and depressive symptoms of middle school students. *Dissertation Abstracts International. Section A: The Humanities and Social Sciences,* 56(9-A), 3507.

Conners, C.K. (1994). Conners' Rating Scales. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (pp. 550–578). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conners, C.K. (1997). *Conners' Rating Scales-Revised: Technical manual* Toronto: Multi-Health Systems Inc.

Craighead, W.E., Curry, J.E, & Ilardi, S.S. (1995). Relationship of Children's Depression Inventory factors to major depression among adolescents. *Psychological Assessment, 7, 171–176.*

Curry, J. E, & Craighead, W.E. (1990). Attributional style and self-reported depression among adolescent inpatients. *Child and Family Behavior Therapy, 12,* 89–93.

Dale, E., & Chall, J.S. (1948). *A formula for predicting readability.* Columbia, OH: Ohio State University Bureau of Educational Research. Research reprinted from *Educational Research Bulletin, 27, 11–20, 37–54.*

Davis, S. (1996). A study of depression and self-esteem in moderately gifted and nongifted children. *Dissertation Abstracts International Section A: The Humanities and Social Sciences,* 56(10-A), 3886.
DeVellis, R.F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.
Devine, D., Kempton, T., & Forehand, R. (1994). Adolescent depressed mood and young adult functioning: A longitudinal study. *Journal of Abnormal Child Psychology, 22,*629–640.
Donnelly, M. (1995). Depression among adolescents in Northern Ireland. *Adolescence, 30,* 339–350.
Donnelly, T.F. (1995). Effects of parental reaction on depression, anxiety, and self-esteem in sexually abused children. *Dissertation Abstracts International. Section B: The Sciences and Engineering,* 56(5-B), 2896.
Drucker, P.M., & Greco-Vigorito, C. (2002). An exploratory factor analysis of Children's Depression Inventory scores in young children of substance abusers. *Psychological Reports, 91,* 131–141.
DuBois, D.L., Felner, R.D., Bartels, C.L., & Silverman, M.M. (1995). Stability of self-reported depressive symptoms in a community sample of children and adolescents. *Journal of Clinical Child Psychology, 24,* 386–396.
DuRant, R.H., Getts, A., Cadenhead, C., Emans, S.J., & Woods, E.R. (1995). Exposure to violence and victimization and depression, hopelessness, and purpose in life among adolescents living in and around public housing. *Journal of Developmental and Behavioral Pediatrics, 16,* 233–237.
Dyer, L.C. (1995). Assessing depression in American Indian children. *Dissertation Abstracts International Section B: The Sciences and Engineering,* 55(11-B), 5064.
Eason, L.J., Finch, A.J., Jr., Brasted, W., & Saylor, C. E (1985). The assessment of depression and anxiety in hospitalized pediatric patients. *Child Psychiatry and Human Development, 16,* 57–64.
Elliott, D.J., & Tarnowski, K.J. (1990). Depressive characteristics of sexually abused children. *Child Psychiatry and Human Development, 21,* 37–48.
Fauber, R., Forehand, R., Long, N., Burke, M., & Faust, J. (1987). The relationship of young adolescent Children's Depression Inventory (CDI) scores to their social and cognitive functioning. *Journal of Psy-chopathology and Behavioral Assessment, 9,* 161–172.
Faulstich, M.E., Carey, M.P., Ruggiero, L., Enyart, P., & Gresham, F. (1986). Assessment of depression in childhood and adolescence: An evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). *American Journal of Psychiatry, 143,* 1024–1027.
Felner, R.D., Rowlison, R.T., Raley, P.A., & Evans, E. (1988). Depression in children and adolescents: A comparative analysis of the utility and construct validity of two assessment measures. *Journal of Consulting and Clinical Psychology, 56,* 769–772.
Finch, A.J., Saylor, C.F., & Edwards, G.L. (1985). Children's Depression Inventory: Sex and grade norms for normal children. *Journal of Consulting and Clinical Psychology, 53,* 424–425.
Finch, A.J., Saylor, C. F, Edwards, G.L., &, Mclntosh, J.A. (1987). Children's Depression Inventory: Reliability over repeated administrations. *Journal of Consulting and Clinical Psychology, 16,* 339–341.
Fine, S., Moretti, M., Haley, G., & Marriage, K. (1985). Affective disorders in children and adolescents: The dysthymic disorder dilemma. *Canadian Journal of Psychiatry, 30,* 173–177.
Finkelstein, R. (1996). Depression and loneliness in the early adolescent, learning-disabled population. *Dissertation Abstracts International. Section A: The Humanities and Social Sciences, 56(12-A),* 4703.
Fitzpatrick, K.M. (1993). Exposure to violence and presence of depression among low-income, AfricanAmerican youth. *Journal of Consulting and Clinical Psychology, 61,* 528–531.
Frias, D., Mestre, .V, del Barrio, V., & Garcia-Ros, R. (1992). Estructura familiar y depresion infantil [Family structure and childhood depression]. *Anuario-de-Psicologia, 52,* 121–131.
Friedman, R.J., & Butler, L.F. (1979). *Development and evaluation of a test battery to assess childhood depression.* Final report to Health and Welfare, Canada, for Project #606–1533–44. Ottawa: Health and Welfare Canada.
Frigerio, A., Pesenti. S., Molteni, M., Snider, J., & Battaglia, M. (2001). Depressive symptoms as measured by the CDI in a population of northern Italian children. *European Psychiatry, 16(1),* 33–37.
Fristad, M.A., Emery, B.L., & Beck, S.J. (1997). Use and abuse of the Children's Depression Inventory. *Journal of Consulting and Clinical Psychology, 65, 699–702.*
Fristad, M, A., Weller, E.B., Weller, R.A., Teare, M., & Preskorn, S.H. (1988). Self-report vs. biological markers in assessment of childhood depression. *Journal of Affective Disorders, 15,* 339–345.
Fristad, M.A., Weller, E.B., Weller, R.A., Teare, M., & Preskorn, S.H. (1991). Comparisons of the parent and child versions of the Children's Depression Inventory (CDI). *Annals of Clinical Psychiatry, 3,* 341–346.
Garvin, V., Leber, D., & Kalter, N. (1991). Children of divorce: Predictors of change following preventive intervention. *American Journal of Orthopsychiatry, 61,* 438–447.

Ghareeb, G.A., & Beshai, J.A. (1989). Arabic version of the Children's Depression Inventory: Reliability and validity. *Journal of Clinical Child Psychology, 18,* 323–326.

Gillick, T.A. (1997). Depression in adolescent female victims of intrafamilial child sexual abuse. *Dissertation Abstracts International. Section B: The Sciences and Engineering, 57(7-B),* 4706.

Gladstone, T R.G., & Kaslow, N.I. (1995). Depression and attributions in children and adolescents: A meta-analytic review. *Journal of Abnormal Child Psychology, 23,* 597–606.

Goldstein, D., Paul, G.G., & Sanfilippo-Cohn, S. (1985). Depression and achievement in subgroups of children with learning disabilities. *Journal of Applied Developmental Psychology, 6,* 263–275.

