THESIS
2
2001

LIBRARY
Michigan State
University

This is to certify that the

dissertation entitled

DISCRIMINANT VALIDITY OF THE
DEVEREUX BEHAVIOR RATING SCALE-SCHOOL FORM
WITH SPECIAL EDUCATION INITIAL REFERRALS

presented by

Barbara Sullivan Dunn

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in School Psychology

Major professor

Date 5-1-00

MSU is an Affirmative Action/Equal Opportunity Institution

O-12771

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
DEC 2 1 2004		
AUG 0 2 2 2 0 1 4		

11/00 c:/CIRC/DateDue.p65-p.14

DISCRIMINANT VALIDITY OF THE DEVEREUX BEHAVIOR RATING SCALE-SCHOOL FORM WITH SPECIAL EDUCATION INITIAL REFERRALS

By

Barbara Sullivan Dunn

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology,
and Special Education

2000

ABSTRACT

DISCRIMINANT VALIDITY OF THE
DEVEREUX BEHAVIOR RATING SCALE-SCHOOL FORM
WITH SPECIAL EDUCATION INITIAL REFERRALS

By

Barbara Sullivan Dunn

One hundred forty-seven students referred for an initial special education evaluation were rated by general education teachers using the Devereux Behavior Rating Scale-School Form (DSF). The entire sample was divided into four groups based on each student's Individualized Educational Planning Committee eligibility determination: learning disability (LD); emotional impairment (EI); learning disability and emotional impairment (LD/EI); and not eligible (NE). The DSF mean Total Test standard scores earned by each subsample were compared. Post-hoc analyses provided support for reclassifying the subsamples as EI (formerly EI or LD/EI) or Not-EI (formerly LD or NE). At an optimum Total scale cut-off score of slightly less than 1 SD above the mean, the DSF correctly identified 77.6% of the EI subsamples and 78.8% of the Not-EI subsamples. At Subscale cutoff scores of 1 SD above the mean, 91.8% of the EI and LD/EI students, as expected, had at least one elevated Subscale score, and 53.1% of all of the not-EI students, as expected, had no elevated Subscale scores. Implications of these and other findings are discussed.

ACKNOWLEDGMENTS

Many people have contributed to the completion of this dissertation through their assistance and support through the years.

I appreciate all of the staff at the Devereux Foundation, who gave me the opportunity, fresh out of college, to participate in the major undertaking of revising the Devereux Behavior Rating Scales. A special thanks to Paul LeBuffe for his wonderful mentorship and continued support.

Dr. Harvey Clarizio, my program advisor and dissertation committee chairperson, has my sincere appreciation for initially challenging me to consider pursuing a PhD in school psychology, and then providing me with the ongoing encouragement and guidance to help me see that goal through completion.

My other committee members, Dr. Linda Patriarca, Dr. Susan Phillips, Dr. John Paul McKinney, and Dr. Mark Reckase also deserve my thanks. Their time, effort, and input have helped improve this dissertation. Similarly, Dr. Len Bianchi's input on the statistical aspects of this study has been much appreciated.

I would also like to thank all of my colleagues at Eaton Intermediate School District for their personal and professional support. Eaton ISD was a wonderful place to start my career as a practicing school psychologist, and the

help from the school psychologists and school social workers in distributing and collecting the behavior rating scales was invaluable for this project. Connie Mitchell and Annette Gordon also helped with compiling demographic data and other computer-related tasks.

Special thanks to my family, as well. My parents, Paul and Dorothy, provided me with excellent educational opportunities. My husband, Kurt, has always encouraged me to become a "Doc," even though our leisure plans often were interrupted by my work on graduate courses and the dissertation. My young son, Charles, has also given me inspiration to complete my dissertation so I can spend more time with him.

Lastly, thanks to the unknown author of the quote that has sustained me during this endeavor: "The race goes not always to the swift, but to those that keep on running."

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES ix

CHAPTER 1
INTRODUCTION 1
The Problem 1

CHAPTER 2
LITERATURE REVIEW 4
Behavior Rating Scales 4
Advantages 4
Disadvantages 8
Guidelines for Selecting and Using Behavior Rating Scales 11
The Devereux School Form (DSF) 15
Description 15
Existing DSF Validity Studies 16
Limitations of DSF Validity Studies 21
Addressing the Limitations of DSF Validity Studies 28

CHAPTER 3
METHODS AND PROCEDURES 30
General Design of Study 30
Data Collection 30
Dependent Variables 31
Independent Variables 32
Research Instrument 32
Study Population and Sample Selection 35
Characteristics of the Sample 38
Research Questions 40
Data Analysis Procedures 41

CHAPTER 4
RESULTS 46
Questions Related to Four IEPC Eligibility Groups . 46
Questions Related to Gender and SES 52
Questions Related to ADHD-Status 54
Questions Related to Subscale Scores as Predictors . 58
Questions Related to Classification Accuracy 61

CHAPTER 5
DISCUSSION 75
DSF Score Differences for the Four IEPC Eligibility Groups 75
DSF Score Differences for Gender and SES Groups . . 81
DSF Score Differences for the ADHD-Status Groups . . 83

TABLE OF CONTENTS (cont.)

Utility of DSF Subscale Scores as Predictors of EI Status 85
Classification Accuracy of the DSF 86

CHAPTER 6
SUMMARY AND IMPLICATIONS 97
Summary of Results 97
Limitations 101
Sampling Concerns 101
Rater Bias 102
Assessment Agreement 104
Perceptions versus Behavior 105
Diagnostic and Statistical Errors 105
Multiple Gating 107
Unreliability of Criterion Variables 108
Implications 110
Practice and Policy 110
Future Research 116

APPENDICES
A. Devereux Behavior Rating Scale-School Form (DSF) 120
B. The Psychological Corporation Study Permission Letter 121
C. Letter to School Social Workers 122
D. Letter to School Psychologists 123
E. Eaton Intermediate School District Study Permission Letter 124
F. Michigan State Board of Education (1997) Learning Disability Criteria 125
G. Michigan State Board of Education Emotional Impairment Criteria 126

LIST OF REFERENCES 127

LIST OF TABLES

1. Examples of Assessment Techniques Categorized by the Time Data are Obtained and Response Time . . . 3

2. Overview of Broad Parent Rating Scales 5

3. Overview of Broad Teacher Rating Scales 6

4. Types of Error Variance Found with Behavior Rating Scales 10

5. Summary of Validity Studies' Correct Classification Rates with DSF Total Cutoff Score of 1 SD Above the Mean 21

6. DSF Subscale Abbreviated Item Types 34

7. Characteristics of the Four IEPC Eligibility Groups 39

8. Sensitivity, Specificity, and Predictive Value Formulas 43

9. Illustrative Example of Sensitivity, Specificity, and Predictive Value Formulas 45

10. Means and Standard Deviations of the DSF Scores for Four IEPC Eligibility Samples 47

11. Tukey's HSD Significance Levels of Means for IEPC Eligibility Groups in Homogeneous Subsets 47

12. Characteristics of the Two Recombined Samples . . . 49

13. Means and Standard Deviations of the DSF Scores for Two Recombined Samples 50

14. Subscale Intercorrelation Coefficients for All EI and All Not-EI Students 51

15. Means and Standard Deviations of the DSF Scores for Gender and SES 53

16. Means and Standard Deviations of the DSF Scores for Three ADHD-Status Samples 55

17. Tukey's HSD Significance Levels and Means for ADHD-Status Groups In Homogeneous Subsets 55

LIST OF TABLES (cont.)

18. Means and Standard Deviations of the DSF Score for Four EI/ADHD-Status Samples 57

19. Tukey's HSD Significance Levels and Means for EI/ADHD-Status Groups for IBF Subscale Scores . . . 58

20. Tukey's HSD Significance Levels and Means for EI/ADHD-Status Groups for PSF Subscale Scores . . . 58

21. Results of Discriminant Function Analysis 60

22. Discriminant Analysis Classification Results 60

23. Sensitivity and Specificity Rates with DSF Total Cutoff Scores of 0 to 2 SD Above the Mean 62

24. Incorrect Classification Rates with DSF Total Cutoff Scores of 0 to 2 SD Above the Mean 64

25. Sensitivity and Specificity Rates with DSF Subscale Cutoff Scores of 0 to 2 SD Above the Mean 67

26. Incorrect Classification Rates with DSF Subscale Optimum Cutoff Scores 71

27. Predictive Values for DSF Subscale Optimum Cutoff Scores 73

28. Comparison of Abbreviated Items from Selected Depression Subscales and Inventories 79

LIST OF FIGURES

1. Research Design Matrix 32

2. Sensitivity and Specificity Rates with DSF Total Cutoff Scores of 0 to 2 SD Above the Mean 62

3. Sensitivity and Specificity Rates with DSF IP Subscale Cutoff Scores of 0 to 2 SD Above the Mean 67

4. Sensitivity and Specificity Rates with DSF IBF Subscale Cutoff Scores of 0 to 2 SD Above the Mean 68

5. Sensitivity and Specificity Rates with DSF DEP Subscale Cutoff Scores of 0 to 2 SD Above the Mean 68

6. Sensitivity and Specificity Rates with DSF PSF Subscale Cutoff Scores of 0 to 2 SD Above the Mean 69

CHAPTER I

INTRODUCTION

The Problem

Differentiating between students with learning disabilities, emotional impairments, both learning disabilities and emotional impairments, and academic/emotional problems not qualifying for special education continues to be an important but oftentimes difficult task for school psychologists. For instance, within the last decade, researchers have suggested that children with emotional disturbances are underidentified (Forness & Knitzer, 1992). Similarly, many children with learning disabilities have unrecognized and untreated social-emotional problems (Handwerk & Marshall, 1998; Kavale & Forness, 1996; Naglieri & Gottling, 1995), particularly as they relate to depression (Hall & Haws, 1989; Wright-Strawderman & Watson, 1992) and nonverbal learning disabilities (Gross-Tsur, Shalev, Manor, & Amir, 1995; Harnadek & Rourke, 1994; Little, 1993). Many low-achieving students who are not receiving special education services also exhibit social-emotional problems at a level sometimes indistinguishable from their peers in special education (Bursuck, 1989; Haager & Vaughn, 1995; Tur-Kaspa & Bryan, 1995; Vaughn & Haager, 1994).

When assessing academic and social-emotional-behavioral problems, an assessment model which consists of multiple

sources of assessment data, obtained from multiple sources and in multiple settings, is generally considered best practice (Merrell, 1994). School psychologists use a variety of diagnostic tools to assist them in making these often difficult differential diagnoses within such a multi-method, multi-informant, multi-setting model. These tools include cognitive, achievement, and personality testing with the referred student; interviews with the student, teacher and parent for their perceptions of the student's difficulties; direct observations of the child's behavior, which yield a sample of the frequency and duration of specific behaviors; and behavior rating scales, for which parents and/or teachers are asked to reflect upon a certain time period (e.g., the last four weeks) and report how often the child has demonstrated various behaviors. These data-gathering techniques can be conceptualized as falling into one of four categories (see Table 1), based upon two dimensions: 1) whether the data-gathering occurs at the same time the behavior occurs ("simultaneous") <u>or</u> at a later time ("retrospective"); and 2) whether the questions posed will result in narrative and anecdotal answers ("open-ended") <u>or</u> yes/no or Likert-style responses ("closed ended").

Table 1.

Examples of Assessment Techniques Categorized by the Time Data are Obtained and Response Type

	OPEN-ENDED	**CLOSED-ENDED**
SIMULTANEOUS	projective personality tests	frequency counts of specified behaviors obtained during classroom observations by an evaluator
RETROSPECTIVE	teacher interviews	behavior rating scales

CHAPTER II

LITERATURE REVIEW

Behavior Rating Scales

For the last 15 years, school psychologists have increasingly included behavior rating scales as part of their assessment battery (Clarizio & Higgins, 1989; Hutton, Dubes, & Muir, 1992). A number of broadly focused parent and teacher behavior rating scales have been used to obtain reports of students' behaviors and emotions (Kamphaus & Frick, 1996), a sampling of which are summarized in Table 2 and Table 3, respectively. As shown, there are parent and teacher report versions for each of the following measures: the Devereux Behavior Rating Scale-School Form (DSF; Naglieri, LeBuffe, & Pfeiffer, 1993), the Behavioral Assessment for Children (BASC; Reynolds & Kamphaus, 1992), the Child Behavior Checklist (CBCL; Achenbach, 1991), and Conners Rating Scale (Conners, 1990). Most scales assess both externalizing and internalizing behaviors, and some also assess adaptive behavior. Ages of children covered typically range from 4 to 18 years. Reliability and validity information vary by scale.

Advantages

When compared to other data-gathering techniques, there are many features of behavior rating scales which are appealing to evaluators. First and foremost, behavior rating scales give teachers and parents an opportunity to be

Table 2.

Overview of Broad Parent Rating Scales

Scale	Authors	Publisher	Content	Ages	Reliability	Validity
Devereux Behavior Rating Scale-School Form (DSF)	Naglieri, LeBuffe, & Pfeiffer (1993)	The Psychological Corporation	Interpersonal Problems, Inappropriate Behavior/Feelings, Depression, and Physical Symptoms/Fears (40 items)	5-18 years	Internal consistency, test-retest (. 70s-.90s) and interrater and intrarater (all p < .01)	
Behavioral Assessment for Children, Parent Rating Scale (BASC-PRS)	Reynolds & Kamphaus (1992)	American Guidance Service	Clinical Maladjustment, School Maladjustment, Personal Adjustment, Emotional Symptoms Index composites, 10 clinical scales, 4 adaptive scales, 5 validity indexes (152-186 items)	4-18 years	Internal consistency , test-retest and interrater (.70s to .90s)	Support for factor structure, modest to high correlations with existing scales, lawful profiles for clinical samples
Child Behavior Checklist (CBCL)	Achenbach (1991)	Author, University of Vermont	Internalizing, Externalizing, Total T, Competence, 8 clinical scales (113 items)	4-18 years	Internal consistency , test-retest (.70s -.90s), and interrater	Support for factor structure, differential validity (nondisabled vs. clinical groups)
Conner's Parent Rating Scale (CPRS)	Conners (1990)	Multi-Health Systems, Inc.	CPRS-93, 9 clinical scales (93 items) CPRS-48, 5 clinical scales (48 items)	3-17 years	Poorly documented	Scales based on factor analysis and Hyperactivity Index of both scales have consistently been proven to be sensitive to treatment effects

Adapted from Kamphaus & Frick (1996, p. 148)

Table 3.

Overview of Broad Teacher Rating Scales

Scale	Authors	Publisher	Content	Ages	Reliability	Validity
Devereux Behavior Rating Scale-School Form (DSF)	Naglieri, LeBuffe, & Pfeiffer (1993)	The Psychological Corporation	Interpersonal Problems, Inappropriate Behavior/Feelings, Depression, and Physical Symptoms/Fears (40 items)	5-18 years	Internal consistency, test-retest (. 70s-.90s) and interrater and intrarater (all p < .01)	Support for factor structure, differential validity (special education/clinical vs. general education/control groups)
Behavioral Assessment for Children, Teacher Rating Scale (BASC-TRS)	Reynolds & Kamphaus (1992)	American Guidance Service	Externalizing, Internalizing, Learning, Attention, and Adaptive (109-148 items)	4-18 years	Internal consistency (m = .80), test-retest (m = .89), and interrater (m = .71)	Support for factor structure and high correlations with existing scales
Child Behavior Checklist-91-Teacher Rating Scale (CBCL-91-TRS)	Achenbach (1991)	Author, University of Vermont	Internalizing, Externalizing, Social, and Attention (113 items)	4-18 years	Internal consistency (m = .87), test-retest (m = .92), and interrater (m = .55)	Support for factor structure and extensive research on correlations with important criteria
Conner's Teacher Rating Scale - 39 (CTRS-39)	Conners (1990)	Multi-Health Systems, Inc.	Attention, Hyperactivity, Conduct, and Anxiety (39 items)	4-12 years	Internal consistency (median = .84) and test-retest (.72-.91)	Scales based on factor analysis; Hyperactivity Index of both scales has consistently been proven to be sensitive to treatment effects
Conner's Teacher Rating Scale - 28 (CTRS-28)			Conduct, Hyperactivity, and Attention (28 items)	3-17 years	no reliability data	

involved in the evaluation process (Elliott, Busse, & Gresham, 1993) by capitalizing on their judgments and observations of what occurs over a period of time in the child's natural environment, namely school and home (Martin, Hooper, & Snow, 1986; McMahon, 1984). Further, behavior rating scales are inexpensive with regard to professional time and training requirements (Merrell, 1994), and parents and teachers are able to elaborate their referral concerns in a time-efficient, easy to understand manner. Indeed, behavior rating scales can streamline the often lengthy evaluation process by complementing the information gathered by other means. For example, the evaluator can first sample a broad range of potentially relevant behavior issues with a teacher- and/or parent-completed behavior rating scale (McConaughy, 1993), and then pursue selected issues during a follow-up interview. Likewise, behavioral frequency and intensity data obtained by the evaluator during a brief direct classroom observation (often 30 minutes or less) can be compared to the teacher's ratings of similar data based on a longer time period (often four weeks or more) in the search for quantitative distinctions regarding qualitative aspects of students' behaviors (Barkley, 1993) and to make sure infrequent but important behaviors, such as violent and assaultive behaviors, are not overlooked (Sattler, 1988). Much like standardized cognitive and academic testing, behavior rating scales often have normative data to aid in the process of determining the statistical deviance of

children's behavior. Also, the assessment of comorbid conditions (which are common for students with behavioral problems) is facilitated by the aggregation of items into empirically derived scales (McConaughy, 1993), as well as by the fact that a broad range of problems are included. As Merrell (1994, p. 68) concludes, with all of these advantages, it is no wonder that behavior rating scales are so popular--they help capture the "`big picture' of the assessment problem" in short order, with limited expense, and with considerable face and clinical validity.

Disadvantages

Despite all the potential advantages, the limitations and disadvantages of behavior rating scales are numerous. Behavior rating scales are often of limited use in the functional analysis and management of behavior problems, due to the scales' insensitivity to antecedents and consequents of behavior (Barkley, 1993; McConaughy, 1993). Moreover, because rating scales reflect the parent's or teacher's perceptions of problems, direct observations and clinical interviews are necessary adjuncts to clarify the behavioral issues at hand (McConaughy, 1993). Also, some relatively rare conditions (e.g., autism) are often not included in the broad-band scales. Further, rating scales also typically use rather simple Likert-style frequency descriptors (e.g., *not at all, just a little, pretty much, very much*), but do not define the exact frequency, intensity, and duration of behaviors that correspond with these descriptors (Elliot et

al., 1993; Reid & Maag, 1994). Raters often differ significantly in their interpretation of the amount of a behavior that corresponds to each of these frequency ratings. Some of the factors impacting raters' interpretation of the Likert scales include their tolerance for disruptive/deviant behavior, differing views of what types of behaviors require intervention, and raters' own perceptions of self-competence, quality and availability of assistance, and their difficulty managing a student (Reid & Maag, 1994). Merrell (1994, p. 69) discusses other, more specific, types of response bias which include *halo effects* (endorsing primarily positive ratings for a student due to some of his/her positive traits that are not relevant to the rated item), *leniency or severity* (a rater's tendency to endorse in an overly generous or critical manner for all students to be rated), and *central tendency effects* (a rater's tendency to avoid end points such as "never" or "always" and to select midpoints instead). Consequently, different raters could observe the same behavior and yet produce ratings that differ with regard to the presence/absence and frequency/intensity, and the difference between ratings would be due to the raters' characteristics rather than the child's actual behavior.

On a related note, Martin, Hooper, and Snow (1986) discuss four types of error variance that can affect the results obtained through a rating scale assessment, as summarized in Table 4. *Source variance* refers to the

Table 4.

Types of Error Variance Found With Behavior Rating Scales

Type of Error Variance	Examples
Source Variance	Similar to *response bias*, in that different raters may have different ways of responding to the rating format
Setting Variance	Related to *situational specificity* of behavior; the eliciting and reinforcing variables present in one environment (e.g., reading class with Ms. Jones) may not be present in a closely related environment (e.g., math class with Ms. Smith)
Temporal Variance	Behavior is likely to change over time, and a rater's approach to the rating scale task may also change over time
Instrument Variance	Different rating scales may be measuring different hypothetical constructs; there is a range of continuity (from close to disparate) between constructs measured by different scales

Adapted from Merrell (1994, p. 69).

rater's subjectivity and his/her unique way of responding to the rating scale. *Setting variance* is due to the situational specificity of behavior, which reflects the fact that humans behave differently in various situations because of the different types of eliciting and reinforcing variables present in each. *Temporal variance* takes into account both the tendency for students' behavior *and* raters' approaches to the rating scale task to change over time. Fourth, *instrument variance* reflects the wide range of different hypothetical constructs assessed by different rating scales. In sum, despite the deceiving appearance of simplicity for most behavior rating scales, these

instruments are susceptible to many sources of error, and thus should meet the same standards of reliability and validity as must standardized measures of cognitive and academic skills (Elliot et al., 1993).

Guidelines for Selecting and Using Behavior Rating Scales

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1996), several criteria should be considered by school psychologists when selecting a behavior rating scale for use in special education assessments. These criteria are the same as are applied to other standardized measures. Specifically, evidence of an appropriate standardization sample, reliability, and validity should be included in the scale manual. Furthermore, validity research should be ongoing, particularly with new scales.

