The Diagnostic Validity of the
Draw A Person: Screening Procedure for Emotional Disturbance
and the Devereux Behavior Rating Scales

DISSESSATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Ted James Andrews, B.A.

**********
The Ohio State University
1993

Dissertation Committee:
Jack A. Naglieri, Ph.D.
Linda Perosa, Ph.D
Michael W. Vasey, Ph.D
Pamela D. Warrick, Ph.D.

Approved by

Advisor
College of Education
To My Parents -

The Greatest Contributors To My Education

and

Sandra Hooge -

Educator, Humanitarian, and Friend
ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Dr. Jack A. Naglieri for his guidance and support throughout this research. Thanks go to the other members of my advisory committee, Drs. Linda Perosa, Michael W. Vasey, and Pamela D. Warrick. The technical assistance of Dr. Dawn Flanagan and Ms. Denise Broadman is gratefully acknowledged. To my fiancee, Laura, I offer sincere thanks for her never ending support and words of encouragement throughout this endeavor. She is truly my best friend.
VITA


1989..............................B.A., Psychology
State University
of New York,
College at Fredonia

1990-1993.........................Psychology Assistant,
The OSU
Psychoeducational Clinic

1993-Present......................Adjunct Professor
Psychology,
Columbus State Community
College

FIELDS OF STUDY

Major Field: Education

Studies in School Psychology and Child Clinical
Psychology
TABLE OF CONTENTS

DEDICATION......................................................... ii
ACKNOWLEDGEMENTS........................................... iii
VITA................................................................. iv
LIST OF TABLES.................................................... viii
LIST OF FIGURES.................................................. x

CHAPTER PAGE
I. INTRODUCTION............................................. 01

   Emotional Disturbance
      in the Schools............................................. 01
   Defining Emotional Disturbance......................... 02
   Rational and Purpose of The Study...................... 08
   Research Questions......................................... 08

II. REVIEW OF RELATED LITERATURE....................... 10

   The Nature of Emotional Disturbance
      in the Schools........................................... 10
   Behavior Rating Scales.................................... 15
   Child Behavior Checklist................................ 17
   Teacher's Report Form.................................... 19
   Direct Observation Form.................................. 20
   Revised Behavior Problem Checklist.................... 21
   Conners' Parent Rating Scale - 48....................... 24
   Conners' Teacher Rating Scale - 39..................... 25
   Shortened Version......................................... 26
   Devereux Behavior Rating Scale -
      School Form............................................ 26
   Devereux Scales of Psychopathology..................... 28
   Conclusions................................................. 31
   Projectives................................................. 35
   Projective Hypothesis of Behavior....................... 36
   Projective Assessment Research......................... 37
Diagnosis | 38
Clinical versus Normal | 38
Comparisons Between Diagnostic Groups | 39
Projective Assessment in the Schools | 41
Verbal Techniques | 42
Visual Techniques | 44
Drawing Techniques | 46
An Objective Drawing Measure | 49
Objective Techniques | 50
Anxiety Scales | 51
Depression Scales | 53
Children's Apperceptive Storytelling Test | 53
Compatibility of Projective Assessment and School Psychologists | 55
Conclusions | 58
DSM Structured Diagnostic Interviews | 60
Psychometric Properties of the DSM | 61
Diagnostic Interview Schedule for Children | 63
Diagnostic Interview for Children and Adolescents | 66
Disruptive Behavior Disorders Rating Scale | 68
DSM in the Schools | 70
Conclusions | 73

III. METHODOLOGY | 75
Subject Selection | 75
Procedure | 77
Assessment Measures | 78
Statistical Transformations | 80
Data Analysis | 81

IV. RESULTS | 85
Summary of Descriptive Statistics | 85
Summary of Multivariate Analysis | 92
Description of Discriminant Analysis | 93
Summary of Discriminant Analysis | 95
Description of Analysis of Sensitivity & Specificity | 108
Summary of Sensitivity and Specificity
  Analysis........................................ 109
  Summary of Results........................... 137

V. DISCUSSION..................................... 143
  Summary of Study............................... 143
  Within Group Differences: DSF............... 146
  Within Group Differences: DSF.............. 147
  Sensitivity and Specificity................... 149
  Limitations of the Present Study............ 151
  Suggestions for Future Research............... 153

LIST OF REFERENCES............................... 155
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age, Race, and Gender Characteristics of the Clinical Group</td>
<td>86</td>
</tr>
<tr>
<td>2. Total Test Means, Standard Deviations, and Minimum/Maximum Scores by Age and Sex for the Clinical and Contrast Groups</td>
<td>87</td>
</tr>
<tr>
<td>3. DSM-III-R Diagnoses of Subjects in Clinical Group (N=142)</td>
<td>88</td>
</tr>
<tr>
<td>4. Obtained Pearson r Correlations Between DAP:SPED, DSP, and DSF Factor, Scale Composite and Total Scores for the Total Sample (N = 301)</td>
<td>91</td>
</tr>
<tr>
<td>5. Summary of Multivariate Analysis of DAP:SPED, DSP, and DSF Total Test Scores by Group</td>
<td>92</td>
</tr>
<tr>
<td>6. Unweighted Marginal Means From One-Way MANOVA of DSF 4 Scale Scores</td>
<td>95</td>
</tr>
<tr>
<td>7. Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSF Scale Scores</td>
<td>97</td>
</tr>
<tr>
<td>8. Summary of Univariate and Multivariate Analyses of DSF 4 Scale Scores</td>
<td>98</td>
</tr>
<tr>
<td>9. Unweighted Marginal Means From One-Way MANOVA of DSP 6 Factor Scores</td>
<td>100</td>
</tr>
</tbody>
</table>
10. Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSP Factor Scores.................... 102
11. Summary of Univariate and Multivariate Analyses of DSP 6 Factor Scores............ 103
12. Unweighted Marginal Means From One-Way MANOVA of DSP 3 Composite Scores............................... 104
13. Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSP Composite Scores............. 106
14. Summary of Univariate and Multivariate Analyses of DSP 3 Composite Scores........ 107
15. Significant Predictors of Discriminant Function................................. 139
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sensitivity and Specificity for the DSP Conduct Factor for Contrast and Clinical Groups Using Various Cut-off Scores</td>
<td>111</td>
</tr>
<tr>
<td>2. Sensitivity and Specificity for the DSP Anxiety Factor for Contrast and Clinical Groups Using Various Cut-off Scores</td>
<td>112</td>
</tr>
<tr>
<td>3. Sensitivity and Specificity for the DSP Depression Factor for Contrast and Clinical Groups Using Various Cut-off Scores</td>
<td>115</td>
</tr>
<tr>
<td>5. Sensitivity and Specificity for the DSP Acute Problem Factor for Contrast and Clinical Groups Using Various Cut-off Scores</td>
<td>117</td>
</tr>
<tr>
<td>7. Sensitivity and Specificity for the DSP Externalizing Factor for Contrast and Clinical Groups Using Various Cut-off Scores</td>
<td>122</td>
</tr>
</tbody>
</table>
8. Sensitivity and Specificity for the DSP Internalizing Factor for Contrast and Clinical Groups Using Various Cut-off Scores.............................................. 123