Gouveia, V.V., Barbosa, G.A., de Almeida, H.J.E, & de Andrade-Gaiao, A. (1995). Inventario de depressao infantil-CDI: Estudo de adptacao com escolares de Joao Pessoa [Children's Depression Inventory-GDI: Adaptation study with students of Joao Pessoal. *Jornal Brasileiro de Psiquiatria, 44,* 345–349.

Gray, K. (1999). Prenatal substance exposure and childhood depressive symptoms. *Dissertation Abstracts International. Section B: The Sciences and Engineering, 60(4-B),* 1524.

Haley, G.M.T., Fine, S., Marriage, K., Moretti, M.M., & Freeman, R.J. (1985). Cognitive bias and depression in psychiatrically disturbed children and adolescents. *Journal of Consulting and Clinical Psychology, 53,* 535–537.

Hammen, C., Adrian, C., Gordon, D., Burge, D., Jaenicke, C., & Hiroto, D. (1987). Children of depressed mothers: Maternal strain and symptom predictors of dysfunction. *Journal of Abnormal Psychology, 96,* 190–198,

Hammen, C., Adrian, C, & Hiroto, D. (1988). A longitudinal test of the attributional vulnerability model in children at risk for depression. *British Journal of Clinical Psychology, 27,* 37–46.

Harrison, C. (1980). *Readability in the classroom*. Cambridge: Cambridge University Press.

Helsel, W.J., & Matson, J.L. (1984). The assessment of depression in children: The internal structure of the Child Depression Inventory (CDI). *Behaviour Research and Therapy, 22,* 289–298.

Hepperlin, C.M., Stewart, G.W., & Rey, J.M. (1990). Extraction of depression scores in adolescents from a general-purpose behaviour checklist. *Journal of Affective Disorders, 18,* 105–112.

Hodges, K. (1990). Depression and anxiety in children: A comparison of self-report questionnaires to clinical interview. *Psychological Assessment, 2,* 376–381.

Hodges, K., & Craighead, W.E. (1990). Relationship of Children's Depression Inventory factors to diagnosed depression. *Psychological Assessment, 2,* 489–492.

Hoffman, K.B., Cole, D. A,, Martin, J.M., Tram, J., & Seroczynski, A.D. (2000). Are the discrepancies between self- and others' appraisals of competence predictive or reflective of depressive symptoms in children and adolescents: A longitudinal study. Part II. *Journal of Abnormal Psychology, 109,* 651–662.

Houghton, S., O'Connell, M., & O'Flaherty, A. (1998). The use of the Children's Depression Inventory in an Irish context. *Irish Journal of Psychology, 19,* 313–331.

Howard, K.L., Kopta, S.M., Krause, M.S., & Orlinsky, D.E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41,* 159–164.

Huddleston, E.N., & Rust, J.O. (1994). A comparison of child and parent ratings of depression and anxiety in clinically referred children. *Research Communications in Psychology, Psychiatry and Behavior, 19,* 101–112.

Ines, T M., & Sacco, W.P. (1992). Factors related to correspondence between teacher ratings of elementry student depression and student self-ratings. *Journal of Consulting and Clinical Psychology, 60,* 140–142.

Jacobs, M.L. (1990). Diagnosis of depression in school-aged exceptional family member children: The Children's Depression Inventory as a screening tool. *Dissertation Abstracts International. 50(9-B),* 4205–4206.

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12–19.

Jensen, P.S., Bloedau, *L.,* Degroot, J., Ussery, T, & Davis, H. (1990). Children at risk: L Risk factors and child symptomatology. *Journal of the American Academy of Child and Adolescent Psychiatry,* 29, 51–59.

Joiner, T.E., Schmidt, K.L., & Schmidt, N.B. (1996). Low-end specificity of childhood measures of emotional distress: Differential effects for depression and anxiety. *Journal of Personality Assessment, 67,* 258–271.

Kashani, J.H., & Carlson, G.A. (1985). Major depressive disorder in a preschooler. *Journal of the American Academy of Child Psychiatry, 24,* 490–494.

Kashani, J.H., Husain, A., Shekim, W.O., Hodges, K.K., Cytryn, L., & McKnew, D.H. (1981). Current perspective on childhood depression: An overview. *American Journal of Psychiatry, 138,* 143–153.

Kaslow, N.J., Rehm, L.P., & Siegel, A.W. (1984). Social-cognitive and cognitive correlates of depression in children. *Journal of Abnormal Child Psychology, 12,* 605–620.

Kazdin, A.E., Colbus, D., & Rodgers, A. (1986). Assessment of depression and diagnosis of depressive disorder among psychiatrically disturbed children. *Journal of Abnormal Child Psychology, 14,* 499–515.

Kazdin, A.E., Esveldt-Dawson, K., Unis, A.S., & Rancurello, M.D. (1983). Child and parent evaluations of depression and aggression in psychiatric inpatient children. *Journal of Abnormal Child Psychology, 11,* 401–413.

Kazdin, A.E., French, N.H., Unis, A.S., & Esveldt-Dawson, K. (1983). Assessment of childhood depression: Correspondence of child and parent ratings. *Journal of the American Academy of Child Psychiatry,* 22, 157–164.

Kazdin, A.E., French, N.H., Unis, A.S., Esveldt-Dawson, K., & Sherick, R.B. (1983), Hopelessness, depression, and suicidal intent among psychiatrically disturbed inpatient children. *Journal of Consulting and Clinical Psychology, 51,* 504–510.

Kazdin, A.E., & Petti, T.A. (1982). Self-report and interview measures of childhood and adolescent depression. *Journal of Child Psychology and Psychiatry, 23,* 437–457.

Kazdin, A.E., Rodgers, A., & Colbus, D. (1986). The Hopelessness Scale for Children: Psychometric characteristics and concurrent validity. *Journal of Consulting and Clinical Psychology, 54,* 241–245.

Knight, D., Hensley, V.R., & Waters, B, (1998). Validation of the Children's Depression Scale and the Children's Depression Inventory in a prepubertal sample. *Journal of Child Psychology and Psychiatry, 29,* 853–863.

Koizumi, S. (1991). The standardization of Children's Depression Inventory. *Syoni Hoken Kenkyu (The Journal of Child Health), 50,* 717–721.

Kovacs, M. (1985). The Children's Depression Inventory. *Psychopharmacology Bulletin, 21,* 995–998.

Kovacs, M. (1992). *The Children's Depression Inventory (CDI) manual* Toronto: Multi-Health Systems Inc.

Kovacs, M. (1995). *The Children's Depression Inventory (CDI) software manual.* Toronto: Multi-Health Systems Inc.

Kovacs, M. (1996a). Presentation and course of major depressive disorder during childhood and later years of the life span. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 705–715.

Kovacs, M. (1996b). The course of childhood-onset depressive disorders. *Psychiatric Annals, 26, 326–330.*

Kovacs, M. (1997a). *The Children's Depression Inventory Parent Version (CDI-P).* Toronto: Multi-Health Systems Inc.

Kovacs, M. (1997b). *The Children's Depression Inventory Teacher Version (CDI-T).* Toronto: Multi-Health Systems Inc.

Kovacs, M. (in press). *Emotional Regulation Scales (ERS).* Toronto: Multi-Health Systems Inc.