Edelbrock's (1983) discussion of considerations regarding behavior rating scales and the raters themselves provides helpful guidelines when selecting a particular instrument. With regard to the instrument, one must first consider which behavioral characteristics the instrument purports to assess. Different domains (e.g., personality, maladjustment, social-emotional functioning, problem behavior) are tapped by different instruments. Moreover, even scales designed to assess "problem behavior," for instance, often vary considerably by focusing on behaviors

as diverse as peer relationships, playground behavior, or more specific syndromes such as depression, anxiety, or hyperactivity.

Because rating scales are used for a myriad of purposes (e.g., screening, identification, clinical diagnosis, school placement, treatment evaluation), Edelbrock (1983, p. 294) cautions, "When selecting a rating scale, it is essential to judge whether the specific target phenomena are appropriate to the application." Of course, individual behavior rating scales are not equally suited to all applications, and thus the evaluation of each instrument rests largely on the intended application.

On a related note, Merrell (1994) concluded that after considering the characteristics, advantages, and disadvantages of behavior rating scales, they are usually best used in the types of decision making associated with screening and additional assessment. His conclusion was in the context of an assessment model known as "multiple gating," in which "a large population is sequentially narrowed down to a small population of individuals who are likely to exhibit the behavioral syndromes in question across settings and over time ... through a series of assessment and decision steps (gates)" (p. 37).

Edelbrock (1983) also elaborated on many of the technical considerations to review when selecting a behavior rating scale. First, one must evaluate the degree to which

individual items, and the entire item pool, reflects the target phenomena being assessed. Common problems with item selection include items that do not directly assess the child's behavior (e.g., *parents are divorced)*, or that tap the consequences of the behavior rather than the behavior itself (e.g., *student has been suspended)*. Next, one must assess which level of behavioral analysis (global or more specific) the items reflect, and which level of analysis is best for the type of assessment being undertaken and the raters involved. Third, Edelbrock considers standardization to be absolutely necessary to the accurate interpretation of behavior rating scales. He added that the appropriate level of stratification of norms (e.g., based on gender, age, race, socioeconomic status, region of the country, etc.) depends on the extent to which these variables account for the variance in scores.

As discussed earlier, ratings are not only a function of the child's behavior and the assessment device, but also the informant. Thus, when selecting a behavior rating scale, there are several considerations regarding who the informant is, using more than one informant, and the influence of informant characteristics (Edelbrock, 1983). At a bare minimum, a rating scale should specify who the informant should be, above and beyond the general direction of "someone familiar with the child's behavior." This is because informants typically differ in the types of behaviors they are most qualified to report. For instance,

teachers tend to be best qualified to rate items addressing classroom behavior and peer interactions, whereas parents tend to be better at reporting behaviors most often seen at home (e.g., enuresis, sibling rivalry, somatic complaints) and rare but clinically significant behaviors (e.g., running away, suicidal behavior). Research also indicates that the frequency and patterning of child behaviors vary as a function of the informant, and consequently, different scales for parents and teachers are typically needed. Even then, it is difficult to determine which parent (e.g., mother, father, or step-parent) or teacher (e.g., reading, math, or gym) is the best informant.

Multiple informants are often used when assessing children's behavior, in part because of the attributes and limitations of each type of informant, but also because this permits a more thorough portrayal of a child's behavior across settings. Edelbrock (1983) notes that the typically low rates of inter-rater agreement for global, broad-band rating scales do not necessarily mean that data from any one rater is invalid or unreliable. Instead, different ratings from parents and teachers may just as likely reflect the situational specificity of behavior for individual children. Thus, high levels of agreement among informants (or evidence of high levels of inter-rater reliability in the scale manual) usually is not a crucial consideration when selecting a behavior rating scale.

The Devereux School Form (DSF)

Description

Some of the better known behavior rating scales frequently used in schools for diagnostic purposes which appear to meet psychometric standards are those developed by the Devereux Institute of Clinical Training (Sattler, 1988). Devereux's most recently developed behavior rating scale for use in the schools is the Devereux Behavior Rating Scale-School Form (hereafter called the DSF; Naglieri, LeBuffe, & Pfeiffer, 1993; see Appendices A and B). The DSF is a revision of the widely used Devereux Child Behavior Rating Scale (Spivak & Spotts, 1966) and the Devereux Adolescent Behavior Rating Scale (Spivak, Spotts, & Haimes, 1967). This particular behavior rating scale was chosen for study over some of its competitors (e.g., Achenbach and Edelbrock's *Teacher Report Form*, 1986; Reynolds and Kamphaus' *Behavior Assessment System for Children*, 1992) for two main reasons: 1) its item pool and scoring format, which are based on the four subcategories of the federal definition of emotional impairment (EI), are unique and potentially attractive to school psychology practitioners; and 2) it is currently being used in many schools to aid in differential diagnosis for initial special education referrals, but no research studies validating it for that purpose have been conducted.

According to the scale's authors, the DSF "was developed to provide the professional a structured system of

determining the extent to which a child's or adolescent's behaviors fall within or outside the normal range ... and to formalize the assessment of a set of behaviors that are indicative of moderate or severe emotional difficulties" (Naglieri, LeBuffe, & Pfeiffer, 1993, p. 2). As part of a comprehensive evaluation, the DSF was developed as a screening measure to identify children who need further evaluation for special education services due to a suspected emotional impairment. The items selected for inclusion in the DSF are intended to measure the four areas of problem behaviors (i.e., interpersonal problems, inappropriate behaviors/feelings, depression, and physical symptoms/fears) that are cited in the federal definition of severe emotional disturbance/impairment, per the Education for All Handicapped Children Act of 1975 (PL 94-142) and the Individual with Disabilities Act of 1990 (PL 101-476). In a recent review, Goh (1995, p. 329) concluded that the DSF manual "provides sufficient information on development, standardization, administration, scoring, interpretation, reliability, and validity" and that this particular scale, unlike many of its competitors, can be considered a "systematic and psychometrically sound behavioral instrument."

Existing DSF Validity Studies

Not many validity studies exist on the DSF. Six criterion-related validity studies are cited in the manual, two of which have since been published in refereed journals

(Naglieri, Bardos, & LeBuffe, 1995; Naglieri & Gottling, 1995). More recently, Goh (1997) investigated the validity of the DSF with culturally diverse samples. Each of these seven studies examined the extent to which the DSF discriminates between general education students and special education students/clinical patients. Most present results for the two age groups (i.e., 5-12 years and 13-18 years), but none have analyzed data by other potentially important variables such as socioeconomic status and gender (perhaps because separate male/female scoring norms are provided in the manual). In each discussion of the validity studies, a Total score cutoff of 1 SD above the mean was used for classification purposes.

One of the published DSF studies (Naglieri, Bardos, & LeBuffe, 1995) examined the extent to which the scale has adequate discriminant validity for differentiating special education students identified as having serious emotional disturbance from nonreferred students in general education. A sample of general education students rated by their classroom teachers was compared to a sample of emotionally impaired students rated by their special education teachers. In the 5-12 year old group (roughly 4:1 males:females), 87% of the general education sample and 63% of the special education sample was classified accurately. In the 13-18 year old group, 93% of the general education sample (3:1 males:females) was classified accurately, compared to only 47% of the special education sample (4:1 males:females).

The second published DSF validity study (Naglieri & Gottling, 1995) compared the scores obtained for dually diagnosed students (with both learning disabilities and emotional impairments: LD/EI) and a matched control sample of general education students on both the DSF and The Teacher Report Form (Achenbach & Edelbrock, 1986). Students (roughly 3:1 male:female ratio) ranged in age from 7 years to 16 years, 7 months, but results were not stratified for the two age groups (i.e., 5-2 years and 13-18 years). The DSF was successful in accurately identifying 90% of the dually diagnosed group, and 96% of the control group, using a Total Scale cutoff score of 1 SD above the mean.

One of the remaining four validity studies published in the scale manual compared teacher ratings for emotionally impaired students in special education with teacher ratings for general education students. In the 5-12 year old group (roughly 4:1 males:females), 97% of the general education students and 53% of the special education students were classified accurately. In the 13-18 year old group (also roughly 4:1 males:females), 90% of the general education sample was classified accurately, compared to only 48% of the special education sample (4:1 males:females).

In another validity study discussed in the manual, ratings for students in residential treatment and special education were compared with ratings for children involved in public schools, sports leagues, and recreational clubs. The DSF correctly identified 92% of the 5-12 year old

students in the control sample (nearly 1:1 males:females), and 54% of the 5-12 year old students in the clinical sample (3:1 males:females). For the 13-18 year old group (roughly 1:1 males:females), the DSF correctly identified 92% of the control sample and 49% of the clinical sample.

In the fifth validity study, the clinical sample was composed of youth with severe emotional disturbances who were receiving treatment in psychiatric hospitals. The control sample consisted of general education students. There was considerable variation in classification accuracy rates (using a Total Scale cutoff score of 1 SD above the mean) depending on the age of the students. In the 5 to 12 year old group,93% of the general education sample (nearly 1:1 males:females) and 55% of the clinical sample (3:1 males:females) were correctly classified. In contrast, for the 13-18 year old group (roughly 3:2 males:females), 86% of the general education sample and 93% of the clinical sample were correctly classified.

The sixth validity study compared youth in residential treatment centers for severe emotional disturbances with general education students. In the 5-12 year old group, 83% of the control sample (3:1 males:females) and 76% of the clinical sample (4:1 males:females) were correctly classified. In the 13-18 year old group, 100% of the control group (1:1 males:females) and 65% of the clinical sample (7:3 males:females) were correctly classified.

The seventh and final validity study (Goh, 1997) is the only one conducted by a researcher other than those involved in the development of the new DSF. This study compared elementary aged general education students with same-aged peers in special education due to emotional disturbances. The sample(roughly 2.5:1 males: females) consisted of a culturally diverse group of Caucasian, African American and Hispanic children ages 5 to 12 years. Different optimum cutoff scores were found for the Caucasian (115), African American (114) and Hispanic (113) groups, with uniformly high accurate classification rates for both the general education and special education samples (about 80%).

A summary of the seven validity investigations is presented in Table 5. For age level 5-12 years, classification rates ranged from 83 to 97% for the control samples, and from 53% to 75% for the clinical samples. Somewhat more variation was seen for the age level 13-18 years, with 86 to 100% accurate classification rates for the control samples, and 47 to 93% accurate classification rates for the clinical samples. The authors of the DSF Scale Manual (Naglieri, LeBuffe, & Pfeiffer, 1993, p. 66) concluded, "These percentages compare quite favorably with those reported for other widely used behavior rating scales (e.g., Achenbach, 1991).

Table 5.

Summary of Validity Studies' Correct Classification Rates with DSF Total Cutoff Score of 1 SD Above the Mean

Validity Study	5-12 Years		13-18 Years		6-16 Years	
	Control	Clinical	Control	Clinical	Control	Clinical
1 (Naglieri, Bardos, & LeBuffe, 1994)	87%	63%	93%	47%	n/a	
2 (Naglieri & Gottling, 1995)	n/a		n/a		96%	90%
3 (DSF Manual, 1993)	97%	53%	90%	48%	n/a	
4 (DSF Manual, 1993)	92%	54%	92%	49%	n/a	
5 (DSF Manual, 1993)	93%	55%	86%	93%	n/a	
6 (DSF Manual, 1993)	83%	75%	100%	65%	n/a	
7 (Goh, 1997)			n/a		n/a	
Total Sample	80%	81%				
Caucasians	84%	84%				
African-Americans	80%	78%				
Hispanics	82%	73%				

Note: n/a means data are not available.

Limitations of DSF Validity Studies

Several methodological and substantive limitations are present in the above validity studies. For instance, these validity studies compared students with severe emotional impairments with non-disabled students. The DSF appears to

be most successful at correctly classifying control group students (i.e., those in general education), which should be a relatively easy discriminatory task. As Elliot et al. (1993, p. 318) state, "The question in making a diagnosis is typically *not* whether a group of referred children's mean behavior ratings differ significantly from a group of nonreferred children's mean ratings." A more difficult discrimination, and one which school psychologists are typically called upon to make, involves comparing not-yet-identified emotionally impaired students with students who have learning and behavioral problems due to other (or co-existing) not-yet-identified "clinical" conditions (e.g., learning disabilities, attention deficit hyperactivity disorder--ADHD, social maladjustment, borderline or low average cognitive skills, etc.).

Indeed, the validity studies' absence of any mention of the impact ADHD may have on the DSF scores is noteworthy, given the characteristic overlapping learning and behavioral problems in children with ADHD. In their review of concomitance, Smith, Dowdy, Polloway, & Blalock (1997) suggest that from 25 to 80% of all students with ADHD experience academic difficulty, and perhaps 10-20% of them may qualify under the learning disability label. At the same time, an estimated 30% of students with learning disabilities may also have ADHD. Rock, Fessler, and Church (1997) concluded that a subgroup of students with ADHD present a similar profile to those students with both

learning disabilities and social-emotional-behavioral impairments. Further, a recent study (Bussing, Zima, Belin, & Forness, 1998) provided additional evidence for the impact of ADHD on special education service delivery, as the researchers found that students within both LD and SED programs who met diagnostic criteria for ADHD generally had more severe impairments than children who met only initial screening criteria for ADHD.

Next, the DSF scale's authors suggested that the sensitivity of the scale may have been underestimated by these studies, in part because the students with emotional impairments were "being assessed while they were in an intervention program, not at the point of referral when abnormal behaviors may be more prevalent. If this is the case, the DSF would likely show higher sensitivity when used at referral" (Naglieri, Bardos, & LeBuffe, 1995, p. 109). Similarly, one might argue that special educators may not perceive children's behavioral problems to be quite as deviant as would a general education teacher, in part because the classroom norms differ (i.e., a special education class with many students with emotional and behavioral problems vs. a general education class with only one student with emotional and behavioral problems). Conversely, because the ratings completed by the special education teachers were not done blind (i.e., they knew the students had emotional impairments), one might reasonably predict that their ratings would have reflected the

students' emotional difficulties. Moreover, since emotionally impaired students tend to have poor outcomes even with special education intervention (Rock et al., 1997), the argument that the ratings would be more discriminating at the point of referral is questionable. The issue of discriminant validity during the referral period is very important, however, because this is when the scale would most often be used in the schools.

The inclusion of 13-18 year old students in many of the validity studies may also contribute to methodological limitations. For instance, whereas children in developmental kindergarten through grade five typically have one main teacher who teaches them several subjects, older children typically have a different teacher for each of five or more subject matters. The variation of these older students' scores that would be contributed by the hour of the day, subject matter taught, personality of each teacher, etc. would make it difficult to interpret the scores. It also becomes more difficult to disentangle learning from emotional problems when children are in middle and high school (e.g., did five or more years of learning difficulties lead to these behavioral problems, or vice versa?). Note, too, that in the Midwestern school districts under study for this investigation, approximately two-thirds of students referred for an initial special education evaluation were in the 5 to 12 year old age bracket, which provides some justification for targeting this age group.

Other methodological limitations of the existing validity studies involve the selection of a Total cutoff score of only 1 SD above the mean. First, the use of the Total scale score is suspect, as a significant impairment in only one subcategory of emotional impairment (i.e., interpersonal problems, inappropriate behaviors/feelings, depression, or physical symptoms/fears) is needed for eligibility. Thus, data on the validity of the Subscale scores is critically needed. Second, the Devereux-School Form authors justify the 1 SD above the mean cutoff score by referring to the above validity studies, as that was the point at which the percentage of true positives and true negatives was maximized. However, since several of the studies had correct classification rates of approximately 50% for the clinical samples, which is basically the same as a coin-toss, the practical utility of the scale is called into question. Further, this cutoff score translates to an identification rate of 16% of the population as having a severe emotional impairment, which is high by any standards used in the schools and norms reflected in some epidemiological studies (Handwerk & Marshall, 1998). Even cutoff scores of 2 SD above the mean would identify roughly 2% of the general population as EI, a figure twice that yielded by current practices. For comparison's sake, just over 1% of students from this study's school districts' total school population have been labeled "emotionally impaired," which corresponds to roughly 8% of all special

education students from these same school districts (1998 data). On the other hand, these data differ from some national epidemiological studies (Costello & Angelo, 1995; Handwerk & Marshall, 1998) which suggest that while fewer than 2% of children receive mental health treatment at any point in time, approximately 7 to 10% of children have a psychiatric problem that significantly impairs functioning. Costello & Angelo (1995, p. 29) make the point that "receipt of services is a poor criterion for defining need for services." These latter prevalence estimates suggest that a 1 ½ SD cutoff score may be more appropriate.

On a related note, another limitation of the data presented in the previously discussed validity studies involves the term "base rates." According to Glutting, McDermotte, Watkins, and Konold (1997, p. 177), "base rates refer to the frequency, or percentage, of a population that falls within a particular diagnostic category." As the epidemiological studies cited above indicate, the base rate for psychiatric normality is at least 90%. Consequently, while at first glance the DSF validity studies' consistently correct classification of roughly 90% of the control samples may seem impressive, it actually may offer no advantage to identifying students as "normal" over simply knowing the base rate.

The limited discussion of diagnostic misses, and the implications thereof, is also a limitation of the validity studies to date. Although the data are available (e.g., in

Validity study 1, if 87% of the control group was correctly classified, that means 13% of the control group was incorrectly classified), the implications of such incorrect classifications within the general and special education samples does not appear to have been given the attention it deserves.

Another issue that received scant consideration in the validity studies to date is that of multiple gating in assessment. As mentioned previously, multiple gating refers to a sequential assessment process wherein at each decision making point (or gate), the number of target students keeps getting smaller until only those with the most significant problems are identified. If the DSF were to be used as a screening instrument with the control groups, then it may be reasonable to find 10-20% of that population (as shown in several of the validity studies' control groups) as potentially having significant social-emotional-behavioral problems that are worthy of further assessment.

Also, the existing validity studies do not analyze how cutoff scores and classification accuracy may differ for students on the basis of important variables such as gender and socioeconomic status (SES). For instance, at present there is a large discrepancy in the male to female ratio for students with the learning disability (LD) eligibility label. This discrepancy is also reflected in most of the DSF validity study samples. However, Anderson (1997) argues that gender bias among referring agents (which potentially

would be reflected in the ways referring agents rate the DSF), is a major factor in the unequal gender distribution in learning disability programs. The impact of SES on DSF scores is also important, given the impact of poverty on emotional and behavioral disorders, as well as learning disabilities (Mamlin & Harris, 1998).

Lastly, the use of educational placement (i.e., general education class vs. special education class) and/or diagnostic label (i.e., no label vs. emotionally impaired) as the criterion variables for the validity studies to date is potentially problematic. As is true for all criterion-related validity studies, the reliability of the criterion variable must be considered (Crocker & Algina, 1986). Many authors (e.g., Macht, 1998; Ysseldyke & Algozzine, 1990) have argued compellingly that special education classification and classroom placements are determined arbitrary and inconsistently. Thus, the unreliability of the criterion variables may have been as big of a problem as any unreliability in the DSF itself.

Addressing the Limitations of DSF Validity Studies

The present study attempts to address the methodological and substantive limitations of the existing Devereux-School Form validity studies in six important ways: 1) in order to assess the scale's utility in making differential diagnoses, the children being included in this study will be those in general education experiencing enough academic and/or behavioral problems to have warranted a

special education referral, rather than comparing "normal" children with those already receiving services due to an identified emotional impairment; 2) in order to reflect how the scale is typically used in the schools, ratings will be completed by general education teachers during the initial special education evaluation process, rather than by special education teachers after a student has already been identified as having an emotional impairment; 3) differences in scores based on gender and SES will be investigated; 4) the effect of a students' ADHD status (i.e., ADHD is diagnosed, suspected, or nonexistent) on DSF scores will be investigated; 5) only children aged 5-12 years will be studied; and 6) in addition to the sensitivity and specificity data reported for various Total scale cutoff scores, similar data will be reported for the four subscales, and discriminant analyses will be presented.

CHAPTER III

METHODS AND PROCEDURES

General Design of Study

Data Collection

Eight school social workers and six school psychologists from a Midwestern Intermediate School District (ISD) were asked to distribute a Devereux-School Form to one general education teacher for each student referred to them for an initial special education evaluation due to a suspected learning disability and/or emotional impairment during the 1994 through 1997 school years (see Appendices C and D). Parental consent was obtained for the initial special education evaluation. This ISD's policy allows for limited disclosure of student records and reports without additional parental consent for studies such as this which are conducted for the ISD for the purposes of validating evaluation instruments (see Appendix E). For students referred to both a school social worker and a school psychologist, the school social worker distributed the rating scales. The general education teachers, who were unaware of the purpose of this investigation, rated the students using the Devereux-School Form, and then returned the completed scale to the school social worker or school psychologist conducting the evaluation. The evaluator then gave the teacher-rated scale to this researcher, who made a copy of the scale, substituted the student's referral number

and a code number for the student's name on the copy, and then returned the original to the evaluator. After Multidisciplinary Evaluation Team (MET) and IEPC meetings were held for students for whom a Devereux-School Form had been completed, the researcher obtained the following information from the central files for each of the students: MET evaluators' initials and title; MET and IEPC eligibility recommendations; Full Scale, Verbal, and Nonverbal IQ's; standardized achievement test scores; Medicaid eligibility (yes/no); and ADHD status (diagnosed, suspected, or nonexistent). Once these data were collected, the researcher deleted the referral number and thereafter used the code number for data analysis purposes. Thus, no one was able to associate responses or other data with individual subjects during the data analysis phase.