10. Sensitivity and Specificity for the DSP Total Score for Contrast and Clinical Groups Using Various Cut-off Scores.................................................. 126


12. Sensitivity and Specificity for the DSF Inappropriate Behaviors Scale for Contrast and Clinical Groups Using Various Cut-off Scores......................... 130

13. Sensitivity and Specificity for the DSF Depression Scale for Contrast and Clinical Groups Using Various Cut-off Scores................................. 131


15. Sensitivity and Specificity for the DSF Total Score for Contrast and Clinical Groups Using Various Cut-off Scores.................................................. 135

Chapter I
Introduction

Emotional Disturbance in the Schools

One of the important responsibilities of psychologists in the schools is the identification of children with emotional disturbance. Despite this responsibility, children with emotional disturbance are underidentified in American schools (Smith, Wood, & Grimes, 1988). According to Brandenburg, Friedman, & Silver (1990) conservative estimates indicate that 7% of all children may have emotional problems sufficient enough to warrant intervention. Forness, Bennett, & Tose (1983) have reported that 1/3 to 1/2 of children with an emotional disturbance in American schools have academic and/or related behavioral problems which require special education. The U.S. Department of Education indicates that less than 1% of all children enrolled in special education classes are identified under the federal category of Serious Emotional Disturbance (SED) (1991, cf. Forness & Knitzer, 1992). Given these statistics it
is apparent that the assessment and identification of children with emotional disturbance needs to become a clinical priority for psychologists.

**Defining Emotional Disturbance**

As one begins to read the research literature addressing the assessment of emotional disturbance in the schools it becomes readily apparent that there is a great deal of controversy as to what constitutes emotional disturbance. The current definition of emotional disturbance causing a great deal of debate among psychologists is put forth by the federal government and contained within Public Law 94-142. P.L. 94-142 was originally known as the Education of the Handicapped Act and has since been changed to the Individuals with Disabilities Education Act [IDEA] (Forness & Knitzer, 1992). IDEA defines a Serious Emotional Disturbance as

...a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree which adversely affects school performance: (a) an inability to learn which cannot be explained by intellectual, sensory, or health factors; (b) an inability to build or maintain satisfactory relationships with peers and teachers; (c) inappropriate types of behavior or feelings under normal circumstances; (d) a general pervasive mood of unhappiness or depression; or (e) a tendency to develop physical symptoms or fears associated with personal or school problems (Federal Register, 42, 474, 1977)

Bower has argued that the federal definition is a modification of his 1957 definition. These modifications, according to Bower, do "serious damage to the integrity of the
research and conceptual base from which the definition is drawn" (p. 55, 1982).

The federal definition combines a clinical, psychodynamic perspective of emotional disturbance with a school-related, behavioral perspective (Bower, 1982). Bower points out that the federal definition is seriously limiting in its recommendation that aggressive, acting-out children be distinguished from children with personality problems despite the fact that such a distinction is not possible in view of current psychological and educational research. Those children deemed to be acting-out are considered socially maladjusted and are not eligible for services under IDEA.

There is no conventional definition of what constitutes social maladjustment. Current practice in most states is to equate conduct disorder and juvenile delinquency with social maladjustment. Several authors have criticized this practice (Forness & Knitzer, 1992, Weinberg & Weinberg, 1990, Skiba & Grizzle, 1992, & Cheney & Simpson, 1990). Weinberg & Weinberg (1990) assert that "...a clear lack of familiarity with the research literature on the nature of both conduct disorder and juvenile delinquency is reflected in the work of those authors who automatically equate these
conditions with social maladjustment" (p. 157). Bower has stated that this definitional dilemma is a result of the fact that policy makers did not utilize the behavioral science research data in a consistent manner or with integrity. He further suggests that the federal definition emerged out of fiscal and resource limitations to serve children with emotional problems in the schools (Bower, 1982).

Forness & Knitzer (1992), as co-chairs of the Workgroup on Definition of the National Mental Health and Special Education Coalition, have proposed a new definition and terminology to replace the current federal label of Serious Emotional Disturbance. Their proposed plan replaces the term Serious Emotional Disturbance with "Emotional or Behavioral Disorder" and requires that the disability be consistently exhibited in two different settings, of which at least one is not school-related. The disability should adversely affect educational performance which includes academic, social, vocational, and personal skills. Emotional and behavioral disorders can co-exist with other disabilities and includes children or adolescents with schizophrenic disorders, affective disorders, anxiety disorders, or other
sustained disorders of conduct or adjustment which adversely influence educational performance.