Kovacs, M., Akiskal, H.S., Gatsonis, C, & Parrone, P.L. (1994). Childhood onset dysthymic disorder: Clinical features and prospective naturalistic outcome. *Archives of General Psychiatry, 51,* 365–374.

Kovacs, M., & Beck, A.T. (1977), An empirical-clinical approach toward a definition of childhood depression. In J.G. Shulterbrandt & A. Raskin (Eds.), *Depression in childhood: Diagnosis, treatment, and conceptual models* (pp. I-25). New York: Raven Press.

Kovacs, M., Gatsonis, C., Paulauskas, S.L., & Richards, C. (1989). Depressive disorders in childhood: IV. A longitudinal study of comorbidity with and risk for anxiety disorders. *Archives of General Psychiatry, 46,* 776–782.

Kovacs, M., & Goldston, D. (1991). Cognitive and social cognitive development of depressed children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 388–392.

Kovacs, M., lyengar, S., Stewart, J., Obrosky, S., & Marsh, J. (1990). Psychological functioning of children with insulin-dependent diabetes mellitus: A longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 388–392.

Kovacs, M., Obrosky, S., Gatsonis, C., & Richards, C. (1997). First-episode major depressive and dysthymic disorder in childhood: Clinical and sociodemographic factors in recovery. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 777–784.

Krane, N.J. (1996). A comparative study of effortful information processing in subclinically depressed and nondepressed elementary school children. *Dissertation Abstracts International. Section A: The Humanities and Social Sciences, 56(11-A),* 4326.

Kuttner, M.J., Delamater, A.M., & Santiago, J.V. (1989). Learned helplessness in diabetic youths. *Journal of Pediatric Psychology, 15,* 581–594.

Lachar, D., & Gdowski, C. L, (1979). *Actuarial assessment of child and adolescent personality:*

An interpretive guide for the Personality Inventory for Children Profile. Los Angeles: Western Psychological Services.

Lam, K. N, (2000). An etic-emic approach to validation of the Chinese version of the Children's Depression Inventory. *Dissertation Abstracts International Section B: The Sciences and Engineering, 60(11–8),* 5780.

Lanktree, C.B., & Briere, J. (1995). Outcome of therapy for sexually abused children: A repeated measures study. *Child Abuse and Neglect, 19, 1145–1155.*

Linna, S., Moilanen, L, Ebeling, H., Piha, J., Kumpulainen, K., Tamminen, T., et al. (1999). Psychiatric symptoms in children with intellectual disability *European Child and Adolescent Psychiatry, 8,* 77–82.

Lipovsky, J.A., Finch, A.J., & Belter, R.W. (1989). Assessment of depression in adolescents: Objective and projective measures. *Journal of Personality Assessment, 53, 449–458.*

Liss, H., Phares, V., & Liljequist, L. (2001). Symptom endorsement differences on the Children's Depression Inventory with children and adolescents on an inpatient unit. *Journal of Personality Assessment, 76,* 396–411.

Llabre, M.M., & Hadi, F. (1997). Social support and psychological distress in Kuwaiti boys and girls exposed to the Gulf crisis, *Journal of Clinical Child Psychology, 26,* 247–255.

Lobert, W. (1989). Untersuchung von Merkmalen depress!ver Verstimmung in der Pubertat mit dem Kinder-Depressions-Inventar nach Kovacs [Investigation of symptoms of depressive moodiness during puberty with the Children's Depression Inventory according to Kovacs]. *Zeitschrift fur Kinder und Jugendpsychiatrie, 17,* 194–201.

Lobert, W. (1990). Untersuchung zur Struktur der depressiven Verstimmung in der Pubertat mit dem GCDI (German Children's Depression Inventory) [Investigation of the structure of depressive moodiness during puberty with the GCDI (German Children's Depression Inventory)]. *Zeitschrift fur Kinder und Jugendpsychiatrie, 18,* 18–22.

Lobovits, D.A., & Randal P.J- (1985). Childhood depression: Prevalence using DSM-III criteria and validity of parent and child depression scales. *Journal of Pediatric Psychology, 10,* 45–54.

Lord, F.M., & NTovick, M.R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

March, J.S. (1997). *The Multidimensional Anxiety Scale for Children (MASC).* Toronto: Multi-Health Systems Inc.

Marciano, P.L., & Kazdin, A.E. (1994). Self-esteem, depression, hopelessness, and suicidal intent among psychiatrically disturbed inpatient children. *Journal of Clinical Child Psychology, 23,* 151–160.

Marriage, K., Fine, S., Moretti, M., & Haley, G. (1986). Relationship between depression and conduct disorder in children and adolescents. *Journal of the American Academy of Child Psychiatry, 25,* 687–691.

Matthey, S., & Petrovski, P. (2002). The Children's Depression Inventory: Error in cutoff scores for screening purposes. *Psychological Assessment, 41,* 146–149.

Mattison, R.E., Handford, H.A., Kales, H.C., Goodman, A.L., & McLaughlin, R.E. (1990). Four-year predictive value of the Children's Depression Inventory. *Psychological Assessment, 2,* 169–174.

McCauley, E., Mitchell, J.R., Burke, P., & Moss S. (1988). Cognitive attributes of depression in children and adolescents. *Journal of Consulting and Clinical Psychology, 56,* 903–908.

Meins, W. (1993). Assessment of depression in mentally retarded adults: Reliability and validity of the Children's Depression Inventory (CDI). *Research in Developmental Disabilities, 14,* 299–312.

Mestre, V., Frias, D., & Garcia-Ros, R. (1992). Propiedades psicomettricas del Children's Depression Inventory (CDI) en poblacion adolescente: Fiabilidad y validez [Psychometric properties of the Children's Depression Inventory (CDI) in the adolescent population: Reliability and validity]. *Psicologica, 13,* 149–159.

Meyer, N.E., Dyck, D.G., & Petrinack, R.J. (1989). Cognitive appraisal and attributional correlates of depressive symptoms in children. *Journal of Abnormal Child Psychology, 17,* 325–336.

Milich, R., Roberts, M.A., Loney, J., & Caputo, J. (1980). Differentiating practice effects and statistical regression on the Coroners' Hyperkinesis Index. *Journal of Abnormal Psychology, 8,* 549–552.

Milne, J., & Spence, S.H. (1987). Training social perception skills with primary school children: A cautionary note. *Behavioural Psychotherapy, 15,* 144–157.

Moretti, M.M., Fine, S., Haley, G., & Marriage, K. (1985). Childhood and adolescent depression: Child-report versus parent-report information. *Journal of the American Academy of Child Psychiatry, 24,* 298–302.

Nelson, W, M., & Politano, P.D. (1990). Children's Depression Inventory: Stability over repeated administrations in psychiatric inpatient children. *Journal of Clinical Child Psychiatry, 19,* 254–256.