Dependent Variables

The main dependent variable used in this study is the IEPC eligibility determination. The four categories of IEPC eligibility are learning disability (LD), emotional impairment (EI), learning disability and emotional impairment (LD/EI), and referred but not eligible (NE) for special education services (see Appendices F and G). A secondary dependent variable is the ADHD status, which is split into three categories: "diagnosed" (a diagnosis of ADHD was noted on the IEPC), "suspected" (the IEPC indicated that an ADHD evaluation was recommended or in progress), and

DEPENDENT VARIABLES

INDEPENDENT VARIABLES	IEPC Eligibility				ADHD Status		
	LD	EI	LD/EI	NE	Diagnosed	Suspected	Nonexistent
DSF Total Score							
IP Subscale Score							
IBF Subscale Score							
DEP Subscale Score							
PSF Subscale Score							

Figure 1. Research Design Matrix

"nonexistent" (there was no mention of an ADHD diagnosis or evaluation on the IEPC).

Independent Variables

The first independent variable is the DSF Total Score, which is a continuous variable recorded in linear standard scores (mean of 100 and SD of 15). The remaining independent variables represent each of the four DSF subscale scores (Interpersonal Problems, Inappropriate Behaviors/Feeling, Depression, and Physical Symptoms/Fears), which are also continuous variables recorded in linear standard scores (mean of 10 and SD of 3). The relationship between the IEPC eligibility determination and the DSF Total and Subscale scores was examined.

Research Instrument

The DSF (Naglieri, LeBuffe, & Pfeiffer, 1993) is a screening measure to identify children who need further

evaluation for a suspected emotional impairment (Goh, 1995). It consists of 40 items rated on a scale of 0 to 4 (0=never, 1=rarely, 2=occasionally, 3=frequently, and 4=very frequently). Each item begins with the stem, "During the past four weeks, how often did the child ... "The authors selected the items based on a review of: the original Devereux scales (Spivak & Spotts, 1966; Spivak et al., 1967); the diagnostic criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition-Revised* (DSM-III-R; American Psychiatric Association, 1987) as well as revisions proposed for the DSM-IV (American Psychiatric Association, 1994); other behavior rating scales; and the literature on behaviors in children that indicate social-emotional-behavioral problems. The items were grouped into four 10-item subscales representing each of the four dimensions of the federal definition: Interpersonal Problems, Inappropriate Behaviors/Feelings, Depression, and Physical Symptoms/Fears (see Table 6). The scale authors organized the items into subscales conceptually and empirically (e.g., item discrimination between the clinical and regular education students, and item-total raw score correlations). Raw scores for each of the subscales yield standard scores (mean=10, SD=3). The sum of the 40 items is used to obtain a Total Test standard score (mean=100, SD=15). Different norms tables are provided on the basis of age level (5-12 years and 13-18

Table 6.

DSF Subscales Abbreviated Item Types

DSF Subscale	Abbreviated Item Types
Interpersonal Problems	*difficulty making/maintaining friends (2 items) *verbal/physical aggression (2 items) *lacking awareness/concern how others feel towards him/her (2 items) *feeling disliked (1 item) *bothers others (1 item) *uneasy around others (1 item) *manipulative (1 item)
Inappropriate Behaviors/Feelings	*inappropriately expresses anger, including physical aggression (3 items) *noncompliant, with and without subsequent emotional upset (2 items) *problems while playing or working (2 items) *property damage (1 item) *lack of regret (1 item) *incites others into retaliating (1 item)
Depression	*diminished interest, pleasure, pride (3 items) *sadness (2 items) *flat affect (2 items) *worthlessness (1 item) *crying (1 item) *isolates self (1 item)
Physical Fears/Symptoms	*unusual reaction to sensory stimuli (4 items) *feels or fears peer rejection (2 items) *resists change (1 item) *school attendance refusal (1 item) *complains of aches/pains (1 item) *irrational fears (1 item)

years), informant (parent and teacher), and student gender (male and female).

The Devereux-School Form was standardized in 1991 at more than 30 sites across the United States. The standardization sample consisted of 3,153 children and adolescents age 5-18 years, and was representative of the

U.S. population in terms of age, sex, geographic region, race, socioeconomic status, and Hispanic origin according to 1990 census data. Ratings were obtained from parents (60%) and teachers (40%) for students in general education and special education (excluding those in classes for students with emotional impairments). The internal reliability coefficients for the 40 item scale for students aged 5-12 years and 13-18 years, respectively, were: parent raters of males .94 and .93, parents raters of females .93 and .92, teacher raters of males .97 and .95, and teacher raters of females .97 and .96. The median internal reliabilities (across ages) for the four subscales were .83 (Interpersonal Relationships), .82 (Inappropriate Behaviors/Feelings), .82 (Depression), and .79 (Physical Symptoms/Fears). As noted earlier, in a recent critical review of the DSF, Goh (1995, p. 329) concluded that the DSF manual "provides sufficient information on development, standardization, administration, scoring, interpretation, reliability, and validity" and that this particular scale, unlike many of its competitors, can be considered a "systematic and psychometrically sound behavioral instrument."

Study Population and Sample Selection

A total sample of 147 (97 male and 50 female) students was included in this study. Students' ages ranged from 5 to 12 years (M=8.83, SD=1.52 years), and grades attended included kindergarten through fifth (M=2.76, SD=1.59 grade). These students attended their neighborhood public elementary

schools within a single rural/suburban intermediate school district from the Midwestern region of the United States. The sample consists of a small portion (approximately 9%) of the general education students referred for an initial special education evaluation due to a suspected learning disability and/or emotional impairment during the 1994-95, 1995-96, and 1996-97 school years. Based on an eligibility decision made by the Individualized Educational Planning Committee (IEPC) for each student in the sample, the students were grouped according to one of the following four eligibility groups: learning disability (LD); emotional impairment (EI); learning disability and emotional impairment (LD/EI); or referred but not eligible (NE) for special education services. Criteria for a diagnosis of LD and/or EI followed state and federal guidelines (see Appendices F and G, respectively). Students for whom ratings were completed but who had IEPC eligibility determinations (e.g., mental impairment, autism impairment, physical or other health impairment, speech and language impairment, etc.) other than LD, EI, LD/EI, or not eligible, were not included in the data analysis.

One of the realities in the schools is that on occasion parents, teachers, and/or evaluators will refuse to formally qualify a student under "EI" eligibility, particularly if that student will qualify for special education services anyway under another label, such as "LD," which is perceived as less stigmatizing. This tendency was accounted for by

considering students called "LD-only" by the IEPC, to be actually LD/EI for the purposes of this study, if all of the following conditions were met: 1) the student had been referred to both the school psychologist and school social worker (which indicates that the general education teacher's had concerns about the student's social-emotional-behavioral functioning), and 2) direct or consultative school social worker services were assigned by the IEP Committee (which suggests that significant emotional issues were identified and required intervention) and 3) both the school social worker involved and this researcher concurred that the student technically met EI eligibility criteria. Of the 20 LD/EI students in this study, 11 (55%) were actually LD-only, and were recategorized as LD/EI for the purposes of this study by meeting the above criteria.

Note, too, that of the 29 EI students in this study, 6 (21%) demonstrated a significant ability-achievement score discrepancy ($\geq$ 18 points with regression in the county under study) in one or more subject areas. Another 6 students (21%) had insufficient data on ability and/or achievement standard scores included on their MET/IEPC paperwork for this researcher to determine if a significant ability-achievement score discrepancy existed. These findings suggest that perhaps some of the EI-only students may have also qualified as LD/EI for the purposes of this study. A decision was not made in this direction, however, due to the assumption that most MET/IEPC members would willingly either

opt for an LD-only or a combined LD/EI label if there was evidence that the ability-achievement discrepancy was due to a specific learning disability (e.g., because of processing problems, etc). Similarly, for that undeterminable subset of MET/IEPC members who interpret the LD and EI eligibility guidelines as being mutually exclusive, their choice of EI over LD suggests that the evidence for EI must have been more compelling than the evidence for LD.

Characteristics of the Sample

In Table 7, the four subsamples are described on the basis of sample size, gender, ethnicity, age, grade, Full Scale IQ, Verbal IQ, Nonverbal IQ, socioeconomic status (indicated by whether or not the student was Medicaid eligible), and ADHD status (regarding a diagnosis or evaluation for attention deficit hyperactivity disorder).

As presented in Table 7, the four subsamples' gender composition ranged from roughly 2:1 in the LD and EI subsamples to nearly 6:1 in the LD/EI subsample. These ratios are consistent with national data for students referred for and identified as LD (e.g., Anderson, 1997), as well as with the validity studies to date. The gender compositions of the four subsamples were essentially similar [chi-squared (9) = 4.4615, $p>.05$]. Each sample is also predominantly European American (ranging from 93.7% to 96.6%), which is commensurate with the ethnic composition of the entire county from which this sample was obtained. Students' mean ages ranged from 8 to 9 years, and mean

Table 7.

Characteristics of the Four IEPC Eligibility Groups

	LD (n = 63)	EI (n = 29)	LD/EI (n = 20)	NE (n = 35)
GENDER				
% male	65.1%	65.5%	85.0%	57.1%
% female	34.9%	34.5%	15.0%	42.9%
ETHNICITY				
% European American	93.7%	96.6%	95.0%	94.3%
% African American	1.6%	0.0%	0.0%	0.0%
% Hispanic American	4.8%	0.0%	5.0%	2.9%
% Other	0.0%	3.4%	0.0%	2.9%
AGE				
mean	8.88	9.19	8.24	8.78
standard deviation	1.65	1.63	1.09	1.32
GRADE				
mean	2.65	3.24	2.35	2.80
standard deviation	1.58	1.85	1.39	1.45
FULL SCALE IQ				
mean	94.62	97.59	94.85	98.42
standard deviation	10.78	11.13	9.34	12.00
VERBAL IQ				
mean	93.12	96.64	95.15	97.21
standard deviation	9.67	10.22	9.86	10.95
PERFORMANCE IQ				
mean	98.17	98.46	96.65	100.48
standard deviation	13.85	12.41	11.82	13.87
SOCIOECONOMIC STATUS				
% eligible for Medicaid	9.5%	41.4%	5.0%	11.4%
ADHD STATUS				
% diagnosed ADHD	19.0%	27.6%	20.0%	11.4%
% ADHD evaluation in	6.3%	10.3%	20.0%	22.9%
progress/recommended	74.6%	62.1%	60.0%	65.7%
% no ADHD diagnosis or evaluation				

grade levels were at the early second to early third grade, which is consistent with the ages/grades at which a large majority of students in this county (and many others) are

first referred for a special education evaluation. All groups' mean Full Scale, Verbal, and Performance IQs fell within the 90 to 100 IQ range. Socioeconomic status (as indicated by percentage of students eligible for Medicaid) varied significantly by group [chi-squared (1) = 18.81, $p<.001$]. Whereas over 40% of the EI group were eligible for Medicaid (or low SES), the other three groups had approximately 10% or fewer students eligible for Medicaid. Lastly, ADHD status rates did not differ significantly by group [chi-squared (6) = 7.24, $p>.05$], and ranged from 60 to 75% with no ADHD diagnosis or evaluation in progress/recommended.

Research Questions

1. Is there a difference between the four IEPC eligibility groups (LD, EI, LD/EI, and NE) on:
   a) the DSF Total score?
   b) the DSF Subscale scores?
2. Is there a difference between the gender and SES groups on:
   a) the DSF Total score?
   b) the DSF Subscale scores?
3. Is there a difference between the three ADHD-status groups (diagnosed,suspected, or nonexistent) on:
   a) the DSF Total score?
   b) the DSF Subscale scores?
4. Are the four DSF Subscale scores useful as predictors of EI status?

5. What is the classification accuracy of the DSF?

   a) What are the sensitivity (true positive) and specificity (true negative) rates for DSF Total cutoff scores of 0 to 2 SD above the mean?

   b) What are the false positive and false negative rates for DSF Total cutoff scores of 0 to 2 SD above the mean?

   c) What are the positive and negative predictive values for the DSF Total optimum cutoff score?

   d) What are the sensitivity (true positive) and specificity (true negative) rates for DSF Subscale cutoff scores of 0 to 2 SD above the mean?

   e) What are the false positive and false negative rates for DSF Subscale optimum cutoff scores?

   f) What are the positive and negative predictive values for the DSF Subscale optimum cutoff scores?

Data Analysis Procedures

The differences between the four subsamples' (LD, EI, LD/EI, and NE) Total Scale and Subscale standard scores were examined using ANOVA. Post-hoc analyses (Tukey's HSD) were performed to isolate where the differences exist. Similarly, ANOVA was used to examine the differences in Total and Subscale standard scores for the two recombined subsamples (all EI and all not-EI). The effect of gender and SES on Total and Subscale standard scores were also analyzed using ANOVA. The impact of students' ADHD status on Total Scale and Subscale standard scores was also

explored using ANOVA and post-hoc analyses (Tukey's HSD). The four Subscale scores were used as predictors of membership (for the four subsamples and the two recombined) by direct discriminant analyses, followed by chi-square analysis.

Note that school psychologists usually choose a cut-off score for decision-making, as opposed to using discriminant analyses. Thus, as presented in Table 8, sensitivity and specificity analyses were performed to test the efficiency of various possible cutoff scores (Total Scale scores from 110 to 120, and Subscale scores from 10 to 16). Sensitivity refers to the conditional probability of identifying an emotional impairment on the DSF when it is in fact present according to the IEPC (*true positives*). Specificity refers to the conditional probability of rejecting an identification of an emotional impairment on the DSF when it is in fact not warranted according to the IEPC (*true negatives*). Also of interest are analyses related to *false negatives* (the conditional probability of rejecting an EI identification on the DSF when EI is in fact present according to the IEPC) and *false positives* (the conditional probability of identifying EI on the DSF when EI is in fact not present according to the IEPC). Lastly, the positive and negative predictive values for the optimum total scale cutoff score are also presented. The positive predictive value refers to the percentage of children identified as EI on the DSF who also later were found EI by the IEPC. The

Table 8.

Sensitivity, Specificity, and Predictive Value Formulas

IEPC DECISION	DSF SCORE AT OR ABOVE CUTOFF (EI)	DSF SCORE BELOW CUTOFF (NOT EI)	INDICES	INDICES
EI (EI or LD/EI)	(a) Number identified as EI by both DSF and IEPC *Valid Positive*	(b) Number identified as not EI by DSF, but IEPC determined EI eligibility *Invalid Negative*	*SENSITIVITY* Conditional probability of identifying EI on the DSF when EI is in fact present according to the IEPC: *True Positives* $\frac{a}{a+b}$	Conditional probability of rejecting an EI identification on the DSF when EI is in fact present according to the IEPC: *False Negative* $\frac{b}{a+b}$
NOT EI (LD or Not Eligible)	(c)Number identified as EI by DSF, but IEPC determined a non-EI eligibility *Invalid Positive*	(d) Number identified as not EI by both DSF and IEPC *Valid Negative*	*SPECIFICITY* Conditional probability of rejecting an EI identification on the DSF when it is in fact not warranted according to the IEPC: *True Negatives* $\frac{d}{c+d}$	Conditional probability of identifying EI on the DSF when EI is in fact not present according to the IEPC: *False Positive* $\frac{c}{c+d}$
TOTALS	*POSITIVE PREDICTIVE VALUE* Conditional probability of the IEPC determining EI eligibility when the DSF predicted EI $\frac{a}{a+c}$	*NEGATIVE PREDICTIVE VALUE* Conditional probability of the IEPC rejecting EI eligibility when the DSF predicted not EI $\frac{d}{b+d}$		

(Based on Gredler, 1997, p. 101)

negative predictive value refers to the percentage of children identified as not-EI on the DSF who also later were found not-EI by the IEPC. These results will provide a basis for comparing predicted group membership to actual

group membership (in Table 9, the manner in which these calculations relate to standard scores and IEPC eligibiilty are presented for illustrative purposes only; i.e., these data do not represent this study's total sample size nor scores.)

In this purely illustrative example of 20 subjects:

a = number of EI or LD/EI at or above cutoff = 5

b = number of EI or LD/EI below cutoff = 2

c = number of all not EI (LD and NE) at or above cutoff = 5

d = number of all not EI (LD and NE) below cutoff = 8

Calculations:

(1) **True Positive (Sensitivity):**

$$\frac{a}{a + b} = \frac{5}{7} = 71\%$$

(2) **False Negative:**

$$\frac{b}{a + b} = \frac{2}{7} = 29\%$$

(3) **True Negative (Specificity):**

$$\frac{d}{c + d} = \frac{8}{13} = 62\%$$

(4) **False Positive:**

$$\frac{c}{c + d} = \frac{5}{13} = 38\%$$

(5) **Positive Predictive Value:**

$$\frac{a}{a + c} = \frac{5}{10} = 50\%$$

(6) **Neg Predictive Value:**

$$\frac{d}{b + d} = \frac{8}{10} = 80\%$$

Table 9.

Illustrative Example of Sensitivity, Specificity, and Predictive Value Formulas (with cutoff score of 115, as shown in bold)

DSF Total Score	90	90	95	95	100	100	105	105	110	110	**115**	**115**	**120**	**120**	**125**	**130**	**135**	**140**	**145**	**150**
IEPC Label	NE	NE	LD	NE	EI	LD	LD	LD/ EI	LD	NE	**LD**	**LD**	**EI**	**LD**	**LD/ EI**	**LD**	**LD**	**EI**	**EI**	**EI**

CHAPTER IV

RESULTS

Questions Related to Four IEPC Eligibility Groups

1. **Is there a difference between the four IEPC eligibility groups (LD, EI, LD/EI, and NE) on:**
   a) **the DSF Total score?**
   b) **the DSF Subscale scores?**

The means and standard deviations of the DSF Total and Subscale scores for the four subsamples are shown in Table 10. Oneway ANOVAs indicate significant differences for all scores: Interpersonal Problems, $F(3, 143) = 19.137$, $p < .001$; Inappropriate Behavior/Feelings, $F(3, 143) = 13.86$, $p < .001$; Depression, $F(3, 143) = 18.624$, $p < .001$; Physical Symptoms/Fears, $F(3, 143) = 11.44$, $p < .001$; and Total Score, $F(3, 143) = 21.85$, $p < .001$. Thus, there are significantly different DSF Total scores and DSF Subscale scores for the four IEPC eligibility groups.

Tukey's HSD post-hoc analyses, shown in Table 11, revealed that for all of the Subscale and Total scores, there are two homogeneous subsets: 1) all of the EI students (i.e., the EI and LD/EI subgroups) and 2) all of the not-EI students (i.e., the LD and NE).

These findings prompted the following post-hoc questions and analyses:

Can the four groups be recombined based on similarities and differences of DSF Total and Subscale scores? If so, what are the sample characteristics of the recombined groups, and are there differences between the recombined groups' DSF Total and Subscale scores?

Table 10.

Means and Standard Deviations (in Parenthesis) of the DSF Scores for Four Subsamples

DSF SCORES	LD (n = 63)	EI (n = 29)	LD/EI (n = 20)	NE (n = 35)
Interpersonal Problems	10.68 (2.61)	14.45 (3.03)	14.40 (3.23)	10.46 (3.04)
Inappropriate Behaviors/Feelings	10.16 (2.52)	13.31 (3.29)	13.45 (3.36)	10.20 (2.63)
Depression	11.51 (3.06)	15.93 (3.46)	15.45 (3.93)	11.00 (3.58)
Physical Symptoms/Fears	10.19 (2.39)	13.28 (3.75)	13.40 (4.25)	10.09 (2.88)
Total	103.73 (12.51)	122.72 (12.72)	122.85 (16.63)	102.51 (14.98)

Table 11.

Tukey's HSD Significance Levels of Means for IEPC Eligibility Groups in Homogeneous Subsets

DSF SCORES	EI and LD/EI (n = 49)	LD and NE (n= 98)
Interpersonal Problems	1.000	0.990
Inappropriate Behaviors/Feelings	0.997	1.000
Depression	0.944	0.935
Physical Symptoms/Fears	0.999	0.999
Total	1.000	0.986

Uses Harmonic Mean Sample Size = 31.025

Based on the Tukey's HSD post-hoc analyses, there is justification for recombining the four groups into two groups (all EI and all not-EI) based on similarities and

differences of DSF Total and Subscale scores. In Table 12, the two recombined groups (all EI and all Not-EI) are described on the basis of sample size, gender, ethnicity, age, grade, Full Scale IQ, Verbal IQ, Performance IQ, socioeconomic status, and ADHD status. The all EI group has a male to female ratio of roughly 3:1, whereas the all not-EI group has a gender ratio of 8:5. Ethnicity is homogeneous in both groups, with over 93% European American membership. Mean age (roughly 8 ½ years) and mean grade level (late second) are quite similar as well. Mean Full Scale, Verbal, and Performance IQS range from 94 to 99. The EI group had a higher proportion of students of low SES (26.5%) than did the not-EI group (10.2%) [chi-squared (1)=5.70, $p<.05$]. The ADHD status proportions ranged from 16 to 25% diagnosed, 12 to 14% evaluation recommended or in progress, and 61 to 71% not ADHD, and these proportions did not differ significantly for the two recombined groups [chi-squared (2)=1.08, $p> .05$].

The means and standard deviations of the DSF Total and Subscale Scores for these two recombined subsamples (all EI vs. all not-EI) are presented in Table 13. Oneway ANOVAs conducted on the two recombined subgroups again indicate significant differences for all scores: Interpersonal Problems, $F (1, 145) = 58.02$, $p < .001$; Inappropriate Behavior/Feelings, $F (1, 145) = 42.11$, $p < .001$; Depression, $F (1, 145) = 55.61$, $p < .001$; Physical Symptoms/Fears, F (1,

Table 12.