Despite the efforts of some authors to modify or change the federal definition of Serious Emotional Disturbance, P.L. 94-142 remains the primary taxonomic system utilized by psychologists in the schools to place children in special education for a behavioral or emotional problem. While there are many psychological measures of emotional disturbance available to the psychologist working with children and adolescents, only two of these measures (Behavior Evaluation Scale and Emotional and Behavioral Problem Scale) are based on P.L. 94-142 criteria (Naglieri & Flanagan, 1993). The assessment efforts of psychologists working with children and adolescents suspected of having an emotional disturbance could be greatly enhanced by psychological measures based on P.L. 94-142 criteria. Naglieri, LeBuffe, & Pfeiffer (in press) have recently introduced a new behavior rating scale (the Devereux Behavior Rating Scale –School Form) which is based on the federal definition of Serious Emotional Disturbance.
As indicated previously, Bower (1982) has stated that the federal definition is a combination of psychodynamic and behavioral perspectives. School psychologists have available to them several objective and standardized behavior rating scales to assess students' observable behavior. Psychodynamic characteristics are typically assessed through the administration of objective personality measures and projective tests. Some objective personality measures have been shown to exhibit adequate psychometric properties, e.g. the Children's Depression Inventory (for a review of this measure see Worchel, 1990). However, these measures are often limited in their range of psychodynamic characteristics assessed, e.g. depression only. Traditional projective tests (e.g. Rorschach and House-Tree-Person) have been shown to have poor psychometric properties (see Gittelman, 1980, Peterson & Batsche, 1983; Worchel, 1990). Naglieri, McNeish, & Bardos (1991) have made available a standardized projective test (Draw A Person: Screening Procedure for Emotional Disturbance) with very good psychometric properties. The DAP:SPED is not designed as a differential diagnostic measure. It is intended as a quick, reliable, and objectively scored screening measure
for emotional disturbance. This measure can be used to identify children who are in need of further evaluation for possible placement in special education for a behavioral or emotional problem according to P.L. 94-142 diagnostic criteria.

Children who are placed in SED classes are often seen outside of the school by another mental health professional. These professionals do not utilize the P.L. 94-142 taxonomic system, but rather DSM-III-R (Diagnostic and Statistical Manual of Mental Disorders, Third Edition - Revised). While there are several structured diagnostic interviews based on DSM-III-R (e.g. Diagnostic Interview Schedule for children and Diagnostic Interview for Children and Adolescents), until recently, there were no behavior rating scales based solely on DSM-III-R criteria. Naglieri, LeBuffe, & Pfeiffer (in press) have developed the Devereux Scales of Psychopathology (DSP), a behavior rating scale based on DSM-III-R criteria. Such a measure can be utilized by both school psychologists and child clinical psychologists to enhance communication and coordinated care efforts between these two professionals.
Rational and Purpose of the Study

The purpose of the present study is to explore the diagnostic validity of three recently made available psychological measures of emotional disturbance for use with children and adolescents: the Draw A Person: Screening Procedure for Emotional Disturbance (DAP:SPED), the Devereux Behavior Rating Scale - School Form (DSF), and the Devereux Scales of Psychopathology (DSP). As indicated previously, each of these measures offers a unique contribution when compared to existing psychological measures of emotional disturbance.

Research Questions

The following research questions are proposed from the preceding discussion.

1. Will there be a significant difference between the mean scores earned by children in a clinical group and a contrast group on DAP:SPED, DSF, and DSP Total Scores?
2. What is the degree of discrimination of the DSF 4 Scale Scores, the DSP 6 Factor Scores, and the DSP 3 Composite Scores for a sample of children in a clinical group and a contrast group?

3. What are the percentages of true positives and true negatives for the DAP:SPED, DSP, and DSF Total Scores, at various cut-off scores, for a sample of children in a clinical group and a contrast group?
   a. What are the percentages of true positives and true negatives for the DSF 6 Factor scores, at various cut-off scores, for a sample of children in a clinical group and a contrast group?
   b. What are the percentages of true positives and true negatives for the DSF 4 Scale Scores, at various cut-off scores, for a sample of children in a clinical and group and a contrast group?
   c. What are the percentages of true positives and true negatives, for the DSP 3 Composite Scores, at various cut-off scores, for a sample of children in a clinical group and a contrast group?
Chapter II
Review of Related Literature

The types of emotional and behavioral disturbances seen in American school aged children and adolescents is discussed in this chapter. An overview of the currently used psychological assessment instruments subsumed under the general categories of projective tests, behavior rating scales, and DSM structured diagnostic interviews is also presented. The psychometric properties and diagnostic validity of these instruments is also discussed.

The Nature of Emotional Disturbance in the Schools

Costello & Janiszewski (1990) compared 89 treated and 126 nontreated school children aged 7 through 11 who scored in the 90th percentile or above on the Child Behavior Checklist (CBCL) on measures of psychopathology, environmental factors, and adaptive functioning. The treated children came from a psychiatric clinic at a Health Maintenance Organization (HMO). The untreated children came from a pediatric clinic at the same HMO. The subjects were assessed using the Diagnostic Interview
Schedule for Children (DISC). Results indicated no significant differences between the two groups in proportions receiving a psychiatric diagnosis, or in the proportions with more than one psychiatric diagnosis. Both groups were just as likely to have attention deficit disorders, anxiety disorders, and oppositional disorders, and to be failing in school. The treated group was twice as likely to have conduct and depressive disorders, and to be poor, male, and black. The most significant finding for psychologists and educators was that twice as many behavioral problems were reported by the teachers of children in the treated group than in the untreated group. The authors concluded that the adults' discomfort with these children's behavior may be more of a factor than the children's failure in school in making a referral for psychological/psychiatric assessment.

Mattison, Humphrey, Kales, Hanford, Finkenbinder, & Hernt (1985) assessed the psychiatric histories and made DSM-III diagnoses of 177 school children aged 6-18 years who were referred for possible inclusion in special education classes for children with social and emotional disturbances. All diagnoses and severity ratings by the authors were conducted with no knowledge of each child's subsequent class placement. These authors found a high
prevalence of psychopathology among all 177 school children. Attention Deficit Disorder was the most common diagnosis in children aged 6-12 years. Affective Disorders were the most common in older children (13-18 years). The psychiatric disorders were similar but of significantly greater severity in the children who were eventually placed in an SED class. It is possible that in the present study, the adult's discomfort with the behavior of the identified children may have been more of a factor than the children's failure in school in making a special education placement. This conclusion is similar to the conclusion drawn by Costello et al. (1990) mentioned above.