Nelson, W.M., Politano, P.M., Finch, A.J., Wendel, N., & Mayhall, C. (1987). Children's Depression

Inventory: Normative data and utility with emotionally disturbed children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 43–48.
Newman, F.L., Ciarlo, J.A., & Carpenter, D. (1999). Guidelines for selecting psychological instruments for treatment planning and outcome assessment. In M.E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed., pp. 153–170). Hillsdale, NJ: Lawrence Erlbaum Associates.
Nieminen, G.S., & Matson, J.L. (1989). Depressive problems in conduct-disordered adolescents. *Journal of School Psychology, 27,* 175–188.
Nolen-Hoeksema, S., Girgus, J.S., & Seligman, M.E.P. (1986). Learned helplessness in children: A longitudinal study of depression, achievement, and explanatory style. *Journal of Personality and Social Psychology, 51,* 435–442.
Norvell, N., Brophy, C., & Finch, A.J. (1985). The relationship of anxiety to childhood depression. *Journal of Personality Assessment, 49,* 150–153.
Ollendick, T.H., & Yule, W. (1990). Depression in British and American children and its relation to anxiety and fear. *Journal of Consulting and Clinical Psychology, 58,* 126–129.
Oy, B. (1991). Children's Depression Inventory: A study of reliability and validity. *Turk Psikiyatri Dergisi, 2,* 132–136.
Pfeffer, C., Karus, D., Siegal, K., & Jiang, H. (2000). Child survivors of parental death from cancer or suicide: Depressive and behavioral outcomes. *Psycho-Oncology, 9,* 1–10.
Polaino-Lorente, A., & del-Pozo-Armentia, A. (1992). Modification de la depresion mediante un programa de intervencion psicopedagogica en ninos cancerosos no hospitalizados [Modification of depression through a psychopedagogic intervention program in childhood cancer hospitalization]. *Analisis y Modification de Conducta, 18, 493–503.*
Polaino-Lorente, A., & Domenech, E. (1993). Prevalence of childhood depression: Results of the first study in Spain. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 34,* 1007–1017.
Politano, P.M., Nelson, W.M., Evans, H.E., Sorenson, S.B., & Zeman, D.J. (1985). Factor analytic evaluation of differences between Black and Caucasian emotionally disturbed children on the Children's Depression Inventory. *Journal of Psychopathology and Behavioral Assessment, 8,* 1–7,
Pons-Salvador, G., & del Barrio, V. (1993). Depresion infantil y divorcio [Child depression and divorce]. *Avances en Psicologia Clinica Latinoamericana, 11,* 95–106.
Ponterotto, J.G., Pace, T.M., & Kavan, M.G. (1989). A counselor's guide to the assessment of depression. *journal of Counseling and Development, 67,* 301–309.
Preiss, M. (1998). Hloubka depresivity v sebeposuzovaci skale CDI u deti po valce v Bosne [Depression depth in self-rating scale CDI in children after the war in Bosnia]. *Ceskoslovenska Psychologic, 42,* 558–564.
Preskorn, S.H., Weller, E.B., Hughes, C.W., Weller, R.A., & Bolte, K. (1987). Depression in prepubertal children: Dexamethasone nonsupression predicts differential response to imipramine vs. placebo. *Psychopharmacology Bulletin, 23,* 128–133.
Puig-Antich, J. (1982). Major depression and conduct disorder in prepuberty. *Journal of the American Academy of Child Psychiatry, 21,* 118–128.
Reicher, H., & Rossmann, P. (1991). Zu den psychometrischen Eigenschaften einer deutschen Version des Children's Depression Inventory [The psychometric properties of a German version of the Children's Depression Inventory]. *Diagnostica, 37,* 236–251.
Reinhard, H.G., Bowi, U., & Rulcovius, G. (1990). Stabilitat, Reliabilitat und Faktorensrruktur einer deutschen Fassung des Children's Depression Inventory [Reliability, stability, and factor structure of a German version of the Children's Depression Inventory]. *Zeitschrift fur Kinder und Jugendpsychiatrie$_f$ 18,* 185–191.
Remherz, H.Z.., Frost, A.K., & Pakiz, B. (1991). Changing faces: Correlates of depressive symptoms in late adolescence. *Family and Community Health, 14,* 52–63.
Renouf, A.G., & Kovacs, M. (1994). Concordance between mothers' reports and children's self-reports of depressive symptoms: A longitudinal study. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 208–216.
Reynolds, C.R., & Richmond, B.O. (1985). *Revised Children's Manifest Anxiety Scale manual.* Los Angeles: Western Psychological Services.
Reynolds, W.M., Anderson, G., & Bartell, N. (1985). Measuring depression in children: A multimethod assessment investigation. *Journal of Abnormal Child Psychology, 13,* 513–526.
Rick, S. (1999). Coping behaviors of sexually abused boys as related to levels of depression and hopelessness, (child sexual abuse). *Dissertation Abstracts International, Section B: The Sciences and Engineering, 60(4-B),* 1535.

Rotundo, N., & Hensley, V.R. (1985). The Children's Depression Scale. A study of its validity. *Journal of Child Psychology and Psychiatry, 26,* 917–927.

Rybolt, Y. (1995). Assessment of depression in school-age children: A cross-cultural comparison of Mexican American and Caucasian students. *Dissertation Abstracts International. Section A: The Humanities and Social Sciences, 56(1-A),* 0146.

Sacco, W.P., & Graves, D.J. (1985). Correspondence between teacher ratings of childhood depression and child self-ratings. *Journal of Clinical Child Psychology, 14,* 353–355.

Saint-Laurent, L. (1990). Psychometric study of Kovac's Children's Depression Inventory with a French-speaking sample. *Canadian Journal of Behavioural Science, 22,* 377–384.

Sakurai, S. (1991). The relation between depression and causal attributional style in Japanese children, *Japanese Journal of Health Psychology, 4,* 23–30.

Saylor, C.F., Finch, A.J., Jr., Baskin, C.H., Furey, W., & Kelly, M.M. (1984). Construct validity for measures of childhood depression: Application of multitrait-multimethod methodology. *Journal of Consulting and Clinical Psychology, 52,* 977–985.

Saylor, C.F., Finch, A.J., Jr., Baskin, C.H., Saylor, C.B., Darnell, G., & Furey, W. (1984). Children's Depression Inventory: Investigation of procedures and correlates. *Journal of the American Academy of Child Psychiatry, 23,* 626–628.

Saylor, C.F., Finch, A.J., Jr., Spirito, A., & Bennett, B. (1984). The Children's Depression Inventory: A systematic evaluation of psychometric properties. *Journal of Consulting and Clinical Psychology, 52,* 955–967.

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

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.

Shah, F., & Morgan, S.B. (1996). Teacher's ratings of social competence of children with high versus low levels of depressive symptoms. *Journal of School Psychology, 34,* 337–349.

Siegel, K., Karus, D., & Raveis, V.H. (1996). Adjustment of children facing the death of a parent due to cancer. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 442–450.

Simmer-Dvonch, L.M. (1999). Development and evaluation of a group treatment for sexually abused male adolescents. *Dissertation Abstracts International. Section B: The Science and Engineering, 59(7-B),* 3715.

Slotkin, J., Forehand, R., Fauber, R., McCombs, A., & Long, N. (1988). Parent-completed and adolescentcompleted GDIs: Relationship to adolescent social and cognitive functioning. *Journal of Abnormal Child Psychology, 16,* 207–217.