Characteristics of the Two Recombined Samples

	ALL EI (n = 49)	ALL NOT-EI (n = 98)
GENDER		
% male	73.5%	62.2%
% female	26.5%	37.8%
ETHNICITY		
% European American	95.9%	93.9%
% African American	0.0%	1.0%
% Hispanic American	2.0%	4.1%
% Other	2.0%	1.0%
AGE		
mean	8.80	8.84
standard deviation	1.50	1.53
GRADE		
mean	2.88	2.70
standard deviation	1.72	1.53
FULL SCALE IQ		
mean	96.47	95.98
standard deviation	12.07	11.31
VERBAL IQ		
mean	96.02	94.59
standard deviation	10.00	10.28
PERFORMANCE IQ		
mean	97.71	99.00
standard deviation	12.07	13.83
SOCIOECONOMIC STATUS		
% eligible for Medicaid	26.5%	10.2
ADHD STATUS		
% diagnosed ADHD	24.5%	16.3%
% ADHD evaluation in progress/recommended	14.3%	12.2%
% no ADHD diagnosis or evaluation	61.2%	71.4%

Table 13.

Means and Standard Deviations (in parenthesis) of the DSF Scores for Two Recombined Samples

DSF Scores	All EI (n = 49)	All Not-EI (n = 98)
Interpersonal Problems	14.43 (3.08)	10.60 (2.76)
Inappropriate Behaviors/Feelings	13.37 (3.28)	10.17 (2.55)
Depression	15.73 (3.63)	11.33 (3.25)
Physical Symptoms/Fears	13.33 (3.92)	10.15 (2.56)
Total	122.78 (14.28)	103.30 (13.38)

145) = 34.73, $p < .001$; Total Score, $F (1, 145) = 66.21$, $p < .001$. Thus, there are significant differences between the recombined groups for all of the DSF Total and Subscale scores, with the all EI sample consistently having higher scores than the all not-EI group.

A review of the subscale means for each group (in Tables 10 and 13) prompts the question of whether or not the four subscales are distinctly different. The total sample's intercorrelation coefficient for each subscale combination is presented in Table 14. (Note that the values for the total group are reported rather than subgroup correlations, because the latter will tend to be reduced in size due to the restricted range.) These data indicate that all six subscale intercorrelation coefficients were statistically significant.

To determine if these correlations were too high, possibly indicating that there are not really four

Table 14.

Subscale Intercorrelation Coefficients (Standard and Corrected) for the Total Sample

	Standard Correlations for Total Sample (n = 147)	Corrected Correlations for Total Sample (n = 147)
Interpersonal Problems - Inappropriate Behaviors/Feelings	.86*	.94*
Interpersonal Problems - Depression	.70*	.79*
Interpersonal Problems - Physical Symptoms/Fears	.75*	.87*
Inappropriate Behaviors/Feelings - Depression	.55*	.60*
Inappropriate Behaviors/Feelings - Physical Symptoms/Fears	.67*	.76*
Depression - Physical Symptoms/Fears	.67*	.77*

Note: * indicates the correlation coefficient is significant at the 0.01 level (2-tailed).

dimensions in the instrument, the full sample correlations were corrected for attenuation. First, internal consistency reliabilities were computed from the present data set. These reliabilities (IP=.89; IBF=.94; DEP=.87; PSF=.84) compared favorably with those obtained for the DSF standardization sample of teacher-rated 5-12 year old students (IP .91; IBF=.94; DEP=.88; PSF=.88), and all exceeded the .80 minimum suggested by Bracken (1987). The corrected correlations are also presented in Table 14. The finding that the Interpersonal Problems-Inappropriate

Behaviors/Feelings corrected correlation (.94) is near 1.0 suggests these two scales may be measuring the same thing. As all the other corrected correlations are below .9, the other two subscales appear to be measuring different constructs.

Questions Related to Gender and SES

2. **Is there a difference between the gender and SES groups on:**
   a) **the DSF Total score?**
   b) **the DSF Subscale scores?**

The means and standard deviations of the DSF Total and Subscale scores for students on the basis of gender (male or female) and SES (Medicaid eligible or ineligible, which corresponds to low SES and moderate to high SES, respectively) are presented in Table 15. Mean DSF scores for male students ranged from 10.81 (Physical Symptoms/Fears) to 12.37 (Depression), and for female students from 11.22 (Inappropriate Behaviors/Feelings) to 13.62 (Depression). The DSF Total scores for males were 108.4, and 112.42 for females. Oneway ANOVAS indicate non-significant differences for all DSF scores on the basis of gender: Interpersonal Problems, $F(1, 145) = 1.814$, $p>.05$; Inappropriate Behaviors/Feelings, $F(1, 145) = .002$, $p>.05$; Depression, $F(1, 145) = 3.333$, $p > .05$; Physical Symptoms/Fears, $F(1, 145) = 3.921$, $p > .05$; and Total Score, $F(1, 145) = 1.949$, $p > .05$. In sum, there are not significant differences between male and female students on the DSF Total and Subscale scores.

Table 15.

Means and Standard Deviations (in parenthesis) of the DSF scores for Gender and SES

DSF Scores	MALE (n = 97)	FEMALE (n = 50)	MEDICAID ELIGIBLE (n = 23)	MEDICAID INELIGIBLE (n = 124)
Interpersonal Problems	11.61 (3.36)	12.40 (3.42)	13.67 (2.86)	11.56 (3.39)
Inappropriate Behaviors/Feelings	11.25 (3.12)	11.22 (3.30)	12.96 (3.05)	10.92 (3.12)
Depression	12.37 (3.72)	13.62 (4.31)	13.70 (4.05)	12.63 (3.94)
Physical Symptoms/Fears	10.81 (3.29)	11.98 (3.55)	12.17 (3.22)	11.03 (3.43)
Total	108.43 (15.86)	112.42 (17.42)	117.57 (14.87)	108.35 (16.39)

As shown in Table 15, mean DSF Subscale scores for Medicaid-eligible (low SES) students ranged from 12.17 (Physical Symptoms/Fears) to 13.70 (Depression), and for Medicaid-ineligible (moderate to high SES) students ranged from 10.92 (Inappropriate Behaviors/Feelings) to 12.63 (Depression). Mean DSF Total scores were 117.57 for Medicaid eligible students, and 108.35 for Medicaid ineligible students. Oneway ANOVAS indicate significant differences for two of the Subscales and the Total score: Interpersonal Problems, $F(1, 145) = 7.055$, p, .05; Inappropriate Behaviors/Feelings, $F(1, 145) = 8.338$, $p < .04$; and Total Score, $F(1, 145) = 6.309$, $P < .05$. In contrast, significant differences were not found for the

remaining two Subscales: Depression, $F(1, 145) = 1.411$, $p > .05$; and Physical Symptoms/Fears, $F(1, 145) = 2.186)$, $p > .05$. Post-hoc tests were not performed because of the small group sizes (note that there were only 23 Medicaid eligible students in the entire sample). In brief, there were significant differences in scores between the Medicaid eligible (low SES) and Medicaid ineligible (moderate to high SES) on the Interpersonal Problems, Inappropriate Behaviors/Feelings, and Total scores, with the low SES students consistently having higher scores. No such differences were found between the two SES groups on the Depression and Physical Symptoms/Fears Subscale scores.

Questions Related to ADHD-Status

3. **Is there a difference between the three ADHD-status groups (diagnosed, suspected, or nonexistent) on:**
   a) **the DSF Total score?**
   b) **the DSF Subscale scores?**

The means and standard deviations of the DSF total and subscale scores for students grouped solely on the basis of ADHD status (i.e., at the IEPC, the student was diagnosed ADHD, or an ADHD evaluation was recommended or in progress, or there was no diagnosis or evaluation of ADHD) are in Table 16. Oneway ANOVAs indicate significant differences only for the Inappropriate Behaviors/Feelings subscale, $F(2, 144) = 6.147$, $p < .01$. Thus, there were not significant differences between the three ADHD-status groups on the DSF Total score and the DSF Subscale scores, with the exception of the Inappropriate Behaviors/Feelings Subscale, for which

Table 16.

Means and Standard Deviations (in Parenthesis) of the DSF Scores for Three ADHD-Status Subsamples

DSF SCORES	ADHD DIAGNOSED (n = 28)	ADHD SUSPECTED (Evaluation in Progress or Recommended) (n = 19)	NOT ADHD (n = 100)
Interpersonal Problems	13.04 (3.65)	12.63 (2.77)	11.41 (3.34)
Inappropriate Behaviors/Feelings	12.25 (3.11)	12.95 (3.17)	10.63 (3.04)
Depression	13.82 (3.67)	12.37 (3.17)	12.59 (4.15)
Physical Symptoms/Fears	12.23 (4.35)	11.95 (2.99)	10.77 (3.13)
Total	115.57(17.06)	113.84(15.14)	107.40(16.12)

Table 17.

Tukey's HSD Significance Levels and Means for ADHD-Status Groups in Homogeneous Subsets

ADHD STATUS	SUBSET FOR ALPHA=.05 1	SUBSET FOR ALPHA=.05 2
NOT ADHD (n =100)	10.63	
ADHD DIAGNOSED (n =28)	12.25	12.25
ADHD SUSPECTED (n = 19)		12.95
SIGNIFICANCE	.098	.649

Uses Harmonic Mean Sample Size = 30.505

These findings prompted the following post-hoc questions and analyses:

Is there a difference on the DSF Total and Subscale scores based on combined EI and ADHD status (i.e., EI and ADHD diagnosed or suspected; EI and Not-ADHD; Not-EI and ADHD diagnosed or suspected; Not-EI and Not-ADHD)?

Table 18 presents the means and standard deviations of the DSF total and subscale scores for the total sample recoded into four groups based on both EI and ADHD status (e.g., EI and ADHD diagnosed/suspected; EI and not-ADHD; not-EI and ADHD diagnosed/suspected; and not-EI and not-ADHD). As shown below, mean DSF Subscale scores for the EI/ADHD group were fairly uniform, ranging from 14.19 (Inappropriate Behaviors/Feelings) to 14.86 (Depression). The EI/ADHD group's mean DSF Subscale scores showed more variation, and ranged from 12.19 (Physical Symptoms/Fears) to 15.58 (Depression). More uniformity in mean DSF Subscale scores was again seen in the Not-EI/ADHD group, with scores ranging from 10.24 (Physical Symptoms/Fears) to 11.92 (Depression). Lastly, the Not-EI/Not-ADHD group had mean DSF Subscale scores which ranged from 9.90 (Inappropriate Behaviors/Feelings) to 11.25 (Depression). DSF Total scores for the four groups ranged from 102.29 (Not-EI/Not-ADHD group) to 123.62 (EI/ADHD). Oneway ANOVAs indicate significant differences for all scores: Interpersonal Problems, $F(3, 143) = 15.773$, $P < .001$; Inappropriate Behavior/Feelings, $F(3, 143) = 14.546$, $p < .001$; Depression, $F(3, 143) = 13.834$, $P < .001$; Physical Symptoms/Fears, $F(3, 143) = 11.213$, $p < .001$; and Total Score, $F(3, 143) = 17.801$, $p < .001$. Thus, there are significant differences on the DSF Total and Subscale scores based on combined EI and ADHD status.

Table 18.

Means and Standard Deviations (in Parenthesis) of the DSF Scores for Four EI/ADHD-Status Subsamples

DSF SCORES	EI AND ADHD DIAGNOSED OR SUSPECTED (n = 21)	EI AND NOT-ADHD (n = 31)	NOT-EI AND ADHD DIAGNOSED OR SUSPECTED (n = 26)	NOT-EI AND NOT-ADHD (n = 69)
Interpersonal Problems	14.33 (3.43)	13.87 (3.35)	11.61 (2.84)	10.33 (2.67)
Inappropriate Behaviors/Feelings	14.19 (3.36)	12.35 (3.24)	11.08 (2.31)	9.90 (2.60)
Depression	14.86 (3.61)	15.58 (4.24)	11.92 (2.88)	11.25 (3.36)
Physical Symptoms/Fears	14.29 (4.34)	12.19 (3.54)	10.24 (2.31)	10.16 (2.69)
Total	123.62 (16.32)	119.13 (15.37)	107.38 (13.01)	102.29 (13.37)

Tukey's HSD post-hoc analyses generally supported the earlier finding of homogeneous groupings based on EI status, with the exception of two Subscales (namely, Inappropriate Behaviors/Feelings and Physical Symptoms/Fears), as shown in Tables 19 and 20. On the Inappropriate Behaviors/Feelings (IBF) Subscale, the "Not-EI and ADHD" subsample and the "EI and Not-ADHD" subsample comprised a homogeneous subset. On the Physical Symptoms/Fears (PSF) Subscale, the "EI and ADHD" subsample was one subset, with the other three EI/ADHD status subsamples comprising another homogeneous subset. In sum, generally the variable that has the greatest significance in terms of DSF Total and Subscale scores is

Table 19.

Tukey's HSD Significance Levels and Means for EI/ADHD-Status Groups in Homogeneous Subsets for IBF Subscale Scores

EI/ADHD-STATUS	SUBSET FOR ALPHA = .05	SUBSET FOR ALPHA = .05	SUBSET FOR ALPHA = .05
NOT-EI and NOT-ADHD (n = 69)	9.90		
NOT-EI and ADHD (n = 26)	11.08	11.08	
EI and NOT-ADHD (n = 31)		12.35	12.35
EI and ADHD (n = 21)			14.12
SIGNIFICANCE	.365	.293	.056

Uses Harmonic Mean Sample Size = 30.113

Table 20.

Tukey's HSD Significance Levels and Means for EI/ADHD-Status Groups in Homogeneous Subsets for PSF Subscale Scores

EI/ADHD STATUS	SUBSET FOR ALPHA = .05	SUBSET FOR ALPHA = .05
NOT-EI and NOT-ADHD (n = 69)	10.16	
NOT-EI and ADHD (n = 26)	10.34	
EI and NOT-ADHD (n = 31)	12.20	
EI and ADHD (n = 21)		14.29
SIGNIFICANCE	.054	1.000

Uses Harmonic Mean Sample Size = 30.113

whether or not a student is EI. Only on the Inappropriate Behaviors/Feelings and Physical Symptoms/Fears Subscales do the ADHD variables have a significant impact on the scores.

Questions Related to Subscale Scores as Predictors

4. **Are the four DSF Subscale scores useful as predictors of EI status?**

A discriminant analysis was performed using the four DSF subscales as predictor variables, and IEPC eligibility (EI or not-EI) as the criterion variable. Table 21 presents the univariate F values, structure coefficients, and standardized canonical discriminant function coefficients. The univariate F values indicate that all four subscales contribute to the differentiation between the two groups. The structure coefficients demonstrate that while the Interpersonal Problems (.89) and Depression (.87) Subscales had the highest coefficients, all four coefficients were high enough (Inappropriate Behaviors/Feelings .76 and Physical Symptoms/Fears .69) to lend support for the use of all four subscales when differentiating the groups. Lastly, the standardized canonical discriminant function coefficients, based on the raw subscale results, indicate which weights should be applied to each scale. Based on this discriminant analysis, the formula for the statistically optimal cutscore for classification purposes, using the subscale standard scores (IP, IBF, DEP, and PSF) in combination, is as follows:

F1* = .1 (IP + IBF) + .2 (DEP) - .02 (PSF) - 4.1

** if F1 < .73, the student most likely is not EI*

** if F1 > .73, the student most likely is EI*

The ability of the discriminant functions to accurately predict EI or not-EI membership was tested by classification analyses, as shown in Table 22. Classification rates were moderately high, with over 85% of the EI students and over

Table 21.

Results of Discriminant Function Analysis

Subscale	r	Univariate F's	Standardized Canonical Discriminant Function Coefficients
Interpersonal Problems	.89	58.02	.40
Inappropriate Behavior/Feelings	.76	42.11	.23
Depression	.87	55.61	.57
Physical Symptoms/Fears	.69	34.73	-.05
Note: r = Predictor variable and discriminant function r.			

Table 22.

Discriminant Analysis Classification Results

Actual Membership	n	Predicted Membership EI	Predicted Membership Not-EI	Overall Accuracy
All EI (EI and LD/EI)	49	33 67.3%	16 32.7%	
All Not EI (LD and NE)	98	14 14.3%	84 85.7%	
Totals	147			117 79.6%

67% of the not-EI students classified correctly. Conversely, this also means that roughly 15% of the EI students and 33% of the not-EI students were classified incorrectly. In all, nearly 80% of the total sample was classified accurately when discriminant functions were used. These discriminant functions were significant [chi-squared (4)=58.8, p<.001).

Questions Related to Classification Accuracy

5. **What is the classification accuracy of the DSF?**
   (a) **What are the sensitivity (true positive) and specificity (true negative) rates for DSF Total cutoff scores of 0 to 2 SD above the mean?**

Because most practitioners do not use a discriminant functions formula when interpreting behavior rating scales for individual students, the efficiency of various possible cutoff scores in predicting EI and not-EI status was analyzed for sensitivity (true positives) and specificity (true negatives). Tables 8 and 9 show how these two values are calculated. The correct classification rates for Total cutoff scores of 0 to 2 SD above the mean are shown in Table 23. Note that traditionally, researchers often identify the "optimum cutoff score" based on the score for which a balance between true positive and true negative hit rates are achieved, such that minimum false positive and false negative errors are obtained. (In contrast, on a practical basis, practitioners may decide to err more on the false positive or false negative side, depending on the purposes of the assessment.) In the present study, such a balance between the true positive (sensitivity=77.6%) and true negative (specificity=78.8%) hit rates was achieved when 114 was used as a cutoff score. Also, the higher cutoff scores were associated with higher specificity values (i.e., higher percentage of not-EI students identified) and lower sensitivity values (i.e., lower percentage of EI students identified). Figure 2 shows this relationship for Total

Table 23.

Sensitivity and Specificity Rates with DSF Total Cutoff Scores of 0 to 2 SD Above the Mean

Cutoff Value	Specificity True Negative (not EI)	Sensitivity True Positive (EI)
110 (0 SD)	68.4%	87.8%
111	69.4%	83.7%
112	71.4%	83.7%
113	74.5%	79.6%
114	*78.6%*	77.6%
115 (1 SD)	79.6%	71.4%
116	81.6%	71.4%
117	81.6%	69.4%
118	81.6%	69.4%
119	85.7%	67.3%
120	86.7%	67.3%
130 (2 SD)	95.9%	26.5%

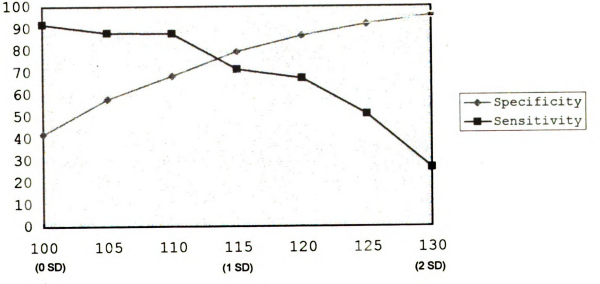

Figure 2. Sensitivity and specificity rates with DSF Total standard cutoff scores of 0 to 2 SD above the mean.

Scores of 100 to 130 (0 to 2 SD above the mean), which covers the range of "Normal" to "Very Significant" in the DSF Manual. In sum, the data in Table 22 and Figure 2 show that over the range of DSF Total cutoff scores of 0 to 2 SD above the mean, specificity rates (true negatives) increase from 68% to nearly 96%, and sensitivity rates (true positives) decrease from nearly 88% to nearly 27%. A balance between the two hit rates was obtained when 114 (nearly 1 SD above the mean) was the cutoff score. However, at that point, over one of five not-EI students and over one of five EI students were not identified accurately.

(b) **What are the false positive and false negative rates for DSF Total cutoff scores of 0 to 2 SD above the mean?**

Practitioners also need to know the likelihood of making a diagnostic error (i.e., rejecting EI on the DSF when EI is present according to the IEPC--false negative, and identifying EI on the DSF when EI is not present according to the IEPC-- false positive). Diagnostic errors are important to consider because, at the very least, practitioners try to adhere to the dictum, "First, do no harm." Moreover, there are potential financial, educational, and emotional consequences to students, parents, and schools when students are incorrectly classified as having or not having significant emotional impairments. The incorrect classification rates for Total Scores of 0 to 2 SD above the mean are shown in Table 24.

Table 24.

Incorrect Classification Rates with DSF Total Cutoff Scores of 0 to 2 SD Above the Mean

Cutoff Value	False Negative	False Positive
110 (0 SD)	12.2%	31.6%
111	16.3%	30.6%
112	16.3%	28.6%
113	20.4%	25.5%
114	**22.4%**	**21.4%**
115 (1 SD)	28.6%	20.4%
116	28.6%	18.4%
117	30.6%	18.4%
118	30.6%	18.4%
119	32.7%	14.3%
120	32.7%	13.3%
130 (2 SD)	73.5%	4.1%

Note that at the cutoff score of 114, previously selected due to the balance between true positive and true negative hit rates, roughly 22% of students found eligible as EI (including LD/EI) by the IEPC were not identified as such by the DSF (false negative), and another 21% of students found not-EI by the IEPC were considered EI based on their DSF scores (false positives). This means that even when there is an optimal balance (psychometrically) between the true positive and true negative hit rates, approximately one of five students would be incorrectly classified as either EI (when he/she was not EI) or not-EI (when he/she was EI) if

eligibility decisions were made solely on the basis of their DSF Total cutoff scores.