Students in a Learning Disability (LD) class and students in a class for children with emotional disturbance (ED) were the groups of comparison in a study by Harris, King, Reifler, and Rosenberg (1984). These authors administered the Teacher Report Form of the Child Behavior Checklist (CBCL-TRF) to the primary teachers of both groups of students. The behavioral profiles for both groups were similar. However, students in the ED class had significantly higher total scores, externalizing, internalizing, social-withdrawal, self-destructive, inattentive, and nervous/overactive scale
scores than the LD group did. The LD group, although less disturbed, fell in the 85th percentile for behavior disturbance. Similar findings regarding behavioral problems in children with a learning disability have been reported by Janke & Lee, 1991; Merrell, 1991; Raviv & Stone, 1991; Costerbader & Keller, 1990; Durrant, Cunningham, & Voelker, 1990). Based on their data, Harris et al. (1984) suggested that special services for behavior disorders are needed for LD groups and specific learning disability instruction is needed for ED groups. Further extending Costello and Janiszewski's (1990) aforementioned hypothesis, it is likely that the children with more severe behavior problems were placed in the ED class, while the more tolerable children, in terms of their behavior, were placed in the LD class.

Mattison, Morales, & Bauer (1991) examined the characteristics of elementary and secondary socially and/or emotionally disturbed girls referred for possible placement in SED classrooms from two schools. The authors were not aware of the students placement by the schools' multidisciplinary teams (MDT) until they had completed the assessment of the students. Compared to girls who were not subsequently placed in an SED class, the SED girls demonstrated a wide range of psychiatric
disorders, and had experienced limited intervention in the school and community. Multiple family stressors, especially abuse, was significantly more common in SED girls than it was in non-SED girls. Significantly more global dysfunction was noted in the SED group on the Schedule for Affective Disorders and Schizophrenia for School-age Children-Epidemiologic Version (SADS-E), the Conners' Teacher Rating Scale (CTRS), and the Child Behavior Checklist (CBCL). Classroom teachers rated the SED girls significantly higher than non-SED girls on three narrow-band factors: Hyperactivity, Conduct Disorder, and Emotional Overindulgent. The non-SED girls were higher on the CTRS Anxious-Passive factor. This finding was not significant. Anxious-passive behavior is often easier to tolerate and manage in a classroom setting than is hyperactive, conduct-disordered behavior and thus may have influenced the placement decisions of the MDTs.
All of the demographic studies mentioned above utilized behavior rating scales in their attempts to describe the characteristics of children placed in SED classes. The research literature on behavior rating scales is presented next.

Behavior Rating Scales

In an effort to improve the utility of diagnostic labels for childhood psychopathology the federal government sponsored the Project on the Classification of Exceptional Children (Hobbs, 1975a, 1975b, cf. Achenbach & Edelbrock, 1978). From this point on, researchers began to utilize advanced multivariate techniques to develop an empirically based child psychopathology taxonomy that attempted to avoid the problems inherent in the adult taxonomic system of psychopathology (Achenbach & Edelbrock, 1978). Behavior rating scales emerged from this research body and continue to receive much attention in the research literatures addressing the assessment and diagnosis of child psychopathology. McMahon defines a behavior rating scale as an assessment instrument that is completed by adults in reference to a child's behavior (1984).
Goh, Teslow, & Fuller conducted a national survey of school psychologists in 1981 and found that the most commonly used assessment instruments were intelligence, achievement, and perceptual-motor tests; little attention was given to the assessment of socio-emotional disorders or behavior problems. When school psychologists did evaluate socio-emotional functioning, they relied on clinical interviews, projective techniques, and direct observation of students more than behavior rating scales (Prout, 1983). Hutton, Dubes, & Muir (1992) used a modified version of the Goh, et al. (1981) survey to describe the assessment practices of school psychologists ten years later. These authors report that the use of intelligence tests has dropped somewhat (21.9% in their study v. 26.9% in the Goh, et al. (1981) study. The use of achievement tests has increased (32.8% v. 22.2% respectively). Furthermore, the use of personality measures (projectives) has dropped (11.5% v. 14.0%) while the use of behavior rating scales and adaptive behavior scales has increased (18.3 v. 8.6%). According to Merrell, (1993) "School-based assessment of student characteristics through the use of teacher rating scales has become a widely utilized form of measurement, with one of the most popular uses of rating scales being to
provide teacher ratings of student behavior" (p. 115). McConaughy (1985) has argued that in view of the poor psychometric properties of many projective techniques, empirically based behavior rating scales are an appropriate measure for school psychologists to employ when conducting an assessment for socio-emotional disturbance or behavior problems. The Child Behavior Checklist is one behavior rating scale which has received much attention in the assessment and diagnostic literatures.

**Child Behavior Checklist**

The Child Behavior Checklist (CBCL) (Achenbach, 1991a) is designed to obtain parents' reports of their children's social competence and behavior problems. It is intended to serve as one part of a multiaxial empirically based assessment (Achenbach, 1991). Separate versions of the CBCL exist for each sex, ages 4-18. The CBCL was originally normed on children and adolescents, ages 4-16. Additional norms were provided for adolescents, ages 17-18 in 1991. The social competence scales are comprised of 20 items which assess a child's activities, involvement with social organizations and friends, and school performance. Symptomatology is rated
on a 3 point scale (0, 1, or 2) and raw scores on each factor are transformed into T scores (mean 50, SD = 10). T Scores falling between 67 and 70 are said to be in the Borderline Clinical Range. T scores above 70 fall in the Clinical Range. The 118 items which comprise the behavior problem scales were derived from factor analyses of 2,300 CBCL's filled out by parents whose children were referred to 42 different child mental health clinics in the eastern United States (Achenbach & Edelbrock, 1983). Second-order factor analysis of the CBCL scores yielded two broad-band syndromes: internalizing and externalizing, and one mixed syndrome for each age and sex group. Principal component factor analysis produced eight narrow-band syndromes for each age and sex group: Withdrawn, Somatic Complaints, Anxious/Depressed (all loading on the internalizing broad-band), Social Problems, Thought Problems, Attention Problems (all comprising the mixed syndrome), Delinquent Behavior, and Aggressive Behavior, (both loading on the externalizing broad-band). A Sex Problems scale for children, ages 4-11, is also available (Achenbach & Edelbrock, 1983). Initial norms for both social competence and behavior problem scales were obtained from 1,400 randomly obtained home interviews in Washington, D.C., Maryland, and
northern Virginia. One-week test-retest and intraclass correlations for agreement between mothers' and fathers' ratings are .90 or above. Good discriminate validity has been shown for the CBCL with children referred for mental health services scoring significantly higher on all the behavior problem scales and significantly lower on the social competence scales (Achenbach & Edelbrock, 1983; Achenbach, 1991). The CBCL provides forms for teacher ratings, direct observation, and youth self-report.