Speer, D.C., & Greenbaum, P.E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63,* 1044–1048,

Spence, S.H., & Milne, J. (1987). The Children's Depression Inventory: Norms and factor analysis from an Australian school population. *Australian Psychologist, 22,* 345–351.

Spirito, A., Overholser, J., & Hart, K. (1991). Cognitive characteristics of adolescent suicide attempters. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 604–608.

Stark, K.D., Kaslow, N.J., & Laurent, J. (1993). The assessment of depression in children: Are we assessing depression or the broad-band construct of negative affectivity? *Journal of Emotional and Behavioral Disorders, 1,* 149–154.

Stavrakaki, C., Williams, E.C., Walker, S., Roberts, N., & Kotsopoulos, S. (1991). Pilot study of anxiety and depression in prepubertal children. *Canadian Journal of Psychiatry, 36,* 332–338.

Sriensmeier-Peister, J., Schumann, M., & Duda, K. (1991). Das Depressionsinventar fur Kinder und Jugendliche (DIKJ): Untersuchungen zu seinen psychometrischen Eigenschaften [The psychometric properties of the German version of the Children's Depression Inventory!. *Diagnostica, 37,* 149–159.

Stiensmeier-Pelster, J., Schurmann, M., & Urhahne, D. (1991). Das Depressionsinventar fur Kinder und Jugendliche (DIKJ): Einsetzbarkeit in der Primarstufe [The Depression Inventory for Children and Adolescents (DICA): Its applicability on the elementary school level]. *Zeitschrift fur Entwicklungspsychohgie und Padagogische Psychologie, 23,* 171–176.

Stacker, C.M. (1994). Children's perceptions of relationships with siblings, friends, and mothers: Compensatory processes and links with adjustment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 35,* 1447–1459.

Strober, S., & Carlson, G. (1982). Bipolar illness in adolescents with major depression: Clinical, genetic, and psychopharmacologic predictors in a three- to four-year prospective follow-up investigation. *Archives of General Psychiatry, 39,* 549–555.

Timbremont, B., & Braet, C. (2001). Psychometric assessment of the Dutch version of the Children's Depression Inventory. *Gedragstherapie, 34,* 229–242.

Twenge, J., & Nolen-Hoeksema, S. (2002). Age, gender, race, sotioeconomic status, and birth cohort differences on the Children's Depression Inventory: A meta-analysis. *Journal of Abnormal Psychology, 4,* 578–588.

Weiss, B., & Weisz, J.R. (1988). Factor structure of self-reported depression: Clinic-referred children versus adolescents. *Journal of Abnormal Psychology, 97,* 492–495.

Wierzbicki, M. (1987). A parent form of the Children's Depression: Reliability and validity in non-clinical populations. *Journal of Clinical Psychology, 43,* 390–397.

Weiss, B., Weisz, J.R., Politano, M., Carey, M., Nelson, W.M., & Finch, A.J. (1991). Developmental differences in the factor structure of the Children's Depression Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 38–45.

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.

Wolfe, V.V, Finch, A.J., Jr., Saylor, C., Blount, R.L., Pallmeyer, T P., & Carek, D.J. (1987). Negative affectivity in children: A multitrait-multimethod investigation. *Journal of Consulting and Clinical Psychology, 55,* 245–250.

Worchel, F.F., Hughes, J.N., Hall, B. M, Stanton, S.B., Stanton, H., & Little, V.Z. (1990), Evaluation of subclinical depression in children using self-, peer-, and teacher-report measures. *Journal of Abnormal Child Psychology, 18,* 271–282.

Worchel, F., Nolan, B., & Willson, V. (1987). New perspectives on child and adolescent depression. *Journal of School Psychology, 25,* 411–414.

Yu, D., & Li, Xu. (2000). Preliminary use of the Children's Depression Inventory (CDI) in China. *Chinese Mental Health journal, 14,* 225–227.

Zivcic, I. (1993). Emotional reactions of children to war stress in Croatia. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 709–713.

2

The Multidimensional Anxiety Scale for Children (MASC)

John S.March
Duke University Medical Center

James D.A.Parker
Trent University

Presumably because pathological anxiety is associated with significant suffering, disruption in normal psychosocial and academic development and family functioning, and increased utilization of medical services, "worry" is among the more common causes of referral to children's mental health care providers (Black, 1995; Simon, Ormel, Von Korff, & Barlow, 1995). Unfortunately, the population prevalence of childhood-onset fears, the structure of anxiety symptoms in the general pediatric population, and the relative importance of specific anxiety dimensions within gender, ethnic, or cultural groupings across time have, until recently, remained unclear (March & Albano, 1998). This is in part because of a lack of acceptable measurement tools (Costello & Angold, 1995; Greenhill, Pine, March, Birmaher, & Riddle, 1998; March & Albano, 1998).

Ideally, instruments intended to assess anxiety in pediatric patients should (a) provide reliable and valid ascertainment of symptoms across multiple symptom domains; (b) discriminate symptom clusters; (c) differentiate normal from pathological anxiety both qualitatively and quantitatively; (d) incorporate and reconcile multiple observations, such as parent and child ratings; and (e) be sensitive to treatment-induced change in symptoms. Other factors that may influence instrument selection include the reasons for the assessment—screening, diagnosis, or monitoring treatment outcome, for example—as well as time required for administration, level of training necessary to administer and/or interpret the instrument, reading level, and cost. Finally, with increasing emphasis on multidisciplinary approaches to assessment and treatment, assessment tools must facilitate communication, not only among clinicians but also between clinicians and regulatory bodies, such as utilization review committees within managed care environments.

Though currently available instruments fall well short of these goals, a complex matrix of tools for assessing normal and pathological fears is now available (Greenhill et al., 1998; March & Albano, 1998). In this chapter, we describe one such instrument, the Multidimensional Anxiety Scale for Children (MASC), which was designed to address the multidimensional assessment of anxiety in children and adolescents in a psychometrically rigorous fashion (March, 1998; March, Parker, Sullivan, Stallings, & Conners, 1997). Excellent reviews of pediatric anxiety disorders in general (March, 1995; Ollendick & King, 1994) and assessment issues in particular (Greenhill et al., 1998) are available.

BACKGROUND

Instruments designed specifically to address anxiety in children and adolescents are required for several reasons. First, children appear to undergo a developmentally sanctioned progression in anxiety symptoms (Keller et al., 1992; Last, Strauss, & Francis, 1987). Second, their day-to-day environments differ from those most typically experienced by adults so that the presentation of anxiety also differs, as in "school phobia." Third, to differentiate normal from pathological anxiety, gender and age norms are necessary. Finally, some fears may be viewed as adaptive protective; only when anxiety is excessive or the context is developmentally inappropriate does anxiety becomes clinically significant (Marks, 1987). Other fears, such those seen in obsessive-compulsive disorder, are developmentally inappropriate under many if not all circumstances (Leonard, Goldberger, Rapoport, Cheslow, & Swedo, 1990). Thus clinicians and researchers interested in childhood anxiety disorders face the challenging task of differentiating pathological anxiety from fears occurring as a part of normal developmental processes . The DSM-III-R (American Psychiatric Association, 1987) addressed this nosological conundrum by introducing a subclass of anxiety disorders of childhood and adolescence. The *DSM-IV* (American Psychiatric Association, 1994) both refines these constructs and establishes a greater degree of continuity— developmental and nosological—with the adult anxiety disorders. The *DSM* taxonomy in essence reflects an expert consensus regarding the actual clustering of anxiety in pediatric samples. Though empirical support for the DSM "factor structure" in some cases is questionable (e.g., generalized anxiety), for other constructs it is more robust (e.g., separation or social anxiety; March et al., 1997).