(c) What are the positive and negative predictive values for the DSF Total optimum cutoff score?

Gredler (1997, p. 102) suggests that the positive predictive value is the most important efficacy index, in that "school districts act mainly on the number of children who are classified as `at risk.'" The positive predictive value is the percentage of children identified as EI on the DSF who also later were found EI (including LD/EI) by the IEPC. Also of interest is the negative predictive value, which is the percentage of children identified as not-EI by the DSF who also later were found not-EI by the IEPC. Tables 8 and 9 review how these two values are calculated.

Using the present study's optimum total scale cutoff score of 114, 64% of students who had high DSF scores were later found eligible as EI (including LD/EI) by the IEPC (positive predictive value). Using the same optimum total scale cutoff score of 114, nearly 88% of students who had low DSF scores were later found not-EI by the IEPC (negative predictive value). In sum, approximately two-thirds of students with high DSF scores (at or above 114) were deemed eligible as EI (including LD/EI) by the IEPC, whereas roughly one-third with similarly high DSF scores were not deemed eligible as EI (i.e., LD-only or not eligible) by the IEPC. Further, almost nine of ten students with low DSF scores (below 114) were subsequently declared not eligible

as EI or LD/EI by the IEPC. Conversely, approximately one of ten students with low DSF scores were later certified as EI (or LD/EI) by the IEPC. Thus, diagnostic errors were made more often for those with high DSF scores than those with low DSF scores.

(d) What are the sensitivity (true positive) and specificity (true negative) rates for DSF Subscale cutoff scores of 0 to 2 SD above the mean?

Because the discriminant function analyses provided support for the use of the Subscale scores, similar sensitivity and specificity analyses were conducted for each of the Subscales. As shown in Table 25, the most effective cutoff values varied from 12 to 14 for the four Subscales, which is consistent with the manual's recommendation of a Subscale cutoff score of 13. Figures 3 through 6 show the increasing specificity values and decreasing sensitivity values as the Subscale cutoff scales increased from 10 to 16 (up to 2 SD above the mean). These data indicate that at the psychometrically optimum Subscale cutoff scores of 12 to 14 (which varied by Subscale), roughly 75% of students identified as not-EI by the DSF were subsequently deemed not-EI by the IEPC, and another 75% students detected as EI by the DSF were later certified as EI (or LD/EI) by the IEPC. The one exception to the above was found on the Physical Symptoms/Fears Subscale, as a lower rate of only 63% of students rated as EI by the DSF on this Subscale were later declared EI (or LD/EI) by the IEPC. This suggests

Table 25.

Sensitivity and Specificity Rates with DSF Subscale Cutoff Scores of 0 to 2 SD Above the Mean

Subscale	10 (0 SD)	11	12	13 (1 SD)	14	15	16 (2 SD)
Interpersonal Problems							
Specificity	42.9	54.1	64.3	*76.5*	81.6	87.8	94.9
Sensitivity	89.9	87.8	79.6	*75.5*	71.4	59.2	40.8
Inappropriate Behavior/Feelings							
Specificity	45.9	59.2	*69.4*	79.6	89.8	94.9	96.9
Sensitivity	85.7	71.4	*69.4*	61.2	49.0	34.7	28.6
Depression							
Specificity	28.6	45.9	61.2	67.3	*76.5*	85.7	89.8
Sensitivity	95.9	89.8	85.7	83.6	*77.6*	69.4	55.1
Physical Symptoms/Fears							
Specificity	50.0	60.2	*75.5*	79.6	88.8	94.9	96.9
Sensitivity	79.6	73.5	*63.2*	59.2	49.0	34.7	30.6

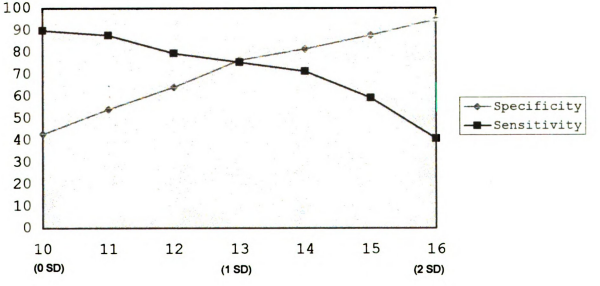

Figure 3. Sensitivity and specificity rates with DSF Interpersonal Problems subscale cutoff scores of 0 to 2 SD above the mean.

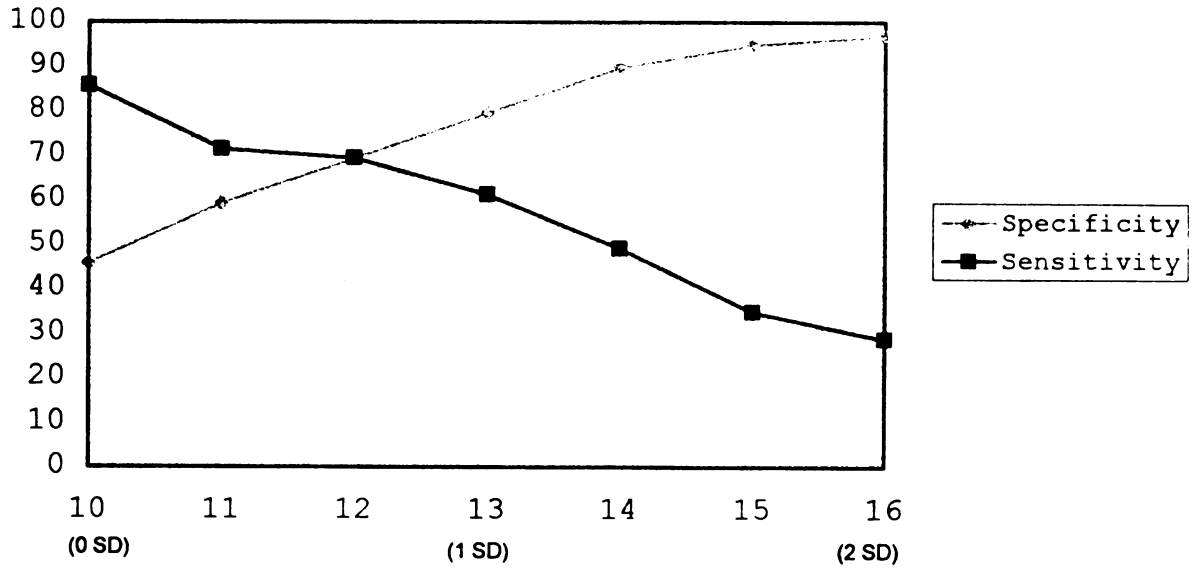

Figure 4. Sensitivity and specificity rates with DSF Inappropriate Behaviors/Feelings subscale cutoff scores of 0 to 2 SD above the mean.

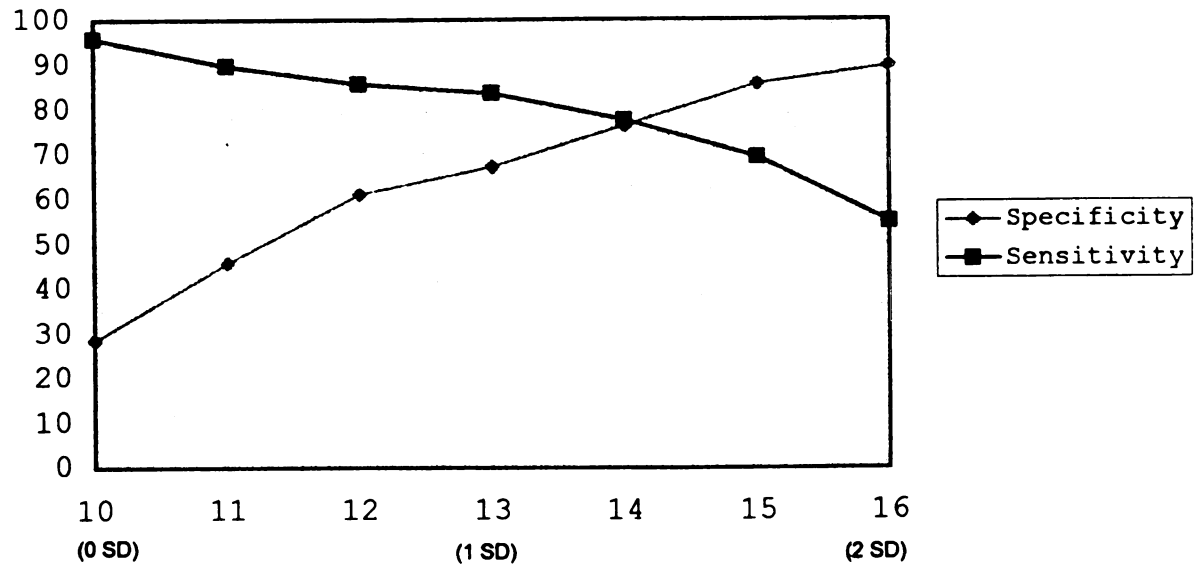

Figure 5. Sensitivity and specificity rates with DSF Depression subscale cutoff scores of 0 to 2 SD above the mean.

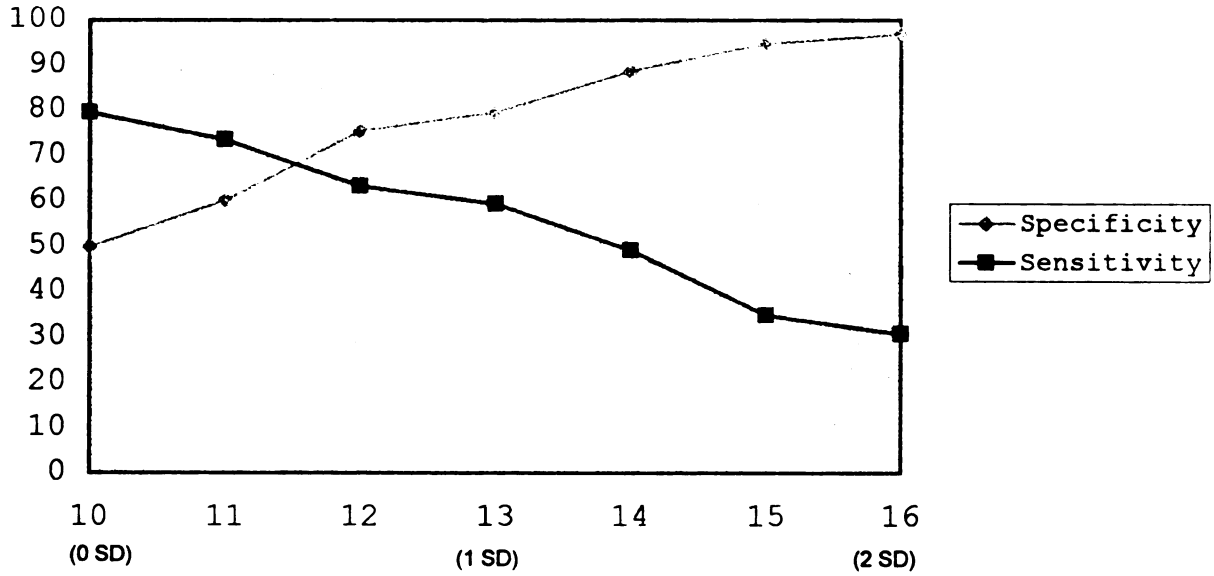

Figure 6. Sensitivity and specificity rates with DSF Physical Symptoms/Fears subscale cutoff scores of 0 to 2 SD above the mean.

that IEPC members are less likely to certify students as EI on the basis of Physical Symptoms/Fears scores than they are for the other problem areas.

Another way to interpret the data is that when the manual's suggested Subscale cutoff score of 13 is used for all Subscales, roughly 53% of the not-EI students had no elevated DSF Subscale scores, as would be expected (specificity--true negatives). Scores were below the cutoff for nearly 80% of the not-EI students on the Interpersonal Problems, Inappropriate Behaviors/Feelings, and Physical Symptoms/Fears Subscales, and roughly 67% of the not-EI

students on the Depression Subscale. Using the same Subscale cutoff of 13, nearly 92% of all of the EI students had at least one elevated DSF Subscale score, as would be expected (sensitivity--true positives), and which is quite good for a screening measure. Scores were most likely to be high for the EI students on the Depression (83.7%) and Intepersonal Problems (75.5%) Subscales. Scores on the other two Subscales (Inappropriate Behaviors/Feelings--61.2% and Physical Symptoms/Fears--59.2%) were high for roughly two of three EI students.

(e) What are the false positive and false negative rates for DSF Subscale optimum cutoff scores?

Again, for every correct classification rate (true negative and true positive), there is a corresponding incorrect classification rate (false negative and false positive). As shown in Table 26, note that at the Subscale cutoff scores of 12 to 14, previously selected due to the optimum balance between true positive and true negative hit rates, up to 36% of students found eligible as EI by the IEPC were not identified as such by at least one of the DSF Subscales (false negatives), and up to 30 % of students found not-EI by the IEPC were considered EI based on at least one DSF Subscale score (false positives).

In other words, when there is an optimal balance (psychometrically) between true positive and true negative hit rates, three of every ten students designated as EI by a DSF Subscale cutoff score (at or above 12 to 14) are not

Table 26.

Incorrect Classification Rates with DSF Subscale Optimum Cutoff Scores

Subscale	12	13	14
Interpersonal Problems			
False Positive	35.7	**23.5**	18.4
False Negative	20.4	**24.5**	28.6
Inappropriate Behaviors/Feelings			
False Positive	**30.6**	20.4	10.2
False Negative	**30.6**	38.8	51.0
Depression			
False Positive	38.8	32.7	**23.5**
False Negative	14.3	16.4	**22.4**
Physical Symptoms/Fears			
False Positive	**24.5**	20.4	11.2
False Negative	**36.8**	40.8	51.0

certified as EI by the IEPC, and more than one in three students scoring low on a DSF Subscale (below 12 to 14) were subsequently deemed eligible as EI by the IEPC. At first glance, these data seem to suggest that numerous diagnostic errors occur even when sensitivity and specificity rates are high. For instance, inspection of the Subscales in Table 26 shows that the rates of false negatives are particularly high for the Subscales assessing Inappropriate Behaviors/Feelings and Physical Symptoms/Fears. In fact, when a cutoff score of 14 is applied, the false negative rates soar to 51% for both of these Subscales. These data indicate that roughly half of the students judged as EI (or

LD/EI) by the IEPC are not detected as emotionally disabled by these two DSF Subscales.

However, it is critical to remember that EI eligibility only requires significant problems in one of the four problem areas. It is very possible for an EI student to have a low score on one Subscale and yet have a high score on another Subscale. Indeed, when the data are aggregated and the manual's recommended cutoff score of 13 is used, only 8% of the EI students had no high DSF Subscale scores (false negatives). This is a very respectable false negative rate for a screening measure. Somewhat less impressive is the finding that nearly 47% of the not-EI students had at least one DSF Subscale score at or above a cutoff score of 13 (false positives). For these students found not-EI by the IEPC, scores were most likely to be elevated on the Depression Subscale (32.7%), and somewhat less likely to be elevated (roughly 20%) on the other three Subscales.

(f) What are the positive and negative predictive values for the DSF Subscale optimum cutoff scores?

As review, the positive predictive value is the percentage of children identified as EI on the DSF who also later were found EI by the IEPC. In contrast, the negative predictive value is the percentage of children identified as not-EI by the DSF who also later were found not-EI by the IEPC. Tables 8 and 9 provide the formulas for calculating these indices. As shown in Table 27, using the present

Table 27.

Predictive Values for DSF Subscale Optimum Cutoff Scores

Subscale	Optimum Cutoff Score	Positive Predictive Value	Negative Predictive Value
Interpersonal Problems	13	61.6%	86.2%
Inappropriate Behaviors/Feelings	12	53.1%	81.9%
Depression	14	62.3%	87.2%
Physical Symptoms/Fears	12	56.4%	80.4%

study's optimum subscale cutoff scores of 12 to 14 (based on specificity and sensitivity rates), 53 to 62% of students who had high DSF subscale scores were later found eligible as EI by the IEPC (positive predictive value). Using the same optimum subscale cutoff scores of 12 to 14, 80 to 87% of students who had low DSF subscale scores were later found not-EI by the IEPC (negative predictive value). In sum, approximately three of five students who scored high on a DSF Subscale (at or above 12 to 14, depending on the Subscale) were deemed eligible as EI (or LD/EI) by the IEPC, as expected. This means that two of five students with at least one similarly high DSF Subscale score were not certified as EI (or LD/EI) by the IEPC, which is contrary to expectations. Over four of five students with a low DSF Subscale score (below 12 to 14, depending on the Subscale)

were subsequently declared not-EI by the IEPC, as expected. Thus, one of five students with at least one similarly low DSF Subscale score were deemed EI (or LD/EI) by the IEPC, which was contrary to expectations. As was the case for the DSF Total scores, diagnostic errors were made more often for those with high DSF Subscale scores than those with low DSF Subscale scores. Also, these rates are affected by the fact that students only need to have significant problems in one of the four problem areas to meet EI criteria.

CHAPTER V

DISCUSSION

This study has provided new information about the validity of the DSF. Using DSF scales completed by general education teachers of elementary-aged students undergoing an initial special education evaluation, this study assessed DSF Total and Subscale score differences among four IEPC eligibility groups (LD, EI, LD/EI, and NE), gender, SES, and three ADHD-status groups (diagnosed, suspected, and nonexistent). The summaries of results, explanations for findings, links to the literature, and conclusions related to the purposes of this study are all discussed in the sections that follow.

DSF Score Differences for the Four IEPC Eligibility Groups

1. **Is there a difference between the four IEPC eligibility groups (LD, EI, LD/EI, and NE) on:**
   (a) **the DSF Total score?**
   (b) **the DSF Subscale scores?**

The present findings revealed significant differences on the DSF Total and Subscale scores for the four IEPC eligibility groups. Specifically, the two samples of students with emotional impairments (EI and LD/EI) had significantly higher DSF Total and Subscale scores than did the two samples of students without emotional impairments (LD and NE) at the point of an initial referral for special education services. Note, however, that a full 55% of the students in the LD/EI group were technically in the LD-only

group based on IEPC decisions alone. These students were recategorized as LD/EI for the purposes of this study because all of the following criteria were met: 1) referrals had been made to both the school psychologist and school social worker; 2) direct or indirect school social worker services were assigned by the IEPC; and 3) both the school social worker involved and this researcher agreed unanimously that the student technically met EI criteria, even though that was not the conclusion of the IEPC. When this recategorization is taken into account, the results suggest that the DSF may be a helpful assessment instrument for use when students are referred to both a school psychologist and a school social worker. Regardless of the IEPC eligibility label the student eventually receives, the DSF appears to be helpful in illuminating social-emotional-behavioral problems experienced by students referred to both specialists.

The mean DSF score analyses were quite similar when the present study's EI and Not-EI samples were compared to previously published DSF validity studies' EI and general education samples, respectively (e.g., Goh, 1997; Naglieri, LeBuffe, & Pfeiffer, 1993; Naglieri, Bardos, & LeBuffe, 1995). Note that all of the prior DSF validity studies provided discriminant validity evidence only for students already in special education versus those in general education. The present study provides the first piece of evidence that the DSF can also help make the more difficult

discrimination between groups of "academically at-risk" students (i.e., warranting a special education referral) with and without emotional impairments prior to placement in special education.

Findings for the dually diagnosed (LD/EI) subgroup present some unique interpretive considerations. In particular, post-hoc analyses revealed that the LD/EI subgroup's scores were much more similar to the EI subgroup than to the LD-only and Not Eligible subgroups. As stated previously, some MET and IEPC members resist determining young children eligible under the EI category, and instead choose less stigmatizing labels (such as "LD-only) whenever possible. Further, the state and federal IDEA classification guidelines discourage LD/EI dual diagnoses. For instance, Michigan's LD criteria state, "The term (LD) does not include children who have learning problems which are primarily the result of ... emotional disturbance ..." (Michigan State Board of Education, 1994, p. 10). Federal regulations also stipulate that LD cannot be caused by emotional factors. Many evaluators interpret this to mean that the eligibility categories of LD and EI are mutually exclusive from the legal standpoint of state and federal guidelines. However, findings from the present study and others provide support for the concomitance of learning disabilities and emotional/behavioral disorders. For instance, Rock et al. (1997) summarized the literature on this issue, and concluded that 25 to 50% of students

identified as LD have significant social-emotional-behavioral problems, and that 40 to 75% of students identified as EI have significant learning problems. While these types of studies do not confirm dual diagnosis in a high percentage of cases, they do highlight the interaction between learning and social-emotional-behavioral problems associated with these two disability groups.

Another finding of some significance is that the mean DSF Depression Subscale scores for the two EI samples (EI-only and LD/EI) were within the "very significant" range. Interestingly, several of the school psychologists and social workers participating in this study remarked that the DSF Depression Subscale scores were higher than they had expected. The DSF Depression Subscale items appear to address several symptoms of clinical childhood depression, rather than just garden-variety demoralization that may accompany academic difficulties. Table 28 compares the DSF Depression Subscale items with items from the Children's Depression Inventory (CDI; Kovacs, 1980), criteria for a Major Depressive Episode (MDE) from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV; American Psychiatric Association, 1994), and the anxious/depressed Subscales from the Achenbach Child Behavior Checklist (CBCL; Achenbach, 1991), Youth Self Report (YSR; Achenbach, 1991), and the Teacher Report Form (TRF; Achenbach, 1991). There was considerable overlap between the DSF Depression Subscale and the other depression

Table 28.