**Teacher's Report Form**

The Teacher's Report Form (TRF) (Achenbach, 1991b) obtains teacher's ratings of children's academic performance, adaptive functioning, and behavior problems. The TRF also includes 118 behavior problem items, 91 of which are similar to or the same as items on the CBCL. Separate versions of the TRF exist for each sex, ages 4-18. The TRF was originally normed on children and adolescents, ages 4-16. Additional norms were provided for adolescents, ages 17-18 in 1991. Five TRF scales are similar to the CBCL; two overlap with CBCL scales; and three are unique to the TRF. As with the CBCL, second-order factor analysis produced two broad-band factors: externalizing and internalizing. Initial norms for the
adaptive functioning and behavior problem scales were obtained from 1,100 randomly selected non-referred children in regular education (Achenbach & Edelbrock, 1986). Good discriminate validity for the TRF is reported with referred children scoring significantly higher on all behavior problem scales, and lower on the adaptive functioning and school performance scales, than non-referred children (Achenbach & Edelbrock, 1986; Achenbach, 1991b).

Direct Observation Form

The observer using the Direct Observation Form (DOF) (Achenbach & Edelbrock, 1983) observes a child for 10 minutes and writes a narrative description of his or her observation. The DOF rating scale is kept in view during the observation period but the actual ratings are not made until the end of the observation period. There are 96 items on the DOF. Eighty-six of the items are similar or equal to 73 of the items on the TRF and 73 on the CBCL. Total behavior problem and on-task scores are computed. An advantage of the DOF is that the observer can take into account situational variables such as classroom activity, teacher behavior, and interactions with other children when interpreting observation results.
(McConaughy, 1985). No norms or subscales have been developed but McConaughy (1985) has pointed out that the two global scores can be compared to the average scores of a small sample of normal and referred children studied by Reed & Edelbrock (1983). The CBCL and its two extensions (the TRF and the DOF) is a psychometrically adequate means for identifying children with socio-emotional disturbance or behavior problems in the schools. Furthermore, the CBCL, TRF, and DOF provide three different sources of information (one of which provides contextual information - DOF) which can be utilized by psychologists in the identification of children with emotional and behavioral problems.

Revised Behavior Problem Checklist

The Revised Behavior Problem Checklist (RBPC) is an extension of the original Behavior Problem Checklist (Quay & Peterson, 1979). Because the BPC has been used in over 100 hundred research studies addressing epidemiology, screening, assessment, and the measurement of the effects of various interventions, the original item pool of the instrument was expanded (Quay, 1983). The RBPC has an augmented item pool of 89 items. The initial 55 item pool of the BPC was generated from a
factor analysis of the presenting complaints of various children seen at a child guidance clinic. The original analysis was conducted on a kindergarten through sixth grade sample of normal children. Four samples were used in the revision of the BPC. Sample 1 consisted of 276 patients in a private psychiatric hospital who were rated by staff members (mean age = 15), sample 2 consisted of 198 out- and inpatients at the same hospital who were rated by their parents at the time of admission (mean age = 11.6), sample 3 was comprised of 114 children attending a private school for children with learning disabilities who were rated by their teachers (mean age = 10), while the 4th sample was comprised of 172 children at a community-sponsored school for children with developmental disabilities who were also rated by their teachers (mean age not reported). Factor analysis of the RBPC yielded five major scales: Conduct Disorder (CD), Socialized Aggression (SA), Attention Problems-Immaturity (AP), and Anxiety-Withdrawal (AW). The AW scale contains the least amount of items (11) and the CD scale contains the most (22). Two minor scales also resulted from the re-factor analysis: Psychotic Behavior (PB) which contains 6 items and Motor Excess (ME) which contains only 5 items.
The CD, SA, AP, and ME scales are all positively intercorrelated, sharing about 27 to 30% of their variance. Quay (1983) reports that these 4 scales represent the major dimensions of "externalizing" psychopathology as introduced by Achenbach in 1966. High coefficients of internal consistency are reported for the RBPC with the average alpha coefficients = .94 for CD, .84 for SA, .92 for AP, and .82 for AW, .72 for PB and .76 for ME. Lower coefficients are to be expected for the last two scales because they have fewer items than the other scales (6 and 5 respectively). Average intercorrelations among raters range from a low of .52 (PB) to a high of .85 (CD). A discriminant function employing all six subscales correctly classified 85.5% of clinical cases versus students in regular education classes (Quay, 1983). For males, 77% of the clinical group was correctly identified with 23% false positives. In the regular class group 87% were correctly classified with 13% false positives. For females, 91% were correctly classified with 23% false negatives in the clinical group and only 7% false positives in the regular class group. The RBPC yields adequate psychometric data, is cost efficient, and requires little time to administer. As such, the RBPC can be a helpful addition
to the psychologists behavioral/emotional problem assessment battery.

Conners' Parent Rating Scale - 48

The Conners' Parent Rating Scale (CPRS) (Conners, 1990) consists of two versions, a 93 item version and a 48 item version. The 48 item version reports sex-by-age normative data and will be discussed here. The CPRS was designed to characterize patterns of child behavior (Conners, 1990). Normative data for the CPRS-48 is based on a study of 578 children, aged 3 to 17 years. Symptomatology is rated on a 4-point scale (0, 1, 2, or 3) and raw scores on each factor are transformed into T scores (mean = 50, SD = 10). Clinical studies have shown that a T score of 70 or greater (2 standard deviations above the mean) may indicate problem areas (Conners, 1990). Factor analysis of the CPRS-48 yielded five factors: Conduct Problem, Learning Problem, Psychosomatic, Impulsive-Hyperactive, and Anxiety. The CPRS-48 can be administered to school age children and adolescents between the ages of 3 and 17. Additional factor analyses of the CPRS-48 items have supported the original five factor structure. Conners (1990) reports
adequate psychometric properties for the CPRS-48. A teacher version is also available.