Some anxiety symptoms, such as refusing to attend school in the patient with panic disorder and agoraphobia, are readily observable; other symptoms are open only to child introspection and thus to child self-report. For this and other reasons, self-report measures of anxiety, which provide an opportunity for children to reveal their internal or "hidden" experiences, have found wide application in both clinical and research settings, Typically, self-report measures use a Likert scale format in which a child is asked to rate each questionnaire using either a frequency or intensity format. For example, a child might be asked to rate "I feel tense" on a fourpoint frequency scale that ranges from almost never to often. Self-report measures are easy to administer, require a minimum of clinician time, and economically capture a wide range of important anxiety dimensions from the child's point of view. Taken together these features make self-report measures ideally suited to gathering data prior to the initial evaluation, as self-report measures used in this fashion increase clinician efficiency by facilitating accurate assessment of the prior probability that a particular child will or will not have symptoms within a specific symptom domain.

For the most part, available self-report rating scales for assessing pediatric anxiety have until now represented age-downward extensions of adult measures that fail to capture or adequately operationalize important dimensions of anxiety in young persons (March & Albano, 1998). Three commonly cited instruments have been in use for over 20 years. The Fear Survey Schedule for Children-Revised (FSSC-R) focuses primarily on phobic symptoms, including fear of failure and criticism, fear of the unknown, fear of injury and small animals, fear of danger and death, and medical fears (Ollendick, 1983). The Revised Children's Manifest Anxiety Scale (RCMAS) provides three factors: physiological manifestations of anxiety, worry and oversensitivity, and fear/concentration (Reynolds & Richmond, 1979). However, the presence of mood, attentional, impulsivity,

TABLE 2.1
Anxiety Rating Scales

	MASC	*RCMAS*	*FSSC-R*	*STAIC*
Broad conceptualization	Yes	Yes	No	Yes
Specific dimensions	Yes	Partial	Phobias	No
Matches *DSM-IIV*	Yes	No	No	No
Reliable	Yes	Yes	Yes	Trait scale
Convergent validity	Yes	Yes	Yes	Yes
Divergent validity	Yes	No	No	No

and peer interaction items on the RCMAS clearly confound other diagnoses, such as ADHD and major depression (Perrin & Last, 1992). Another widely used measure, the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, Gorsuch, & Luchene, 1976), consists of two independent 20-item inventories that assess anxiety symptoms from a variety of domains but do not exhaustively cover the symptom constellations represented in *DSM-IV*. The State scale purports to assess present-state and situation-linked anxiety; the Trait scale addresses temporally stable anxiety across situations. Numerous authors have questioned the validity of the state-trait distinction (Kendall, Finch, Auerbach, Hooke, & Mikulka, 1976) and the nature of item selection for the STAIC (Finch, Kendall, & Montgomery, 1976; Perrin & Last, 1992). Table 2.1 contrasts these older measures with the MASC with respect to construct validity, applicability to *DSM-IV,* reliability, and convergent and divergent validity. Thus, the MASC was developed within the context of broad agreement by clinicians and researchers that new instruments were necessary if the field of pediatric anxiety disorders was to progress scientifically (see, e.g., Jensen, Salzberg, Richters, & Watanabe, 1993; March & Albano, 1996).

OVERVIEW OF THE MASC

The MASC is a 39-item Likert-style self-report measure developed to index a wide range of anxiety symptoms in elementary, junior high, and high school age youngsters (8 to

TABLE 2.2
MASC Factors and Subfactors

Physical Symptoms
Tense
Somatic
Social Anxiety
Humiliation Fears
Performance Fears
Harm Avoidance
Perfectionism
Anxious Coping
Separation Anxiety

19 years old). As shown in Table 2.2, the MASC has four main factors, three of which can be further divided into two subfactors. Taken together, these factors and subfactors capture the central constructs of pediatric anxiety as they emerge in both population and clinical samples.

Procedures for developing and psychometrically validating a new rating scale are complex and time consuming (Cicchetti, 1994). In developing the MASC, the following sequence was used:

- An exhaustive review of available rating scales, diagnostic interviews, and the *DSM-IV* generated over 400 potential items.
- A Q-sort procedure was used to divide these items into cognitive, emotional, physical, and behavioral categories.
- A data reduction procedure generated a 41-item scale representing the four conceptual domains.
- A pilot study of over 1,000 elementary, junior high, and senior high school students was conducted in a school-based community sample.
- Based on results from the pilot study, which yielded a five-factor solution, a 104-item scale (with approximately 20 items per factor) was again piloted in a school-based sample.
- Principle components factor analyses of data from this population survey provided the current MASC factor structure, which shows excellent internal reliability without excessive redundancy in item content.
- Based on further clinical and research experience using the scale with children and adolescents aged 5–18, 39 items were retained for the final version of the MASC.
- Confirmatory factor analyses in clinical and community populations and in a large sample of ADHD children replicated the MASC factor structure.
- Parent-child and parent-parent concordance was low to moderate, depending on the domain of symptomatology being assessed; this finding indicated the clinical usefulness of the MASC as a child self-report measure.
- Convergent and divergent validity of the MASC with respect to parent ratings of externalizing behavior and internalizing symptoms was shown to be high.
- Test-retest reliability (stability over time) has been demonstrated in clinical and epidemiological samples.
- The MASC has been shown to be treatment sensitive.
- The MASC in now in wide use in industry- and foundation-funded studies of pediatric anxiety disorders and studies funded by the National Institute of Mental Health (NIMH),

The theoretical background, initial construction, validation, reliability, and norming of the MASC are extensively discussed in the MASC manual (March, 1998).

DEVELOPMENT OF THE MASC

Preliminary Studies

Although work to date on the taxonomy of anxiety in children and adolescents provides limited support for the *DSM-IV* anxiety clusters (see, e.g., Silverman & Eisen, 1992),

some have suggested that a broader conceptualization is necessary (March & Albano, 1998; Ollendick & King, 1994; Ollendick, Matson, & Helsel, 1985). In contrast to scales that assess a specific *DSM-IV* anxiety construct (see, e.g., Beidel, Turner, & Morris, 1994), the MASC was developed to assess a wide spectrum of common anxiety symptoms in children across the elementary, junior high, and senior high school age range. Thus, when beginning the item selection procedure, we elected not to assume anything about the normative clustering of pediatric anxiety symptoms other than to hypothesize that specific descriptors should, on theoretical grounds (Marks, 1987), fall within emotional, cognitive, physical, and behavioral symptom domains.