Comparison of Abbreviated Items from Selected Depression Inventories and Subscales

DSF Depression Subscale (Ages 5-12 Form)	CDI	DSM-IV MDE	CBCL, YSR, and TRF Anxious/Depressed Subscale
3 items	*anhedonia	*diminished interest or pleasure	---
2 items	*frequent sadness	*depressed or irritable mood	*unhappy, sad, or depressed
1 item	*feels self is not as good as most people	*feelings of worthlessness (or guilt)	*feels worthless
1 item	*frequent crying	---	*cries a lot
1 item	* doesn't want to be with people		
2 items (regarding flat affect)	---	---	---
(2 items on Ages 13-18 Form only)	*thinks about suicide	*recurrent thoughts about death, suicidal ideation	---
(1 item on Ages 13-18 Form only)	*fatigue	*fatigue or loss of energy	---
--- --- ---	*poor appetite *trouble sleeping *indecisive	*weight loss or gain *insomnia or hypersomnia *diminished ability to think or concentrate	--- --- ---

DSF Depression Subscale (Ages 5-12 Form)	CDI	DSM-IV MDE	CBCL, YSR, and TRF Anxious/Depressed Subscale
--- --- --- --- ---	*feels lonely *bothered by things *frequent crying *feels unloved *worries bad things are going to happen	--- --- --- --- ---	*lonely *hurt when criticized (TRF) *cries a lot *feels unloved *worries
---	12 items (regarding school failure, feeling ugly, aches/pains, not having fun, having no friends, fighting, feeling inferior, incompetent, noncompliance, hopelessness, feeing "bad," self-hatred)	---	----
---	---	1 item (psychomotor retardation or agitation)	---
---	----	---	6 items (regarding fearing own impulses, needing to be perfect, feeling persecuted, fear/anxiety, self-consciousness, suspiciousness)

Note: Asterisked items (*) are included in two or more measures. Dashes (--) mean a corresponding item was not present. Table adapted from Compas (1997, p. 214).

have items reflecting other common symptoms of depression, such as problems with eating and sleeping, suicidal ideation, or worrying. The fact that the participating school psychologists and social workers were surprised at the relatively high incidence of depression among students provides additional evidence that depression in children is often underidentified, and that children referred for learning and/or behavioral issues should be assessed for depressive symptoms (Wright-Strawderman & Watson, 1992). The DSF may be one instrument that can help rectify that situation by sampling a limited array of depression indicators, after which a more in-depth evaluation of depression should be undertaken if the DSF Depression Subscale score is elevated.

Lastly, the full sample correlations, corrected for attenuation, suggest that the Interpersonal Problems and Inappropriate Behaviors/Feelings subscales may be measuring the same construct. If the two scores are quite different for an individual student, further assessment may be in order to account for that type of unexpected result.

DSF Score Differences for Gender and SES

2. Is there a difference between the gender and SES groups on:
(a) <u>the DSF Total score?</u>
(b) <u>the DSF Subscale scores?</u>

There were not significant differences in DSF Total scores and DSF Subscale scores on the basis of gender alone. This suggests that the separate norms provided for male and

female students are effective and equitable in transforming raw scores into meaningful standard scores for both gender groups.

There were some significant DSF score differences between the two SES groups of Medicaid eligible (low SES) and Medicaid ineligible (moderate to high SES). More specifically, the Medicaid eligible students had significantly higher scores on the Interpersonal Problems, Inappropriate Behaviors/Feelings, and Total scores than did the Medicaid ineligible students. No such differences were found between the two SES groups on the Depression and Physical/Symptoms Subscale scores. These general findings are consistent with others that indicate a disproportionate number of children with disabilities are identified from impoverished environments (Smith et al., 1997). On one hand, the higher rate of social-emotional problem identification among the low SES students may reflect the outcomes of any number of risk factors often (but not always) associated with poverty (e.g., poor sanitation, air, and water; high lead levels; exposure to alcohol and drugs among expecting and practicing parents; etc.). On the other hand, the elevated DSF scores among the low SES students in this sample may also reflect some teachers' narrow perceptions about what is "normal" with regard to language, culture, ability, and lifestyle (Smith et al., 1997). The Michigan special education criteria for LD and EI contain the exclusionary criteria: "A determination of impairment

shall not be based solely on behaviors related to environmental, cultural, or economic differences" (Appendix G). This is indeed a difficult task, given the implication of environmental, cultural, and economic factors in the academic and social learning process. The exclusionary criteria for EI also include: "The term `emotionally impaired' does not include persons who are socially maladjusted, unless it is determined that such persons are emotionally impaired" (Appendix G). Given that the low SES students in this study had elevated scores on the two Subscales most likely to reflect social maladjustment, one wonders if IEPC teams are sensitive enough to this diagnostic issue. Practitioners are cautioned to carefully review other assessment data when DSF scores are elevated for low SES students.

DSF Score Differences for the ADHD-Status Groups

3. **Is there a difference between the three ADHD-status groups (diagnosed, suspected, or nonexistent) on:**
   (a) **the DSF Total score?**
   (b) **the DSF Subscale scores?**

The findings related to ADHD-status (i.e., diagnosed, suspected, or nonexistent) are also noteworthy. When all scores were compared based on ADHD-status alone, there were significant differences on the Inappropriate Behaviors/Feelings Subscale. This makes some sense intuitively, given the characteristic impulsive/inappropriate emotions and behaviors demonstrated by most individuals with ADHD. When the sample was recoded

based on combinations of EI and ADHD-status (i.e., EI/ADHD, EI-only, Not-EI/ADHD, Not-EI-only), post-hoc analyses revealed essentially similar scores for the EI-only and Not-EI/ADHD subgroups on that same subscale. These analyses suggest that evaluators need to consider an elevated score on the Inappropriate Behaviors/Feelings Subscale as an indicator to rule out ADHD prior to (or in addition to) determining a student's eligibility under the emotional impairment category. Similarly, given the high comorbidity rates of ADHD and conduct disorders (Stahl & Clarizio, 1999), along with the previously discussed finding regarding low SES students, practioners are advised to give serious consideration to whether a student might be "socially maladjusted" rather than emotionally impaired when an elevated Inappropriate Behaviors/Feelings Subscale score is presented. Interestingly, a few teachers wrote on their rating scales that they would have rated some students with ADHD as "very frequently" exhibiting many of the DSF item behaviors prior to these students receiving Ritalin. Indeed, 52% of the false positives in this study (i.e., LD and NE students with total scores exceeding the cutoff) were either diagnosed or under evaluation for ADHD. While the DSF does not explicitly take into account ADHD symptoms, further study is needed on this variable, especially given the small sample sizes in the present study.

The fact that the other DSF Scores (Total, Interpersonal Problems, Depression, and Physical

Symptoms/Fears) did not differ significantly between the three ADHD-status groups is somewhat surprising, given the high rate of comorbid emotional/behavioral problems that individuals with ADHD typically experience (Barkley, 1993). The finding that EI-status-alone has a greater bearing on DSF scores than does ADHD-status-alone does suggest, however, that the DSF items primarily assess severe emotional impairment (which is the purpose of the scale), rather than ADHD symptoms such as inattention, impulsivity, and overactivity. Further, by and large, students with both EI and ADHD had higher DSF scores than students with either EI or ADHD, which speaks to the severity of behavioral symptoms presented when students have comorbid conditions. These findings do indicate the need to use other assessment measures, including other behavior rating scales, to diagnose ADHD, as the DSF has been neither designed nor proven to be up to that task.

Utility of DSF Subscale Scores as Predictors of EI Status

4. **Are the four DSF Subscale scores useful as predictors of EI status?**

A discriminant analysis, using the four DSF subscales as predictor variables and IEPC Eligibility (all EI or all not-EI) as the criterion variable, indicated that all four subscales contribute to the differentiation between the two groups. A statistically optimal cutoff score formula was presented which used the subscale standard scores in combination. In all, nearly 80% of the total sample (over

85% of EI students and over 67% of the not-EI students) were classified accurately when discriminant functions were used. These results compare favorably with prior DSF validity studies, which found overall accuracy rates of roughly 77%, with roughly 75% of the EI and roughly 79% of general education students correctly classified (Goh, 1997; Naglieri, Bardos, & LeBuffe, 1995). These data can be viewed as suggesting that the prior criterion-related validity studies of the DSF (which focused on general education and special education students) can be generalized to "academically at-risk" students undergoing an initial special education evaluation. A word of caution, however, is that the results also indicate that roughly 14% of the EI students and 33% of the not-EI students were classified incorrectly based on DSF scores. Thus, practitioners need to consider the DSF just one part of a multi-method, multi-informant, multi-setting evaluation.

Classification Accuracy of the DSF

5. What is the classification accuracy of the DSF?

(a) What are the sensitivity (true positive) and specificity (true negative) rates for DSF Total cutoff scores of 0 to 2 SD above the mean?

Rather than using the complicated discriminant analysis formula to predict group membership as described above, most school practitioners rely on a cutoff score for diagnostic decision-making. Based on the present study's sensitivity and specificity analyses, the DSF appears to have clinical utility as one part of a multi-method, multi-informant,

multi-setting evaluation of emotional impairment. For the total sample, the most efficient total scale cutoff score (114) was approximately 1 SD above the mean. These high true positive (77.6%) and true negative (78.6%) hit rates are similar to those found in Goh's (1997) study of culturally diverse children (about 80% hit rates at a cutoff score of 115). They contrast markedly from those reported in the DSF manual for other samples of elementary school children (Naglieri, LeBuffe, & Pfeiffer, 1993; 54 to 63% true positives and 87 to 92% true negatives at a cutoff score of 115). The higher sensitivity values found in the present study lend support for the DSF developers' assertion that the DSF "would likely show higher sensitivity when used at referral" rather than when used with emotionally impaired students who are already in an intervention program (Naglieri, Bardos, & LeBuffe, 1995, p. 109). That is, the DSF was better at identifying the not-yet-labeled EI students (who are not receiving special education services and consequently are likely to be exhibiting significant problems) in the present study than it was at identifying already-labeled EI students (who are receiving special education services and consequently are likely to be exhibiting fewer problems). By the same token, the lower specificity rates in the present study also may be accounted for, in part, by the differences in comparison groups. Recall that the DSF manual validity studies used general education (presumably non-disabled) students in the control

group, while the present study's comparison group consisted of LD and referred-but-not-eligible students. As mentioned previously, the confounding factors of potentially under-identified dual diagnoses (LD/EI) and ADHD may also have resulted in students who were identified as having significant problems according to the DSF, but who were not deemed EI by the IEPC.

Further, as was the case in Goh (1997), increasing the percentage of not-EI students correctly identified by the DSF from roughly 80% to 90% would require a cutoff score of 125 (1 2/3 SD above the mean). However, that would reduce the percentage of EI students correctly identified to nearly 50%. This trade-off of a modest gain in not-EI students being identified correctly for a dramatic decrease in EI students being identified is not desirable. As Goh (1997, p. 307) asserts, "One might argue that if the DSF is to be used for screening children who should be further evaluated for SED *(severe emotional disturbance)* diagnosis, then it would be desirable for the cutoff to have higher sensitivity. That is, a lower cutoff score should be used as far as it does not result in dramatic decrease of specificity rates."

(b) <u>What are the false positive and false negative rates for DSF Total cutoff scores of 0 to 2 SD above the mean?</u>

The present study calls specific attention to the incorrect classification rates that are associated with the DSF Total cutoff scores. At the DSF Total cutoff score of

114, described above as most efficient psychometrically due to the balance between true positive and true negative hit rates, there were false negative and false positive error rates of approximately 20%. This means that the DSF failed to correctly identify over 20% of students considered EI by the IEPC, and the DSF incorrectly identified another 20% of students as EI that were not EI according to the IEPC. These error rates seem high, and as was the case for hit rates, any reduction in one type of error rate (e.g., false negatives) by choosing a different cutoff score results in an increase in the other type of error rate (e.g., false positives). What are the "costs" of diagnostic errors (i.e., calling a student EI when he/she is not, and calling him/her "normal" when he/she is EI)? Again, most validity studies justify a cutoff score when the diagnostic hit rates and error rates are most balanced. And, Goh (1997) favors a lower cutoff score to yield higher sensitivity (true positives), even though that also yields more false positives (i.e., students being called EI when they are not), as well.

Obviously, there is more to this issue than just balancing sensitivity and specificity figures. Complicating the issue is the conflict between one's understanding of the philosophy underlying the IDEA law, and the philosophy underlying everyday practice. For instance, is the objective of IDEA to see that all EI students are identified, or to ensure that all students labeled EI and

receiving special education services are truly EI? If one thinks the former, one would be more willing to commit more false positive diagnostic errors. If one believes the former, one would be more willing to commit more false negative diagnostic errors. Similar conflicts arise in everyday practice, as practitioners weigh the relative merits of giving students the help they need, trying to disentangle the effects of social/cultural/economic factors for students exhibiting significant behavior problems, and allocating sparse resources in a cost-effective yet equitable manner. Again, in each of these cases, one would be willing to commit different diagnostic errors with instruments such as the DSF, depending on the practical issue at hand.

Further reason for caution in interpreting the nearly 80% hit rates found in the present study is provided by Reid and Maag (1994). They maintain that an accurate diagnosis does not automatically result from high levels of sensitivity and specificity. Indeed, because sensitivity and specificity consider only the rate of correct diagnoses, these indices are of questionable value when used with the general population in the case of low incidence disorders. Recall that 1998 data indicated that just over 1% of all students in the present study's intermediate school district were diagnosed EI, and just 8% of all special education students were EI. Reid and Maag proposed a formula that corrects for the rate of misdiagnoses by taking into account

the base rate: *sensitivity x base rate/(true positives + false positives)*. They noted that with a hypothetical rating scale with a high sensitivity of 95%, a high specificity of 95%, and a disorder prevalence (base rate) of 5% (e.g., ADHD), the probability of a correct positive diagnosis is only 50%. Using the present study's findings of a DSF sensitivity of 77.6%, a DSF specificity of 78.6%, and a base rate of EI within the special education population of only 8%, the probability of a correct positive diagnosis in the special population using the DSF is less than 25%! However, while we do have varying estimates of EI in the general population, we do not know what the true base rate is in the public school referral population. Many authorities (e.g., Forness & Knitzer, 1992) believe that EI is under-diagnosed in public school populations. Yet, many parents and professionals are reluctant to place an EI label on children. Nonetheless, Reid and Maag's argument is well taken that both rating scale diagnoses and the prevalence estimates derived from rating scales should be viewed with extreme caution.

Note, too, that the IDEA law does not require difficulties in all four subcategories of the EI definition. Instead, a student could be eligible under the EI category if significant problems are experienced in just one area. Thus, the DSF Total score is in many ways irrelevant, and the DSF Subscale scores are of greater diagnostic interest.

More discussion of the validity of the DSF Subscale scores follows later in this chapter.

(c) <u>What are the positive and negative predictive values for the DSF Total optimum cutoff score?</u>

Using this study's optimum total cutoff score of 114, 64% of students with high DSF scores were later found eligible as EI, while nearly 88% of students with low DSF scores were later found not-EI by the IEPC. These DSF positive and negative predictive values (like the sensitivity and specificity rates just discussed) also appear to have limited practical diagnostic significance. Gredler (1997) suggests that positive predictive values are the most important efficiency index, in that "school districts act mainly on the number of children who are classified as `at risk.'" He recommended that positive predictive values should exceed 65% in screening instruments for kindergarten readiness. Perhaps different guidelines are relevant for EI screening instruments. For instance, there are many potential sources of diagnostic errors (e.g., reluctance to identify LD/EI comorbidity, social maladjustment exclusionary criteria, etc.) among the 36% of students with high DSF scores who were not identified as EI. Also, the DSF would be just one part of a multi-method, multi-informant, multi-setting special education evaluation. On the other hand, given the potential ramifications often associated with an EI diagnosis (e.g., separate classroom placement, encouragement to seek psychiatric consultation,

etc.), one might also argue that the positive predictive values should be even higher for the DSF if it is to be used as a diagnostic instrument.

(d) **What are the sensitivity (true positive) and specificity (true negative) rates for DSF Subscale cutoff scores of 0 to 2 SD above the mean?**

The present study's DSF Subscale sensitivity (ranging from 63.2 to 77.6%) and specificity (ranging from 69.4 to 76.5%) rates provide the first empirical data regarding the DSF manual's recommendation for a cutoff value of approximately 1 SD above the mean for the four Subscales. That is, there was a balance between DSF Subscale sensitivity and specificity at approximately 1 SD above the mean for each of the four Subscales.

Note that students need to have significant problems in only one of the four areas to meet EI criteria. Thus, it is not unreasonable to expect that an EI student might have an elevated score on one Subscale, while having scores below the cutoff on the other Subscales. When the manual's recommended Subscale cutoff score of 13 is used, nearly 92% of the EI students had at least one high DSF score (sensitivity), as expected, and roughly 53% of the not-EI students had no high DSF Subscale scores (specificity), as expected. Scores were most likely to be elevated, as expected, for the EI students on the Interpersonal Problems (75.5%) and Depression (83.7%) Subscales. Scores were most likely to be below the cutoff, as expected, for the not-EI students on the Interpersonal Problems (76.5%),

Inappropriate Behaviors/Feelings (79.6%), and Physical Symptoms/Fears (79.6%) Subscales. These overall classification rates suggest respectable diagnostic utility of the DSF Subscales when used as a screening measure. Note, of course, that all of the cautions (e.g., costs of diagnostic errors depending on philosophy, impact of base rates for each of the EI subcategories) described above with regard to DSF Total cutoff scores apply to the Subscale cutoff scores as well.

Further evidence for the diagnostic utility of the DSF Subscales is suggested by the finding that the match between elevated Subscale scores and EI subpart eligibility was exact in 17% of the EI-only cases and was partially consistent in 72% of the cases. The remaining 10% of the EI-only cases did not have an EI subpart specified on the MET and IEPC. As the law only requires EI eligibility in one area, these findings suggest that in up to nine of ten cases, the DSF reflects at least some of the same social-emotional problems in EI students that the IEPC deems significant for eligibility purposes.

(e) <u>What are the false positive and false negative rates for DSF Subscale optimum cutoff scores?</u>

Results show that at the manual's recommended Subscale cutoff score of 13, approximately 47% of the students labeled not-EI by the IEPC had, contrary to expectations, at least one elevated DSF score (false positives). Scores were most likely to be elevated for the not-EI students on the

Depression Subscale (32.7%), and somewhat less likely (roughly 20%) on the other three Subscales. This finding could either be interpreted to mean that the DSF Depression Subscale incorrectly overidentifies depression in students, or that IEPC teams are unwilling to certify students as EI primarily on the basis of significant depressive symptoms.

Further, using the same Subscale cutoff score of 13, only 8% of the students labeled EI by the IEPC had, contrary to expectations, no elevated EI scores (false negatives). This is a very respectable false negative rate for a screening measure. Scores were least likely to be elevated for EI students on the Inappropriate Behaviors/Feelings (38.8%) and Physical Symptoms/Fears (40.8%) Subscales. Consequently, practitioners are advised not to interpret low scores on these two Subscales as evidence that a student does not have significant social-emotional problems.

(f) <u>What are the positive and negative predictive values for the DSF Subscale and optimum cutoff scores?</u>

Using the present study's psychometrically optimum Subscale cutoff scores of 12 to 14, roughly one-half to one-third of students with elevated DSF Subscale scores later were found eligible as EI by the IEPC, as expected (positive predictive value). More than four-fifths of students who had low DSF Subscale scores were later found not-EI by the IEPC, as expected (negative predictive value). Diagnostic errors were more frequent for students with high DSF scores than students with low DSF scores. As previously discussed,

diagnostic errors were most likely on the Inappropriate Behaviors/Feelings and Physical Symptoms/Fears. However, using the predictive value formulas, diagnostic errors this time were more likely with high DSF scores on these two Subscales. That is, roughly 45% of students with elevated scores on these two Subscales were not given an EI label by the IEPC, contrary to predictions. When these results are considered in tandem with those previously discussed, practitioners are strongly advised to treat all scores on the Inappropriate Behaviors/Feelings and Physical Symptoms/Fears with caution, and to use other assessment measures to more accurately assess those aspects of students' social-emotional functioning.

CHAPTER VI

SUMMARY AND IMPLICATIONS

This chapter contains a brief summary of the findings from the present study. Next, limitations of the present study are reported. In the final sections of this chapter, possible implications of this study on practice, policy, and future research are advanced.

Summary of Results

Based on the presentation and discussion of results discussed in previous chapters, the following brief summary of findings is presented:

1) Significant differences between the four IEPC eligibility groups (LD, EI, LD/EI and NE) were found on the DSF Total score and all four DSF Subscale scores.
2) Recombining the four IEPC eligibility groups into two new groups of all-EI (i.e., EI and LD/EI) and all not-EI (i.e., LD-only and NE) was statistically justified based on the DSF Total and Subscale score differences. Significant differences between these two recombined groups were also found on the DSF Total score and all four DSF Subscale scores, with the all-EI students consistently having higher scores.
3) Full sample correlations, corrected for attenuation, suggest that the Interpersonal Problems and Inappropriate Feelings/Behaviors subscales may be measuring essentially the same construct.

4) There were not significant differences between males and females on any of the DSF Total or DSF Subscale scores.