**Conners' Teacher Rating Scale - 39**

The Conners' Teacher Rating Scale (CTRS) (Conners, 1990) is also available in two versions, a 28 item version (for ages 3-17) and a 39 item version (for ages 4-12). The CTRS-39 has normative data available from a Canadian sample of 9,583 children and will be reported here. Separate norms for groups formed by age and gender are available for both the CTRS-28 and CTRS-39. Factor analysis of the CTRS-39 yielded 6 factors: Hyperactivity, Conduct Problem, Emotional Overindulgent, Anxious-Passive, Asocial, and Daydream-Attendance Problem. Symptoms are rated and scored in the same fashion as the CPRS-48. Conners (1985), Epstein & Nieminen (1983), and Schachar, Sandberg, & Rutter (1986) all report adequate psychometric properties of the CTRS-39. Additional factor analyses have supported the original 6 factor solution.
Shortened Version

A shortened version of Conners' scales is available. The scale, referred to as the Hyperactivity Index, is automatically scored from both forms of the CPRS and CTRS. According to Conners, "The Hyperactivity Index was developed to provide an easily measurable, empirical assessment of the extent to which the child performs behaviors which are usually considered as indicative of an underlying diagnosis of hyperkinesis" (1990, p. 2). Conners (1990) points out that this scale is best conceptualized as a general index of child psychopathology, and not as a syndrome specific to the diagnostic category of hyperactivity or attention deficit disorder. This scale is comprised of the most highly loaded symptoms from the factor scales. Norms are available in Sattler (1988).

Devereux Behavior Rating Scale - School Form

The Devereux Behavior Rating Scale - School Form (DSF) (Naglieri, LeBuffe & Pfeiffer, in press) consists of 40 items which were selected in accordance with the federal definition of Severe Emotional Disturbance. Individual items were placed into one of four scales which correspond to the federal definition: Interpersonal
Problems, Inappropriate Behaviors/Feelings, Depression, and Physical Symptoms/Fears. The sum of the 40 items yields a Total Test standard score with a mean of 100 and a standard deviation of 15.

The Devereux School Form was standardized in 1991 in over 30 cities across the United States. The standardization sample was comprised of 3,152 persons aged 5 to 18. Sixty percent of the ratings were obtained from parents and 40% were obtained from teachers of students in regular and special education classes, not including those children in SED classes. Naglieri et al. (1992) report internal reliability coefficients for ages 5-12 and 13-18. Parent's ratings of males aged 5-12 = .94, and .93 for 13-18 year old males. Parents' ratings of females aged 5-12 = .93, and .92 for 13-18 year old females. Teachers' ratings were .97 and .95 for males, and .97 and .96 for females. The median internal reliability for the four subtests across age is .83.

Naglieri & Gottling (1992) compared the Devereux School Form and the TRF (Achenbach & Edelbrock, 1983) with respect to their ability to determine the presence and severity of psychological problems in samples of learning disabled and nondisabled students. These authors examined the mean score differences between the
two groups using a $t$ test. The efficiency of
discrimination was determined by the calculation of $d$
ratios (the differences between the means divided by the
pooled standard deviations). The accuracy of
identification was examined via an analysis of
sensitivity (true positives) and specificity (false
negatives) as outlined by Verhulst & Koot (1992, cf.
Naglieri & Gottling, 1992). Results indicated that when
the cut-off score of 70 as suggested by Achenbach &
Edelbrock (1986, cf. Naglieri & Gottling, 1992) is
utilized, the TRF correctly identified 49% of the total
sample and the Devereux School form correctly identified
77%. Naglieri & Gottling (1992) concluded that
psychologists who need to assess for the prevalence of
emotional disturbance in students with learning
disabilities should find the Devereux School Form to be
an efficient tool in this process.

Devereux Scales of Psychopathology

The Devereux scales of psychopathology (DSP)
(Naglieri, LeBuffe, & Pfeiffer, in press) is comprised of
135 items for ages 5-12 years and 136 items for ages 13-
18 years. The items comprising both age versions overlap
by about 50%. On both versions theses items are broken
down into three levels. The first level is the Scale level. For the 5-12 year old version the scales are: Conduct, Attention, Withdrawal/Depression, Anxiety, Acute Problems, and Autistic. For the 13-18 year old version all the scales are the same except for the Attention scale. A Delinquency scale replaces the Attention scale on the 13-18 year old version. On both versions, the scales are divided into three Composite areas: Externalizing, Internalizing, and Psychotic. The three composites are added together to yield a Total Test Score. Naglieri, LeBuffe, & Pfeiffer (in press) indicate in their test manual that the DSP was developed to assist in the identification of behavior signs related to particular categories (P.L. 94-142 & DSM-III-R) of psychopathology.