The actual procedure followed several steps. First, available self-report anxiety scales covering general and specific symptom domains as well as the *DSM-III-R* criterion items were reviewed. Each of the over 400 resulting items/questions from these measures was then placed on a 3″×5″ card and sorted by two expert clinicians into four symptom domains: cognitive, physical, emotional, and behavioral. Cognitive items were defined as ascertaining a thought, urge, or image, which could be specific, as in a fear of dogs, or general, as in "worry." Physical items were characterized by physiological indicators, such as nausea or a racing heart. Emotional items were defined as ascertaining a subjective feeling, such as fear, or a subjective sensation, such as tension. Behavioral items was defined as ascertaining operant mechanisms of anxiety reduction through approach behaviors, such as reassurance seeking, or avoidance behaviors, such as avoiding public speaking. Disagreements were resolved by forced consensus judgment.

Second, the item pools were reduced by (a) retaining items that were easy to understand, covered the desired age range, and closely reflected one and only one of the four chosen anxiety dimensions and (b) eliminating duplicates and rewording. Third, a Q-sort procedure was used to enhance item-content validity. Expert clinicians, members of an anxiety disorders support group, and lay nonexperts classified 60 items (15 per group) into the four selected domains.

Fourth, based on their comments and the pattern of misclassification, a 41-item, four-point Likert scale—having approximately 10 items per hypothesized symptom domain—was developed and piloted in a population sample of 1,066 fourth- through eighth-grade students. A three-point Likert version was not entertained because of the possibility of excessive midpoint responding—one of the drawbacks, for example, of the RCMAS.

Results from this preliminary study suggested a five-factor solution, which only partially conformed to the hypothesized four-domain model of anxiety: Somatic/Autonomic Arousal (14 items), Fears and Worries (7 items), Social Fears (10 items), Behavioral Avoidance/Approach (6 items), and Separation Anxiety (4 items). The uneven distribution of the items, which attenuated the internal reliability of the smaller factors, coupled with the lack of precision in the model, indicated the need for further scale development.

Based on the results from the first study, additional items (from the initial pool) were added to the five factors to bring each up to a total item pool of approximately 20 items. The resultant 104-item questionnaire was then administered to a population sample of 374 third- through twelfth-grade students. One classroom from each school was chosen at random for each grade; subjects thus were evenly split between Grades 4 to 12. Elementary school students were tested in their usual classroom, junior high school students in their homeroom. Questionnaires were read aloud to students, who had the opportunity to ask questions about individual items but not to seek clarification about how they should respond. Like the earlier questionnaire, this questionnaire also used a four-point Likert scale in which respondents were asked to rate each question as

"Never," "Sometimes," "Rarely," and "Always true about me." Students with reading disabilities were given extra time or reading support as needed. Teachers provided demographic information.

Factor Structure

With these data in hand, we then conducted a series of exploratory principal components factor analyses (using Varimax rotation) on the total sample. A robust four-factor solution emerged: Physical Symptoms, Social Anxiety, Separation Anxiety, and Harm Avoidance (March, 1998; March et al., 1997). All four factors had 9 items except the first, which had 10 items. Specifying a conventional Eigenvalue of 1.0 as the PCA entry criterion generated additional factors. In contrast to the reported factor structure, where between-factor overlap proved minimal at the item level, these smaller factors explained little additional variance and contained items that tended to load across multiple factors.

Each major factor was then subjected to a principal components factor analysis (again using Varimax rotation). Three of the four main factors—all except the Separation Anxiety factor—produced a clear two-factor solution using an Eigenvalue of 1.0 as the entry. Physical Symptoms factored into Tense/Restless and Somatic/Autonomic subfactors, harm Avoidance factored into Perfectionism and Anxious Coping, Social Anxiety factored into Humiliation/Rejection Fears and Performance Anxiety, and the Separation Anxiety factor was found to be unidimensional. In all cases, the first listed subfactor carried the majority of the variance (March et al., 1997).

A large body of literature suggests that anxieties of all sorts are more common in females than males (Benjamin, Costello, & Warren, 1990) and that some symptoms, for example, separation anxiety, vary with age (Francis, Last, & Strauss, 1987). To establish between-group differences for age or gender when using a self-report questionnaire, it is crucial to first establish that the factor structures are identical. To this purpose, a multisample confirmatory factor analysis was conducted using the EQS (Bentler, 1995) statistical program to test whether the four-factor model for the 39 MASC items was equivalent for males and females. All factor loadings were constrained to be equal for males and females, as were the correlations between the four MASC factors. Multiple goodness-of-fit indicators revealed that the four-factor model fit well in both sexes. The nonnormed fit index (*NNFI;* Bentler & Bonnett, 1980) was 0.913, the comparative fit index (*CFI;* Bentler, 1988) was 0.916, and the incremental fit index (IF/; Bolen, 1989) was 0.917. The magnitude of the three indexes (above 0.90, as suggested by Bentler, 1995) suggests that the model had excellent fit to the data regardless of gender.

A multisample confirmatory factor analysis was also conducted to test whether the four-factor model was equivalent for younger and older students. The sample was separated into two age groups: 12 years and under (n=159) and 13 years and over (n=211). As suggested by Weiss et al. (1991) on theoretical grounds, this age cutoff approximates the move from concrete to formal operations in the context of emerging puberty. Multiple goodness-of-fit indicators revealed that the four-factor model fit well in both age groups: *NNFI=.976, CFI=.977,* and *IFI=.978*. Thus, we concluded that the MASC factor structure is invariant across age and gender.

Confirmatory Factor Analyses

Having established the factor structure of the MASC, we then sought to replicate the factor structure in two groups of subjects: a second large school-based sample of 2,698

children and adolescents and a clinical sample of 390 children and adolescents. As before, multiple goodness-of-fit indices were used to evaluate the fit of the data to the measurement model. In both nonclinical and clinical samples, the four-factor model for the 39-item MASC met the criteria standards for adequacy of fit (Bentler, 1988). Parameter estimates for the relationships were statistically significant. Thus, the data had good fit to the MASC model. Confirmatory factor analyses for the four-factor MASC model also have been conducted in a large sample of (mostly nonanxious) young children with ADHD, and these too demonstrated adequacy of fit of the data and thus the extraordinary robustness of the MASC factor structure (March et al., 1999). The overall conclusion to be gained from the confirmatory factor analyses is that the MASC factor structure replicates nicely across diverse samples of children and adolescents.

Reliability

Reliability in psychometric terms has several meanings, Internal reliability represents consistency between items within a group of items composing a discrete factor (Cronbach, 1970). Test-retest reliability represents consistency in a set of scores by the same rater (single-case intraclass correlation coefficient [ICC]) or set of raters (mean ICC) over time (Shrout & Fleiss, 1979). Test-retest reliability varies with the conditions under which the test is administered, practice or memory effects, true change in the variable(s) of interest, plus an instability component due to measurement error attributable to the instrument itself. Without adequate reliability, it is not possible to determine whether differences in scores between individuals or within subject over time are due to "true" differences or to "chance" error.

Internal Reliability. Using a cutoff of 0,6 (below which internal consistency is suspect), total sample α-reliabilities, which range from .6 to .85, are acceptable for all main factors and subfactors for the 39-item MASC (March et al., 1997). Internal reliability for the MASC total score is 0.9. Furthermore, α-reliabilities for the MASC total score are generally comparable for males (.85) and females (.87). Very high reliability coefficients (above .9) indicate excessive redundancy at the item level. Inspection of item content shows individual items within a factor or subfactor to be face valid for the measured construct but not redundant with respect to item content.