5) Significant differences between the Medicaid-eligible (low SES) and Medicaid-ineligible (moderate to high SES) groups were found on the DSF Total score, the Interpersonal Problems Subscale score, and the Inappropriate Behaviors Subscale scores. Scores for these two SES groups did not differ significantly on the Depression and Physical Symptoms/Fears Subscales.

6) A significant difference between ADHD-status groups (diagnosed, suspected, or nonexistent) was found only on the Inappropriate Behaviors/Feelings Subscale. On that Subscale, students with symptoms of ADHD had higher scores than students without symptoms of ADHD. The other DSF Total and Subscale scores did not differ significantly for the ADHD-status groups.

7) When students were reclassified on the basis of combinations of both EI-status and ADHD-status, significant differences were found on the DSF Total and all four DSF Subscale scores, primarily because of the earlier established impact of EI-status on DSF scores. The addition of ADHD-status to EI-status groups was most significant for the Inappropriate Behaviors/Feelings and Physical Symptoms/Fears Subscales.

8) Structure coefficients indicate all four DSF Subscales are useful as predictors of EI status. A discriminant analysis formula was presented which can be used to maximize the use of all four subscale standard scores, with over 85% accuracy for EI students and over 67% accuracy for non-EI students.

9) DSF Total score sensitivity (true positive) rates ranged from 87.8% to 26.5% at cutoff scores from 0 to 2 SD above the mean. DSF Total score specificity (true negative) rates ranged from 68.4% to 95.9% at cutoff scores from 0 to 2 SD above the mean. A balance between the two rates (sensitivity=77.6%; specificity=78.6%) was obtained when 114 (nearly 1 SD above the mean) was selected as the cutoff score.

10) DSF Total score false positive rates ranged from 31.6% to 4.1% at cutoff scores from 0 to 2 SD above the mean. DSF Total score false negative rates ranged from 12.2% to 73.5% at cutoff scores from 0 to 2 SD above the mean. A balance between the two rates (false positives=21.4%; false negatives=22.4%) was again obtained when 114 (nearly 1 SD above the mean) was selected as the cutoff score.

11) Using a DSF Total cutoff score of 114, positive predictive values were 64% and negative predictive values were 88%.

12) DSF Subscales scores sensitivity (true positive) rates ranged from 95.9% to 28.6% at cutoff scores from 0 to 2

SD above the mean. DSF Subscale scores specificity (true negative) rates ranged from 28.6% to 96.9% at cutoff scores from 0 to 2 SD above the mean. A balance between the two rates was found at cutoff scores of 13 for Interpersonal Problems (sensitivity=75.5%; specificity=76.5%), 12 for Inappropriate Behaviors/Feelings (sensitivity=69.4%; specificity=69.4%), 14 for Depression (sensitivity=77.6%; specificity=76.5%), and 12 for Physical Symptoms/Fears (sensitivity=63.2%; specificity=75.5%). At the Subscale cutoff score of 13 recommended in the manual, 91.8% (sensitivity) of EI students had a least one elevated Subscale score, as expected, and 53.1% (specificity) of not-EI students had no elevated Subscale scores, as expected.

13) Using optimum DSF Subscale cutoff scores of 12 to 14, false positive rates ranged from 10.2% (Inappropriate Behaviors/Feelings) to 38.8% (Depression), and false negative rates ranged from 14.3% (Depression) to 51% (Inappropriate Behaviors/Feelings and Physical Symptoms/Fears). Using the Subscale cutoff score of 13 recommended in the manual, 46.9% (false positives) of not-EI students had at least one elevated Subscale score, contrary to expectations, and only 8.2% (false negatives) of EI students had no elevated Subscale scores, contrary to expectations.

14) Using optimum DSF Subscale cutoff scores of 12 to 14, positive predictive values ranged from 53.1% to 62.3%, and negative predictive values ranged from 80.4% to 87.2%.

Limitations

Sampling Concerns

Several limitations to the present study's findings exist. For example, the students who comprise this sample are from five rural and suburban public schools in the Midwest, and thus the results may not be generalizable to students across the United States. Similarly, the 147 students in this study are only a small subsample (roughly 9%) of those referred and evaluated at this particular intermediate school district during a three year period, and thus the results may not be generalizable to all other special education referrals within that area. Further, this study focused only on students aged 5-12 years; the results may be quite different for the 13-18 year old age group. As discussed previously, older children typically have a number of general education teachers. The variation in older students' scores would be affected by the hour and content of each subject matter, personality of each teacher, etc. It also becomes more difficult as the years accumulate to disentangle learning from emotional problems (i.e., did six or more years of emotional problems lead to these learning difficulties, or vice versa?). Also, it was somewhat difficult to find sufficient students for the "referred-but-

not-eligible" subgroup because many of these types of students have already been screened and assisted through the pre-referral process. Combining these pre-referral students with the "referred-but-not-eligible" subgroup was considered, but parental permission for formal evaluation (such as behavior rating scales and special education personnel involvement) is not typically obtained during the pre-referral process. Reasons for the small samples sizes for the EI and LD/EI subgroups (e.g., parent/teacher/evaluator bias against using an EI label unless absolutely necessary for young children) have already been discussed, and also limit the generalizability of the present study's results. On a related note, the focus of this study was on the relationship of IEPC eligibility determination and/or ADHD-status to DSF scores. The effect of such factors as intelligence, academic achievement, age, gender, ethnicity, and socio-economic status on DSF scores, IEPC eligibility, and ADHD-status may have affected this study's results, but were not reviewed herein. Further, for the purposes of this study, ADHD-status was reliant on IEPC members noting such information on the IEPC paperwork. It is conceivable that there were students with ADHD for whom such information was not mentioned, thus affecting the sample composition of this study.

Rater Bias

Mention must also be made of another source of diagnostic error. When reviewing the profiles of LD-only

students with elevated scores, one case stood out. The teacher had rated each item as "rarely" or "occasionally" true for the kindergarten student under consideration. That is, no items were endorsed at either extreme ("never" or "frequently/very frequently"). In this particular case, the total score then fell within the "very significant" range, or 1 ½ SD above the mean. This case highlights a couple of limitations inherent in behavior rating scales in general. Namely, there is often a "horn effect" in that once a rater endorses one type of behavioral item, they are more likely to endorse other behavioral items (even if the student does not demonstrate those behaviors). Secondly, this example also appears to illustrate the response bias of "central tendency effects," due to the rater's selection of midpoint ratings and avoidance of the end points of the scale such as "never" or "very frequently." Third, this case illustrates the importance of assessing item level responses and not just total scores. In this case, one might ask if a student's emotional functioning is truly "very significantly" impaired when he/she only exhibits all behavioral problems rarely or occasionally? Note, too, that the DSF does not have any items which transform into validity scales, such as those on the Minnesota Multiphasic Personality Inventory (MMPI; in Butcher, 1990), which are designed to reflect response bias.

A related limitation is that teacher's ratings on such scales are affected by their personal levels of sensitivity

to and tolerance of children's behavior problems, their levels of experience with students, and their expectations regarding child behavior (Edelbrock, 1983). General education teachers who have referred students for special education services may be more likely to rate these students as having significant social-emotional problems, perhaps in part to justify the referral and increase the chances of the students being found eligible. Raters (as suggested by the case detailed above) also are likely to differ in their interpretation of the Likert-type frequency descriptors (e.g., *never, rarely, occasionally, frequently, and very frequently*) used on the Devereux-School Form (Reid & Maag, 1994). As Barkley (1987, p. 219) states, "Such scales, despite their apparent objectivity, are simply quantifications of adult opinions. As a result they are subject to the same sources of unreliability as those opinions."

Assessment Agreement

Another limitation involves the obvious fact that the DSF is just one piece of data when deciding special education eligibility, and thus a perfect correlation between scale scores and the IEPC eligibility decision is not expected. It is beyond the scope of this study to assess the degree of agreement between the teacher-rated DSF scale scores and other pieces of data (e.g., parent-rated DSF scale scores, standardized individual test results, student and parent interviews, other behavior rating scales,

direct observations, etc.) that also play an important role in the determination of the specific special education eligibility category.

Perceptions versus Behavior

The limitations of behavior rating scales, in general, also result in limitations for the present study. The teacher-rated DSF scales used in the present study, like all rating scales, only reflect teachers' perceptions of problems upon reflection, and not direct, frequency-count observations of students' actual behavior or social-emotional functioning. Relatedly, the DSF's use of a Likert-style frequency scale (i.e., *never, rarely, occasionally, frequently, very frequently*), with no quantifiable descriptors attached, are wide open for differences of interpretation of the amount of a behavior that corresponds to each of the frequency ratings. As has been asked by critics of behavior rating scales, only partly tongue-in-cheek, "Just how many `fidgets' are there in a `frequently?'" (e.g., Reid & Maag, 1994). Other rater factors also highlight the subjectivity of rating scales, and include teacher tolerance for disruptive/different behavior, own perceptions of self-confidence, quality and availability of assistance, and ability to manage students.

Diagnostic and Statistical Errors

The present study's reliance on sensitivity and specificity rates, while replicating prior DSF validity studies analyses, has limitations. As discussed in detail

previously, in cases of disorders with a low prevalence, such as EI, even behavior rating scales with excellent sensitivity and specificity may be of questionable value. This is because a focus on sensitivity/specificity figures might well cause one to overlook the rate of diagnostic errors, which are also of great clinical significance. Namely, in the case of "false positives," the DSF scores would be elevated for students who do not meet EI criteria. In the case of "false negatives," the DSF scores would be below the cutoff for students who do otherwise meet EI criteria. Note that the IDEA appears to favor not exluding any student who might possibly be EI from special education services, even if some non-EI students are inaccurately identified and serviced. Contrastingly, practitioners tend to favor erring in the opposite direction, emphasizing that no students are inaccurately labeled EI, even if some truly EI students are not identifed and served.

In terms of possible statistical errors, many significance tests were conducted in this study. The more such tests done, the greater the likelihood that some results will be significant by chance alone. One cannot know which findings are chance effects and which are not. The overall error in this study was controlled by having more confidence in those results that were significant at the .01 level (rather than the less conservative .05 level) as well as noting that the pattern of results were consistent and interpretable.

Multiple Gating

Results from the present study also do not reflect the fact that the referral-to-assessment-to-IEPC determination procedure is one of "multiple gating." As defined previously, the premise behind a multiple gating procedure is that through a sequential series of assessment and decision steps (or gates), a large population is gradually narrowed down into a smaller population most likely to have significant academic/behavioral problems (Merrell, 1994). The use of an instrument such as the DSF is only one such gate. Other decision points occur when teachers first decide which children to refer to a child study team, when child study teams decide which of these prereferral cases to refer for more involved testing and to whom, and when school psychologists and school social workers administer and interpret a number of instruments for each referred child within a multi-method, multi-informant, multi-setting assessment model. In the present study, the DSF was used within a fairly large population (all children referred for special education), and as such could be considered a screening instrument. A relatively high degree of diagnostic error (particularly high rates of false positives) is often tolerated in such a screening instrument, because students with elevated scores will be evaluated with more precise assessment techniques in subsequent "gates." In this light, then, the moderate to high rates of false positive diagnostic errors found in the

DSF for all referred students may be less alarming. In sum, multiple gating is an important assessment concept because it helps address issues related to time and cost effectiveness, as well as different levels of diagnostic error acceptability, depending on how far along an individual child is on the multiple gating assessment procedure.

Unreliability of Criterion Variables

Lastly, the present study's use of the IEPC eligibility decision (i.e., LD, EI, LD/EI, or NE) as the main criterion variable is a potentially major source of error. For example, the lack of reliable criteria for LD, EI, and LD/EI may be as big a problem as the errors inherent in the DSF. Many researchers (e.g., Merrell, 1994) have noted that the EI criteria (see Appendix G) may at first glance appear objective, but in fact a great deal of subjectivity is involved. For instance, what exactly do the statements "long period of time" and "to a marked degree" mean? Similarly, what are the objective criteria which would signify "`inappropriate' types of behavior or feelings under `normal' circumstances?"

ADHD status was another criterion variable. The reliance on notes within the central records files regarding ADHD status is potentially unreliable, as it is likely that an undeterminate number of students with ADHD were not so described on their IEPC paperwork. Further, a discussion of

the unreliability of ADHD diagnoses has been presented at length by other authors (e.g., Reid and Maag, 1994).

Crocker and Algina (1986) discuss additional practical problems in criterion-related validation. First, *small sample sizes*, such as those used in the present study, are much less likely to demonstrate acceptable validity levels than are larger samples (e.g., n=200 or more). Secondly, the issue of *criterion contamination* may have played a role in the present study, because in most cases the people who were able to influence the students' criterion scores (i.e., IEPC decision) also had access to students' scores on the predictor (i.e., DSF scores). In this case, it is likely that IEPC team members who were aware of elevated DSF scores would tend to use that information as evidence of an emotional impairment, and if there were low DSF scores, IEPC team members would tend to use that data as evidence that an emotional impairment was not present. Third, the *reliability of the predictor and criterion* is an issue because errors of measurement should be held to a minimum whenever one attempts to assess the degree of relationship between the predictor and criterion.

On a more critical note, Ysseldyke and Algozzine (1990) argue that IEPC special education classification is truly an arbitrary process, due only in part to the subjective nature of eligibility definitions in part. Also playing a role in the decision-making process, they assert, are the amount of special education funds available. That is, when funds are

plentiful, more lenient criteria tend to be used. Conversely, when funds are scant, more rigid criteria tend to be used. Similarly, in a recent work entitled *Special Education's Failed System: A Question of Eligibility*, Macht (1998, p. 3) contends, "The eligibility process used by schools to determine exceptionality and to finance special education assistance is inconsistent and often inaccurate." Not only do standards for LD and EI, for example, vary by state lines, the absence of uniformity can occur within an individual school system. For instance, different IEPC members "...can interpret observations differently and influence any number of decisions made by a school's eligibility team." Further, although there are criteria for the disability criteria, in the end the law allows the IEPC members' eligibility decision to be final, even if it does not provide evidence that the student met specific eligibility criteria. Obviously, the unreliability of IEPC eligibility determinations is not only a matter of concern for the present study, but for the practice of school psychology as a whole.

Implications

Bearing in mind the above limitations, the following implications are advanced for practice, policy, and future research:

Practice and Policy

This study is the first to provide validity support for the use of the DSF Total and Subscale scores in

differentiating between students referred for a special education evaluation and ultimately determined to have learning disabilities, emotional impairments, comorbid learning disabilities and emotional impairments, or academic/emotional problems not severe enough to qualify for special education. Thus, special education evaluators (primarily school psychologists and school social workers) can consider the present study's validity results in addition to those validity studies found in the DSF manual when selecting diagnostic instruments. The results suggest that the DSF is a diagnostically useful standardized screening instrument for teachers to complete early on in the assessment process for all students suspected of having significant learning and/or emotional disabilities that require special education support.

Four main implications for practice and policy can be advanced based on the present study's results. First, DSF score analyses provide evidence for the diagnostic difference between LD-only and LD/EI students. In fact, the DSF scores of the LD/EI students were consistently more similar to the EI students than to the LD-only students. Many evaluators are reluctant to use the LD/EI dual diagnosis, either from a legal standpoint of state and federal guidelines that can be interpreted as stating that LD and EI are mutually exclusive diagnostic categories, or from a bias against using an EI label if a less stigmatizing label (such as LD-only) will be enough to qualify a student

for special education services. Those reasons notwithstanding, the present study adds to a growing body of literature on the comorbidity of significant learning and emotional problems. At a minimum, it is suggested that the DSF (or some measure of social-emotional functioning) should be considered as one part of an assessment of all students suspected of having specific learning disabilities. Elevated DSF scores within the LD population (whether or not EI is officially added to these students' special education classification) are likely to add valuable information when developing an appropriate individualized education plan. At a broader policy level, perhaps it is time to revisit the state and federal guidelines dealing with the mutual exclusiveness of the LD and EI categories.

Second, as previously discussed, the mean DSF Depression Subscale scores for the two EI samples (EI-only and LD/EI) were within the "very significant" range. Implications of this finding for practice include, at a minimum, using the DSF to screen for depression in students referred for a special education evaluation, and following up with other depression inventories for students with elevated DSF Depression Subscale scores. Indeed, this is an example of an application of the multiple gating assessment model discussed previously. In addition, anecdotal data from the present study provides support for the documented tendency for childhood depression to be underestimated, as several participating evaluators commented that the DSF

Depression Subscales were higher than they had expected for the EI students in question. To address overall awareness issues, it is suggested that special education evaluators receive more preservice and inservice training on detecting childhood depression.

Third, the present study provides some evidence for the usefulness of the DSF when assessing the social-emotional needs of students who may have ADHD. Three of the four DSF Subscale scores and the DSF Total score were not significantly different when students were grouped according to ADHD-status alone, suggesting that the DSF is measuring social-emotional difficulties beyond those commonly associated with ADHD (i.e., inattention, impulsivity, and hyperactivity). Thus, practitioners may find the DSF a useful adjunct to their ADHD assessments, especially given the high rate of comorbid emotional impairments for students with ADHD. The finding that the DSF Inappropriate Behavior/Feelings Subscale was significantly different among ADHD-status groups implies that practitioners may want to consider an elevated score on that Subscale as an indicator to rule out ADHD prior to (or in addition to) an EI eligibility decision. Similarly, the anecdotal reports of several teachers that their ratings of students would have been different had they rated before a trial of medication was initiated suggests that the DSF may be sensitive to ADHD medication effects. If ADHD and medication are being considered during the course of an initial EI evaluation,

practitioners may want to have teachers rate these students with the DSF before and after the ADHD medication has been tried, in order to help make a differential diagnosis of EI and/or ADHD. Moreover, the DSF may be useful in studies designed to measure improvement due to treatment for ADHD.

Fourth, and most importantly, there are important practical implications of the present study's findings related to the DSF's classification accuracy. At first glance, the specificity (true negative) and sensitivity (true positive) rates for the DSF Total scores (nearly 80% each) and Subscale scores (ranging from 63 to 78% each) are respectable, particularly when compared to prior DSF validity study results. These classification accuracy figures typically determine which cutoff scores evaluators choose to use for a given scale. However, when the important variables of diagnostic misses (i.e., false negatives and false positives) and base rates of EI within the general and special education populations (1% and 8%, respectively, in the intermediate school district under study) are taken into account, the probability that a student with elevated DSF Total scores actually was EI fell to below 25%.

Moreover, the emphasis on the DSF Total score as an indicant of EI on this and other validity studies is very misleading, as students only need to show significant difficulties in at least one (not all four) subcategories of EI. Results from the present study suggest that there is

diagnostic utility for the DSF Subscales. When a Subscale cutoff score of 13 is used, nearly 92% of EI students had at least one elevated Subscale score, as expected. This means that only 8% of the EI students had no elevated DSF Subscale scores, which is a respectable false negative rate for a screening measure such as the DSF. Note that up to 40% of the EI students did not have elevated Inappropriate Behavior/Feelings or Physical Symptoms/Fears Subscale scores, suggesting that low scores on these two Subscales should not rule out EI eligibility in another area. In addition, roughly 53% of the not-EI students had no elevated Subscale scores as expected, meaning 47% of the not-EI students had at least one elevated Subscale score. Scores were most likely to be elevated for the not-EI students on the Depression Subscale (32.7%). This finding suggests that practioners need to decide on a case-by-case basis if the DSF is oversensitive to depressive symptoms, or if the DSF is accurately identifying depressive symptoms for which special educaiton services may or may not be needed.

The practical implication of these analyses is that data obtained from the DSF, and rating scales in general, need to be viewed cautiously. The proliferation of behavior rating scales for EI, of which the DSF is just one of many, and their aura of objectivity unfortunately in many instances have misled evaluators into accepting certain cutoff scores as diagnostic of EI. On the contrary, analyses contained herein have demonstrated that a diagnosis

based solely on a rating scale such as the DSF may, at times, be even less accurate than the flip of a coin. Regrettably, although behavior rating scales tend to be easiest assessment instruments to administer and score, many evaluators overestimate the significance of the scores obtained. A diagnosis of EI can be made with confidence only after other sources of information (e.g., structured diagnostic interviews, case histories, observational measures, functional behavioral analyses, and even medical evaluations in some instances) have been carefully considered in a multiple gating assessment model that emphasizes multi-method, multi-informant, multi-setting techniques. Reid and Maag's (1994, p. 349) warning bears repeating, "School psychologists must be wary of the seductive quality of (behavior rating scales') pseudo objectivity." Preservice and inservice training should more aggressively address this issue. In sum, while the DSF and other behavior rating scales may have their place in a multi-method assessment, they can never replace informed clinical judgment.

Future Research

This study investigated the discriminant validity of the DSF for a select group of initial special education referrals within a single Midwestern intermediate school district. Because any study of classification accuracy is sample specific, additional study is needed with other school samples. In particular, it would be helpful to

determine the DSF's clinical utility with other populations, such as: a more culturally diverse group; students aged 13 through 18 years; students with developmental delays; students who are suspected of being socially maladjusted (given the exclusionary criteria of social maladjustment within the EI definition); specific EI subtypes; and specific ADHD subtypes. Similar studies could also be done based on parent DSF ratings of initial special education referrals, as the present study focused only on teacher DSF ratings. Of course, given the discussion of the limits of specificity/sensitivity data, researchers are advised to take into account the base rates and diagnostic errors when reporting the classification accuracy rates for these proposed studies. More critically, if diagnostic errors are as high in subsequent studies as some were in the present study, the continued use of the DSF in practice and research will be difficult to justify.