The DSP was standardized in 16 states representing the Northeast, Midwest, South, and West regions of the United States in 1991. The standardization sample was comprised of 3,153 individuals aged 5-18 years. The sample was representative of the U.S. population in terms of age, sex, geographic region, race, socioeconomic status, ethnicity, community size, and educational placement. Naglieri et al. (in press) report total score internal reliability coefficients ranging from .966 to
.984. Scale internal reliability coefficients range from .699 (parent/female for ages 13-18 on factor 6) to .990 (parent/male for ages 13-18 on factor 1). The internal reliabilities for the Externalizing, Internalizing, and Psychotic composites ranged from .876 (parent ratings of females aged 5-12) to .984 (teacher ratings of females aged 5-12). Test-retest reliability coefficients are provided for intervals of 24 hours, 2 weeks, and 4 weeks in a study which utilized children who attended either a psychiatric day or residential treatment program at the Devereux Foundation. Test-retest reliability with a regular education sample over a one week interval yielded Total score coefficients of .85 for 5-12 year olds, and .78 for 13-18 year olds. The Total Score Intrarater reliability coefficient for teachers in one study was .775 and .566 for teacher aides. An inter-rater reliability study which compared the ratings of teachers and teacher aides for 45 children residing in an inpatient unit of a mid-Atlantic children's hospital yielded inter-rater reliabilities of .66 (Conduct), .55 (Attention), .48 (Anxiety), .52 (Depression), .46 (Autistic), .44 (Acute Problems), .61 (Externalizing), .44 (Internalizing), and .45 (Psychotic).
Six criterion-related validity studies are presented in the test manual. Naglieri et al. (in press) summarize these studies by reporting that the minimum percent of accuracy when the DSP is used to predict group membership (Normal v. clinical) at ages 5-12 was 58.0 in study #4, and the maximum was 78.1 in study #6; at ages 13-18 the minimum percent of accuracy was 65.5 in study #5, and the maximum was 89.7 in study #1. Naglieri et al. (in press) report that results of these analyses indicate that the DSP performed well in the identification of individuals with serious emotional disturbance.

Conclusions

The behavior rating scales reported above were chosen because they have received much attention in the assessment and diagnosis literature (the DSF and DSP were reported because they are new and unique measures of behavior and emotional problems in children and adolescents). However, the fact that these instruments are so widely used by researchers does not indicate that they are the "best" available measures. Naglieri & Flanagan (1992) compared the psychometric properties of a variety of behavior rating scales. These authors used the following criteria in selecting the behavior rating
scales to be included in their study: 1) the scale should be designed for use with children and adolescents, 2) the scale should focus on behaviors, not personality characteristics, 3) the scale should provide an evaluation of behavior across several domains, 4) the scale should be standardized, and 5) the scale should be completed by an adult in reference to a child's behavior. The following behavior scales were selected: the Child Behavior Checklist (CBCL), Teacher's Report Form (TRF), Emotional and Behavior Problem Scale (EBPS), Revised Behavior Problem Checklist (RBPC), Louisville Behavior Checklist (LBC), Behavior Evaluation Scale (BES), Behavior Rating Profile (BRP), Walker Problem Behavior Identification Checklist, School Behavior Checklist (SBC), Burk's Behavior Rating Scales (BBRS), Hahnemann Elementary School Behavior Rating Scale, Devereux Adolescent Behavior Rating Scale, and the Devereux Child Behavior Scale. Naglieri & Flanagan (1992) provided a summary of these rating scales based on psychometric properties as outlined by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1985, cf. Naglieri & Flanagan, 1992) and the Consumer's Guide to Tests in Print (Hammill, Brown, & Bryant, 1988, cf. Naglieri & Flanagan, 1992). More specifically, the
authors examined the reliability, validity, standardization, and interpretation guidelines as presented by the individual authors of the selected scales.

The results of this study indicated that only the EBPS, BES, and BRP provide adequate norms with regard to size, representativeness, and recency of their standardization sample. The CBCL, TRF, and SBC provide excellent interpretation guidelines while the RBPC, HESB, and DCB provide minimal guidelines. Naglieri & Flanagan (1992) report that only the EBPS (four subscales only), RBPC, and WBPIC provide adequate evidence \( r \geq .80 \) of all three types of reliability (internal consistency, test-retest, and interrater). Only the WBPIC meets the Standards for Educational and Psychological Testing (AERA, et al., 1985, cf. Naglieri & Flanagan, 1992) criteria for construct, criterion-related, concurrent, predictive, and content validity. To summarize the overall adequacy of their fourteen behavior rating scales the authors rejected 1) any scale which did not have an adequate standardization, 2) any scale with inadequate reliability, and 3) any scale with inadequate validity. The BES, BRP, and EBPS were the only scales which passed all three of the aforementioned criteria. Naglieri &
Flanagan (1992) note that "it is a disturbing fact that more than 75% of the behavior rating scales available do not have an adequate standardization sample even though some of those rejected have good reliability and/or validity" (p. 21).

Despite the psychometric limitations of some behavior rating scales, those meeting adequate psychometric properties are an important asset to the psychologist assessing for emotional and behavior problems in students. McConaughy (1985) concludes that behavior rating scales are of value in school-based assessments for the following reasons: 1) the obtained information is quantifiable and therefore subject to psychometric standards of reliability and validity, 2) normative data allow the psychologist to compare ratings of an individual child to those of a larger, non-referred population of the same sex and age range, 3) often the large pool of items enables the psychologist to obtain a wealth of information beyond the referral question, 4) scoring profiles allow the psychologist to organize information in a systematic way, 5) broad- and narrow-band syndromes encompass empirically derived syndromes which is a significant advance over current classification systems (e.g., DSM-III-R), 6) multiple
sources of information can be obtained on the same measure (e.g., mother, father, teacher, etc.), and 7) standardized rating scales are efficient and economical. The use of projective techniques in the schools is discussed next.

Projectives

According to Peterson & Batsche (1983) there has been a "tightly intertwined relationship" between school psychologists and projective assessment. Goh & Fuller (1983) conducted an empirical analysis of behavior and personality assessment measures utilized by school psychologists. These authors found that projective techniques, more specifically human figure drawings and sentence completion tasks, were used far more often than objective personality tests and behavior rating scales by school psychologists. The Bender Visual-Motor Gestalt Test was used most often by the 274 school psychologists surveyed (88%), with sentence completion techniques a close second (87%). The House-Tree-Person (73%) and the Thematic Apperception Test (65%) were the next most utilized measures. It is interesting to note that school psychologists were more likely to use the Thematic Apperception Test than they were the Children's
Apperception Test (59%). In contrast to this finding was the significantly lower usage of empirically based, standardized behavior rating scales: Devereux Behavior Rating Scales (33%), and the Walker Problem Behavior Identification Checklist (24%). However, recent data from Hutton and his colleagues (1992) indicate that school psychologists are using behavior rating scales more, and projective personality techniques less (18.3% v. 11.5%). Despite this change in practice, many school psychologists are still incorporating projective assessment measures into their psychological test battery. Nonetheless, there is a considerable amount of research which has criticized the use of projective techniques. Before this literature is presented, a description of the projective hypothesis is in order.