Test-Retest Reliability. In a clinical population of children and adolescents with a mixture of anxiety disorders and/or ADHD (March et al., 1997), we examined the test-retest reliability of the MASC at 3 weeks and 3 months using the intraclass correlation coefficient (ICC) calculated according to procedures outlined by Shrout and Fleiss (1979). Mean ICCs for the MASC total score were .785 at 3 weeks and .933 at three months, indicating satisfactory to excellent test-retest reliability (March et al., 1997). Similarly, mean ICCs for all factors and subfactors save the Harm Avoidance factor fell in the satisfactory to excellent range at 3 weeks; all factors and subfactors proved satisfactory to excellent at 3 months (March et al., 1997). Mean ICCs for the MASC-10 (an empirically derived short form) and an anxiety disorders index ranged from .64 to .89, again indicating satisfactory to excellent stability. More recently, we examined the test-retest reliability of the MASC in a school-based sample of children and adolescents (March & Sullivan, 1999). For both single-case and mean ICCs, the MASC exhibited satisfactory to excellent stability across all factors and subfactors. Importantly, reliability

was good to excellent for both genders, for younger and older children, and for Caucasian and African American youths. Satisfactory test-retest reliability also was demonstrated for the MASC-10 and for an anxiety disorders index with high discriminant validity. Thus, the MASC (uniquely at this point) can be said to demonstrate excellent test-retest reliability in both clinical and epidemiological samples.

Validity

Correlational Analysis. The factor structure of the MASC also is unique among extant scales in its subdivision of main factors into subfactors that nevertheless explain a meaningful proportion of the variance (March et al., 1997). With the exception of Perfectionism, which shows a weaker relationship to Physical Symptoms in females than in males, the pattern of shared variance as indicated by correlational analysis is similar for males and females. Importantly, although almost all correlations are significant at a Bonferroni-corrected alpha level of .05 or lower, the absolute magnitude of the shared variance is in the low to moderate range for most pairs. This suggests that the MASC is indeed measuring separate dimensions of anxiety, even at the subfactor level, which in turn should make it ideally suited to discriminate patterns of anxiety in subgroups of children with anxiety disorders.

Convergent and Divergent Validity. For the MASC to be useful clinically, the MASC factors would share greater variance with measures in the same symptom domain (convergent validity) than in different domains (divergent validity). In a test of this hypothesis in a clinical sample of children and adolescents with a variety of internalizing and externalizing disorders, we hypothesized that the MASC would be strongly correlated with a measure of anxiety (RCMAS), less so with a measure of depression (CDI), and not all correlated with a measure of disruptive behavior (ASQ-P). In all instances, the results went in the predicted direction, implying that the MASC is a specific indicator of pediatric anxiety symptomatology Notably, the MASC performed significantly better than either the RCMAS or the CDI in this regard (March et al., 1997).

More recently, Muris examined the correlation between the MASC and another new anxiety rating scale, the Screen for Child Anxiety Related Emotional Disorders (SCARED), which was by design keyed to the *DSM-IV* view of anxiety disorders in youths (Birmaher et al,, 1997). Not surprisingly, given the specificity for anxiety of both scales and also their differences in factor structure, the overall correlation between the sales was .72, with correlations between subtests ranging between .35 and .63 (Muris, Gadet, Moulaert, & Merckelbach, 1998). In an extension of these findings, Muris, Merckelbach, Ollendick, King, and Bogie (2002) extended these findings, comparing the psychometrics of three older scales, the RCMAS, STAIC, and FSSC-R, with the psychometrics three newer scales, the MASC, the SCARED, and the Spence Children's Anxiety Scale (SCAS; Spence, 1997) in a large sample of normal adolescents (N=521). In general, internal consistency was superior for the new scales. Reflecing their common origin in the DSM-IV, the SCARED and the SCAS were more strongly associated with each other than either scale was with the MASC, though all correlations were significant. Not surprisingly, subscales intended to measure specific categories of anxiety symptoms proved more strongly associated, with the MASC Harm Avoidance scale showing unique variance.

Predictive Validity. Using the Anxiety Disorders Interview Schedule for Children (ADIS-C) as the reference standard (Silverman & Albano, 1996), Deirker et al. (2001) recently examined the level of diagnostic and discriminative accuracy of three dimensional rating scales for detecting anxiety and depressive disorders in a school-based survey of ninth-grade youths. They concluded that MASC scores were most strongly associated with individual anxiety disorders, particularly among females, and successfully discriminated diagnosed depressed youths from anxious youths. In contrast, the RCMAS was not successful in discriminating anxiety and depression.

Similarly, Wood, Piacentini, Bergman, McCracken, and Barrios (2002) examined the concurrent validity of the ADIS diagnoses of social phobia, separation anxiety disorder (SAD), generalized anxiety disorder (GAD), and panic disorder diagnoses using the MASC as the reference standard. They identified little relationship between MASC scores and GAD diagnoses (though they did not examine the relevant subfactors), but they did notice a strong convergence between ADIS diagnoses and the empirically derived MASC social phobia, separation, and panic symptom constellations.

Discriminant Validity. Discriminant validity has been a persistent problem for older scales, such as the RCMAS. For example, Perrin and colleagues showed that the RCMAS and the STAIC differentiated children with *DSM-IH-R* anxiety and attention deficit disorders from normals but not from each other, whereas the FSSC-R was ineffective at discriminating between any grouping (Perrin & Last, 1992). We examined the discriminant validity of the four central scales from the MASC by using discriminant function analysis to predict group membership in patients with anxiety disorders versus normal controls. Two groups of children and adolescents were used in the present analysis. The first group consisted of children and adolescents who met *DSM-IV* criteria for an anxiety disorder other than obsessive-compulsive disorder. The second group (nonclinical) consisted of children and adolescents randomly selected from a large pool of subjects with normative data on the MASC and matched with the clinical sample on the basis of age and sex. A discriminant function analysis was performed using the four MASC subscales as predictors of membership in two groups (clinical vs. nonclinical). Discriminant function scores from this analysis were used to classify subjects into clinical or nonclinical groups. A variety of diagnostic efficiency statistics were calculated from these classification results: The sensitivity was 90%, the specificity was 84%, the positive predictive power was 85%, the negative predictive power was 89%, the false-positive rate was 16%, the false-negative rate was 11%, kappa was 0.74, and the overall correct classification rate was 87%. Interestingly, in the study by Muris and colleagues (2002), correlations among anxiety questionnaires were generally higher than those between anxiety scales and a measure of depression, with the MASC total score showing slightly better discriminant validity than other scales, again perhaps because of the included Harm Avoidance factor.

Females Are More Anxious Than Males. The literature consistently shows that, across ages and disorders, girls show more anxiety than boys (March, 1995; March & Sullivan, 1999). As expected, females show more anxiety than males in Bonferroni-corrected planned contrasts between item-mean scores for males and females on the 39-item MASC. These differences are significant at the $p<.001$ level, though the absolute magnitude of each difference is typically low.