With regard to diagnostic errors, further investigations into the sources of such errors on the DSF may be best accomplished through qualitative methods. Issues deserving further study include, to name a few: school personnel's willingness to identify EI in young students (particularly when other less stigmatizing options are available), interpretations of the social maladjustment and LD exclusionary criteria within the EI definition, and the impact of EI program availability on the rate of EI diagnoses.

Further research may also be warranted with regard to study participants' anecdotal comments that the DSF seems to be very sensitive (perhaps overly so) to depression in EI and LD/EI students. One possibility is to investigate the relationship between the DSF Depression Subscale scores with other respected, widely used measures of depression, such as the Children's Depression Inventory (Kovacs, 1983).

Interesting information may also be gained by exploring in a more qualitative way the impact of some current trends within education (e.g., inclusion, prereferral intervention, and "wrap-around" services) on the referral process, and the role the DSF may play in each (provided, of course, that more compelling evidence on the diagnostic utility of the DSF is forthcoming). One such study might investigate the merits of matching severity of EI (based on DSF scores) with the special education intervention implemented (e.g., inclusion through co-teaching and/or a personal aide in general education, resource room, categorical room, etc.). Another study could focus on the clinical utility of the DSF in identifying problems and guiding interventions that are addressed through the prereferral intervention and/or "wrap-around" processes. A related research question could determine if the interventions implemented (based, in part, on DSF scores) ultimately reduces the number of students formally referred for and in receipt of special education services.

Lastly, exploring the political, social, and economic issues embedded in these research questions is imperative if we are to improve upon our ability to meet the needs of all students. For example, as Mamlin and Harris (1998) so poignantly write, the whole issue of just what is a disability is called into question when study after study indicates that a large proportion of students referred for and found eligible for special education services are those from impoverished families who do not work with school personnel to address their academic, social, and behavioral needs. Similarly, Macht (1998, p. 203) makes a compelling case for a "problem directed, solution-oriented, and child focused" assistance delivery model that does not require "artificial justification, manipulated formulas, categorical labels, suggestions of pathology, or excruciating weeks of waiting." In conclusion, future research which focuses on expanding the service-as-needed option for all students, their teachers and schools, and their parents, will be involved in a worthy endeavor.

APPENDIX A

DEVEREUX BEHAVIOR RATING SCALE-SCHOOL FORM (DSF)

The 1993 edition of the Devereux Behavior Rating Scale-School Form (DSF) was used in this research project. It is a copyrighted, commercially available instrument that can be purchased from the Psychological Corporation at 555 Academic Court, San Antonio, Texas 78204-2498. The Psychological Corporation may also be reached by phone (210-299-1061) or by fax (210-270-0327).

APPENDIX B

THE PSYCHOLOGICAL CORPORATION STUDY PERMISSION LETTER

THE
PSYCHOLOGICAL
CORPORATION®

The Psychological Corporation
555 Academic Court
San Antonio, Texas 78204-2498
Tel 210-299-1061
Telex 5106015629 TPCSAT
Fax 210-270-0327

January 29, 1996

Ms. Barbara Sullivan Dunn
6464 Peck Road
Eaton Rapids, MI 48827

Dear Ms. Dunn:

Thank you for your letter concerning your use of the *Devereux Behavior Rating Scale - School Form* in your dissertation research.

As a responsible test publisher, we believe it is our duty to protect the security and integrity of our test instruments. Therefore, we cannot allow copies of the test instrument to be included with or stapled in your dissertation. If available, sample items may be included, but actual test items cannot. Also, all testing must be conducted in your presence or that of another qualified individual so that all test materials remain secure.

We will gladly grant permission for the use of this test instrument if the above restrictions will be followed. Please indicate your agreement to these terms by signing and returning this letter for our files. When you sign and return this letter, you may contact Ms. Sarah Sanchez in Customer Service at (800) 228-0752, ext.5427, to order your materials. If you have already placed an order, it will be released upon receipt of this signed letter. As a student, you are eligible for a 50% discount on the purchase of materials; however, you must request the 50% student discount and pay for the materials yourself in order to receive it.

Also, please forward a copy of your final dissertation for our library upon completion.

Thank you for your interest in our test materials. If you have further questions or needs, please contact us.

Sincerely,

Christine Doebbler
Manager
Legal Affairs

AGREED:

Barbara Sullivan Dunn

APPENDIX C

LETTER TO SCHOOL SOCIAL WORKERS

TO: **Eaton ISD School Social Workers**

FROM: **Barbara Sullivan**

RE: *Devereux Behavior Rating Scales* **Research Project**

DATE: **08-28-95**

As many of you know, I am in the process of completing my doctoral degree in school psychology at Michigan State University. My dissertation topic involves determining the usefulness of the new *Devereux Behavior Rating Scale-School Form* *(DSF)* during the initial evaluation process for students (aged 5-12 years) suspected of qualifying under the EI and/or LD categories. This project has been granted approval from Eaton Intermediate.

I would appreciate help from each of you in collecting the needed information for my project. This help would entail the following:

(1) For each **initial EI** evaluation (ages 5-12 years in preschool and elementary grades only), distribute one *DSF* to the student's general education teacher. For confidentiality purposes, upon receiving my copy of the form, I will substitute the student's name with his or her special education referral number, and later a code number.

(2) The (choose one):
* Send me the teacher-rated *DSF*, and I will score it and send the original back to you.
* OR: Score the teacher-rated *DSF*, and send a copy to me.

Please note that for those of you who did not use the *DSF* last year, and/or would not normally use it during you evaluations were it not for this project, I will supply you with sufficient protocols for your initial evaluations during the course of my study.

I think it will be interesting to see how useful the *DSF* is for our initial evaluations, and look forward to sharing the results with you once my project is completed.

Please feel free to contact me with any questions or concerns.

APPENDIX D

LETTER TO SCHOOL PSYCHOLOGISTS

TO: **Eaton ISD School Psychologists**

FROM: **Barbara Sullivan**

RE: *Devereux Behavior Rating Scales* **Research Project**

DATE: **08-28-95**

As many of you know, I am in the process of completing my doctoral degree in school psychology at Michigan State University. My dissertation topic involves determining the usefulness of the new *Devereux Behavior Rating Scale-School Form (DSF)* during the initial evaluation process for students (aged 5-12 years) suspected of qualifying under the EI and/or LD categories. This project has been granted approval from Eaton Intermediate.

I would appreciate help from each of you in collecting the needed information for my project. This help would entail the following:

(1) For each **initial LD** evaluation (ages 5-12 years in preschool and elementary grades only), distribute one *DSF* to the student's general education teacher. If a school social worker is also on the referral, you do not distribute the scale. For confidentiality purposes, upon receiving my copy of the form, I will substitute the student's name with his or her special education referral number, and later a code number.

(2) The (choose one):
* Send me the teacher-rated *DSF*, and I will score it and send the original back to you.
* OR: Score the teacher-rated *DSF*, and send a copy to me.

Please note that for those of you who did not use the *DSF* last year, and/or would not normally use it during you evaluations were it not for this project, I will supply you with sufficient protocols for your initial evaluations during the course of my study.

I think it will be interesting to see how useful the *DSF* is for our initial evaluations, and look forward to sharing the results with you once my project is completed.

Please feel free to contact me with any questions or concerns.

APPENDIX E

EATON INTERMEDIATE SCHOOL DISTRICT
STUDY PERMISSION LETTER

EISD

Eaton Intermediate School District • 1790 East Packard Highway • Charlotte, Michigan 48813

September 8, 1995

To Whom it may concern:

Barbara Sullivan Dunn is employed as a school psychologist at Eaton Intermediate School District, and has requested permission to conduct doctoral dissertation research using the *Devereux Behavior Rating Scale-School Form* (*DBRS-SF*).

Briefly, the proposed study entails having Eaton ISD school social workers and school psychologists distribute a *DBRS-SF* to the general education teacher of students (ages 5-12 years) referred for an initial special education evaluation due to a suspected learning disability and/or emotional impairment.

Confidentiality will be ensured. By using a research code system and deleting students' names, no one will be able to associate scale responses or other data with individual students during the data analysis phase of the project. Students' identities will be kept confidential and reports of research findings will not permit associating students with specific *DBRS-SF* responses or profiles or other research findings. All rating scales in the possession of Ms. Dunn will be kept in locked file cabinets in a locked office during non-office hours.

As the purpose of the proposed study is to validate an instrument used by Eaton ISD evaluators, the data may be collected in the manner detailed above without obtaining additional parental consent beyond that already secured for the initial referral itself.

Please feel free to contact us at (517) 543-5500 or (517) 484-2929 with any questions or concerns.

Wayne Buletza

Dr. Wayne Buletza, Ph.D.
EISD Director of Special Education

Jo Gager

Ms. Jo Gager
EISD Special Education Monitor

/chm

Eaton Intermediate is an equal opportunity employer that offers student programs and services without regard to sex, race, creed, national origin or handicap.

APPENDIX F

MICHIGAN STATE BOARD OF EDUCATION (1997)
LEARNING DISABILITY CRITERIA

R 340.1713 "Specific learning disability" defined; determination

Rule 13. (1) "Specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell , or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain disfunction, dyslexia, and developmental aphasia. The term does not include children who have leaning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, of autism, or of environmental, cultural, or economic disadvantage.

(2) The individualized educational planning committee may determine that a child has a specific learning disability if the child does not achieve commensurate with his or her age and ability levels in one or more of the areas listed in this subrule, when provided with learning experiences appropriate for the child's age and ability levels, and if the multidisciplinary evaluation team finds that the child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas:

(a) Oral expression.
(b) Listening comprehension.
(c) Written expression.
(d) Basic reading skill.
(e) Reading comprehension.
(f) Mathematics calculation.
(g) Mathematics reasoning.

(3) The individualized educational planning committee shall not identify a child as having a specific learning disability if the severe discrepancy between ability and achievement is primarily the result of any of the following:

(a) A visual, hearing, or motor handicap.
(b) Mental retardation.
(c) Emotional disturbance.
(d) Autism.
(e) Environmental, cultural, or economic disadvantage.

(4) A determination of impairment shall be based upon a comprehensive evaluation by a multidisciplinary evaluation team, which shall include at least both of the following:

(a) The child's regular teacher or, if the child does not have a regular teacher, a regular classroom teacher qualified to teach a child of his or her age, or for a child of less than school age, an individual qualified by the state educational agency to teach a child of his or her age.

(b) At least one person qualified to conduct individual diagnostic examinations of children, such as a school psychologist, a teacher of speech and language impaired, or a teacher consultant.

APPENDIX G

MICHIGAN STATE BOARD OF EDUCATION (1997)
EMOTIONAL IMPAIRMENT CRITERIA

MICHIGAN STATE BOARD OF EDUCATION

REVISED ADMINISTRATIVE RULES FOR SPECIAL EDUCATION -- 1997

R 340.1706 Determination of emotionally impaired.

Rule 6. (1)The emotionally impaired shall be determined through manifestation of behavioral problems primarily in the affective domain, over an extended period of time, which adversely affect the person's education to the extent that the person cannot profit from regular learning experiences without special education support. The problems result in behaviors manifested by one or more of the following characteristics:

(a) Inability to build or maintain satisfactory interpersonal relationships within the school environment.

(b) Inappropriate types of behavior or feelings under normal circumstances.

(c) General pervasive mood of unhappiness or depression.

(d) Tendency to develop physical symptoms or fears associated with personal or school problems.

(2) The term "emotionally impaired" also includes person who, in addition to the above characteristics, exhibit maladaptive behaviors related to schizophrenia or similar disorders. The term "emotionally impaired" does not include persons who are socially maladjusted, unless it is determined that such persons are emotionally impaired.

(3) The emotionally impaired shall not include persons whose behaviors are primarily the result of intellectual, sensory, or health factors.

(4) A determination of impairment shall be based on data provided by a multidisciplinary team, which shall include a comprehensive evaluation by both of the following:

(a) A psychologist or psychiatrist.

(b) A school social worker.

(5) A determination of impairment shall not be based solely on behaviors related to environmental, cultural, or economic differences.

LIST OF REFERENCES

LIST OF REFERENCES

Achenbach, T.M. (1991). Integrative guide for the 1991 CBCL/4-18, YSR, and TRF Profiles. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M., & Edelbrock, C. (1986). Manual for the Teacher's Report Form. Burlington, VT: University of Vermont, Department of Psychiatry.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1996). Standards for educational and psychological testing. Washington: American Psychological Association.

American Psychiatric Association (19870. Diagnostic and statistical manual of mental disorders (third edition, revised). Washington, DC: Author.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders (fourth edition). Washington, DC: Author.

Anderson, K.G. (1997). Gender bias and special education referrals. Annals of Dyslexia, 47, 151-162.

Barkley, R.A. (1987). The assessment of attention deficit-hyperactivity disorder. Behavioral Assessment, 9, 201-233.

Barkley, R.A. (1993). The handbook of attention deficit hyperactivity disorder. New York: The Guilford Press.

Bursuck, W. (1989). A comparison of students with learning disabilities of low achieving and higher achieving students on three dimensions of social competence. Journal of Learning Disabilities, 22, 188-194.

Bussing, R., Zima, B.T., Belin, T.R., & Forness, S.R. (1998). Children who qualify for LD and SED programs: Do they differ in level of ADHD symptoms and comorbid psychiatric conditions? Behavioral Disorders, 23(2), 85-97.

Butcher, J.N. (1990). The MMPI-2 in psychological treatment. New York: Oxford University Press.

Clarizio, H.F, & Higgins, M.M. (1989). Assessment of severe emotional impairment: Practices and problems. Psychology in the Schools, 26, 154-162.

Committee to Develop Standards for Educational and Psychological Testing. (1996). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Compas, B.E. (1997). Depression in children and adolescents. In E.J. Mash and L.G. Terdal (Eds.) Assessment of childhood disorders, Third Edition, 197-229.

Conners, C.K. (1990). The Conners Rating Scales Manual. North Tonawanda, NY: Multi-Health Systems.

Costello, E.J., & Angelo, A. (1995). Developmental epidemiology. In D. Cicchetti and D.J. Cohen (Eds.) Developmental psychopathology, Volume 1, Theory and Methods, 23-56.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Fort Worth, IL: Holt, Rinehart & Winston, Inc.

Edelbrock, C. (1983). Problems and issues in using rating scales to assess child personality and psychopathology. School Psychology Review, 12, 293-299.

Education for All Handicapped Children Act of 1975. (Public Law 94-142), 300, 34 U.S.C. 541 (1984).

Elliot, S.N., Busse, R.T., & Gresham, F.M. (1993). Behavior rating scales: Issues of use and development. School Psychology Review, 22, 313-321.

Forness, S.R., & Knitzer, J. (1992). A new proposed definition and terminology to replace "Serious Emotional Disturbance" in Individuals with Disabilities Education Act. School Psychology Review, 21, 12-20.

Glutting, J.J., McDermott, P.A., Watkins, M.M., and Konold, T.R. (1997). The base rate problems and its consequences for interpreting children's ability profiles. School Psychology Review, 26(2), 176-188.

Goh, D.S. (1995). Devereux Behavior Rating Scale-School Form. Journal of Psychoeducational Assessment, 13, 326-331.

Goh, D.S. (1997). Clinical utility of the Devereux Behavior Rating Scale-School Form among culturally diverse children. Psychology in the Schools, 34(4), 301-308.

Gredler, G.R. (1997). Issues in early childhood screening and assessment. Psychology in the Schools, 34(2), 99-106.

Gross-Tsur, V., Shalev, R.S., Manor, O., & Amir, N. (1995). Developmental Right-Hemisphere Syndrome: Clinical spectrum of the nonverbal learning disability. Journal of Learning Disabilities, 28, 80-86.

Haager, D., & Vaughn, S. (1995). Parent, teacher, peer and self-reports of the social competence of students with learning disabilities. Journal of Learning Disabilities, 28, 205-215, 231.

Hall, C.W., & Haws, D. (1989). Depressive symptomatology in learning-disabled and nonlearning-disabled students. Psychology in the Schools, 26, 359-364.

Handwerk, M.L., & Marshall, R.M. (1998). Behavioral and emotional problems of students with learning disabilities, serious emotional disturbance, or both conditions. Journal of Learning Disabilities, 31, 327-338.

Harnadek, M.C.S., & Rourke, B.P. (1994). Principle identifying features of the syndrome of nonverbal learning disabilities in children. Journal of Learning Disabilities, 27, 144-154.

Hutton, J.B., Dubes, R., & Muir, S. (1992). Assessment practices of school psychologists: Ten years later. School Psychology Review, 21, 271-284.

Individuals with Disabilities Education Act of 1990. (Public Law 101-476), 307, 20 USA, 1400 1400 et seq. (1990).

Kamphaus, R., & Frick, P. (1996). Clinical assessment of child and adolescent personality and behavior. Boston: Allyn & Bacon.

Kavale, K.A., & Forness, S.R. (1996). Social skills deficits and learning disabilities: A meta-analysis. Journal of Learning Disabilities, 29, 226-237.

Kovacs, M. (1983). Children's Depression Inventory. Pittsburgh, PA: University of Pittsburgh School of Medicine.

Landau, S., Milich, R., & Widiger, T. (1991). Predictive power methods may be more helpful for making a diagnosis than sensitivity and specificity. Journal of Child and Adolescent Psychopharmacology, 342-351.

Little, S. (1993). Nonverbal learning disabilities and socioemotional functioning: A review of recent literature. Journal of Learning Disabilities, 26, 653-665.

Macht, J. (1998). Special education's failed system: A question of eligibility. Westport, CT: Bergin & Garvey.

Mamlin, N., & Harris, K.R. (1998). Elementary teachers' referral to special education in light of inclusion and prereferral: Every child is here to learn ... but some of these children are in real trouble. Journal of Educational Psychology, 30(3), 385-396.

Martin, R., Hooper, S., & Snow, J. (1986). Behavior rating scale approaches to personality assessment in children and adolescents. In H. Knoff (Ed.), The assessment of child and adolescent personality (pp. 309-351). New York: The Guilford Press.

McConaughy, S.H. (1993). Advances in empirically based assessment of children's behavioral and emotional problems. School Psychology Review, 22, 285-307.

McMahon, R. J. (1984). Behavioral checklists and rating scales. In T. H. Ollendick & M. Herson (Eds.), Behavioral assessment of childhood disorders (2nd ed., pp. 105-153). New York: Guilford Press.

Merrell, K.W. (1994). Assessment of behavioral, social, and emotional problems: Direct and objective methods for use with children and adolescents. White Plains, NY: Longman Publishing Group.

Michigan State Board of Education (1997). Revised Administrative Rules for Special Education.

Milich, R., Widiger, T., & Landau, S. (1987). Differential diagnosis of attention deficit and conduct disorders using conditional probabilities. Journal of Consulting and Clinical Psychology, 55, 760-767.

Naglieri, J.A., Bardos, A.N., & LeBuffe, P.A. (1995). Discrimant validity of the Devereux Behavior Rating Scale-School Form for students with serious emotional disturbance. School Psychology Review, 24, 104-111.

Naglieri, J.A., & Flanagan, D.P. (1993). Psychometric characteristics of commonly used behavior rating scales. Comprehensive Mental Health, 2, 225-239.

Naglieri, J.A., & Gottling, S.H. (1995). Use of the Teacher Report Form and the Devereux Behavior Rating Scale-School Form with learning disordered/emotionally disordered students. Journal of Clinical Child Psychology, 24, 71-76.

Naglieri, J.A., LeBuffe, P.A., Pfeiffer, S.I. (1993). Devereux Behavior Rating Scale-School Form. Devon, PA: The Devereux Foundation.

Reid, R., & Maag, J.W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. Journal of School Psychology, 32, 339-354.

Reynolds, C.R., & Kamphaus, R.W. (1992). Manual: Behavior Assessment System for Children. Circle Pines, MN: American Guidance Service.

Rock, E.E., Fessler, M.A., & Church, R.P. (1997). The concomitance of learning disabilities and emotional/behavioral disorders: A conceptual model. Journal of Learning Disabilities.

Sattler, J.M. (1988). Assessment of children. San Diego: Author.

Smith, T.E.C., Dowdy, C.A., Polloway, E.A., & Blalock, G.E. (1997). Children and adults with learning disabilities. Boston, MA: Allyn and Bacon.

Smith, C.R., Frank, A.R., & Snider, B.C.F. (1984). School psychologists' and teachers' perceptions of data used in the identification of behaviorally disordered students. Behavioral Disorders, 10, 27-32.

Spivak, G., & Spotts, J. (1966). Devereux Child Behavior Rating Scale. Devon, PA: The Devereux Foundation.

Spivak, G., Spotts, J., & Haimes, P.E. (1967). Devereux Adolescent Behavior Rating Scale. Devon, PA: The Devereux Foundation.

Stahl, N.D. & Clarizio, H.F. (1999). Conduct disorder and comorbidity. Psychology in the Schools, 36(1), 41-50.

Tur-Kaspa, H., & Bryan, T. (1995). Teachers' ratings of the social competence and school adjustment of students with LD in elementary and junior high school. Journal of Learning Disabilities, 28, 44-52.

Vaughn, S., & Haager, D. (1994). Social competence as a multifaceted construct: How do students with learning disabilities fare? Learning Disability Quarterly, 17, 253-266.

Wright-Strawderman, C., & Watson, B.L. (1992). The prevalence of depressive symptoms in children with learning disabilities. Journal of Learning Disabilities, 25, 258-264.

Ysseldyke, J.E., & Algnozzine, B. (1990). Introduction to special education. Boston, MA: Houghton Mifflin Company.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02092 9588