**Projective Hypothesis of Behavior**

The projective hypothesis of behavior assumes that when a child is exposed to a stimulus which is ambiguous, his or her interpretations will reveal elements of intrapsychic conflict which are not overtly manifested. The child's interpretations can provide symbolic information about his or her current conflict and eventually lead to conscious and unconscious factors
which influence the psychological functioning of the child as well as revealing the underlying dynamic factors which shape the child's personality and thus influence his or her behavior (Knoff, 1986). As Gittel (1980) points out, diagnosis can be conducted at several different levels. The psychologist can look for specific personality types or characteristics and/or levels of personality organization. Personality syndromes or mental disorders can also be the focus of diagnosis. There is a widely held assumption that dynamic relationships exist between personality characteristics and psychiatric disorders among those professionals who purport the use of projective assessment (Gittel, 1980). Several authors have attempted to empirically assess the utility and efficacy of the various methods of personality and emotional assessment employing the projective hypothesis.

Projective Assessment Research

In her review of the literature on projective psychological testing with children Gittel (1980) found seven studies which attempted to detect aggressive tendencies in children. These studies used one of the following measures: the TAT, Bender-Gestalt, Draw-a-
Person test, and the Rorschach. Two of the studies were conducted prior to 1960 and the remaining studies were conducted in the mid to late 70's. Gittelmann concluded that based on the empirical results of these studies, there is no reason to believe that these measures can provide a valid estimate of aggressive tendencies in children.

Gittelmann (1980) found 7 studies, 6 of which used the Draw-a-Person test and 1 which used the Rorschach to assess self-esteem and self-image in children. She concludes that most of the studies yielded negative results. The few studies which found expected significant differences were not to the degree which would allow the application of projective tests as measures of self-image in individual children (Gittelmann, 1980).

Diagnosis

Clinical versus Normal Groups

Most of the research on projective assessment and diagnosis has attempted to differentiate normal children from those children who are either referred, or already in treatment for, emotional or behavior problems. According to Gittelmann (1980), this "approach may well be
fruitless if different pathological groups have opposite response patterns (the combining of values which are opposite in various disturbed groups could lead to negative results in mean differences between disturbed and normal groups, and only the score variances might reveal group contrasts)" (p. 423). The advantage of this research methodology is that researchers do not have to reliably define their patient group. Gittelman reviewed 16 studies utilizing the aforementioned methodology and found that these studies did not do one or more of the following: 1) keep the examiners "blind" to the research groups (clinical or normal control), or 2) control for age, IQ, and/or socioeconomic status of the subjects. These methodological inadequacies seriously flaw this research literature and render its results uninterpretable (Gittelman, 1980).

Comparisons Between Diagnostic Groups

Gittelman (1980) could not find one study which offered independent reliability and validity data for the predicted variable. Furthermore, none of the studies comparing diagnostic groups based their diagnoses on objective and reliable clinical criteria (Gittelman, 1980). Based on her review of 15 studies, Gittelman
(1980) found all 15 studies to be flawed by the same methodological inadequacies as previously mentioned. With regard to reliability, the Bender-Gestalt has a mean test-retest reliability coefficient of .60 (Koppitz, 1975, cf Gittelman, 1980). The TAT has not been shown to have adequate test-retest reliability (Kagan, 1959, 1960; Reddy, 1960, cf. Gittelman, 1980). The Rorschach's test-retest coefficients have been shown to be positively related to age with the correlation for children under age 8 being poor and above age 8 very good (Ames et al., 1974, cf. Gittelman, 1980). Gittelman (1980) concluded that only the Rorschach appears to tap some meaningful "human characteristics" in children over age 8. Peterson & Batsche (1983) report that there is some research evidence which demonstrates that the Draw-a-Person, the House-Tree-Person, and the Kinetic-Family-Drawing measures have satisfactory reliability. However, given a lack of empirical validity data for projective measures as a group, the question remains, what are these tests measuring? According to Knoff (1983) "...responses to projective stimuli reflect the student's [cognitive] style and contents of thinking and perception; these data are not readily available from behavioral observations" (p. 396). Knoff believes that projective assessment in a
school environment can 1) identify psychological issues which may be troubling a student or a significant person in his or her ecological system, 2) provide a better understanding of the referral problem, especially when integrated with behavioral data, and 3) indicate appropriate interventions.

**Projective Assessment in the Schools**

Knoff (1983) points out that the school psychologist needs to view a socio-emotional referral question in an ecological context (both in and out of school) which considers the present and future mental health of the child. A projective battery can be part of a personality assessment but does not have to be (Knoff, 1983). A projective battery is indicated if 1) the referral question is so complicated that projectives can offer a better understanding of the problem, 2) the referred student is resistant to diagnostic interviewing, 3) another mental health professional requires confirmation of his or her diagnostic findings and/or clinical impressions. As Koppitz has stated (1982), in many instances it is possible to conduct a personality assessment on the basis of student observations, teacher reports, parent conferences, and school records. When a
projective battery is necessary, Koppitz (1982) and Worchel (1990) have suggested the following techniques for use with school aged children and adolescents (these techniques are divided into verbal, visual, drawing, and objective measures):

**Verbal Techniques**

*Projective questions* are quick and easy to administer and can be slipped informally into a formal assessment, one or two at a time. These questions can offer the psychologist useful information about the student's attitudes and feelings. Examples of projective questions are "suppose you had three wishes, what would you wish for?", "if you had the choice, how old would you like to be, would you like to be younger, older, or the same age you are now?" The psychologist needs to call on his or her knowledge of child development to formulate age appropriate questions. *Rotter's Incomplete Sentence Blank-High School Form (ISB)* consists of 40 sentence stems which can be quickly and easily administered. The ISB can serve as an "icebreaker" with shy or reluctant students. Rotter has provided a scoring system which evaluates the quality of the student's responses on a seven point scale for the degree of conflict each
response reflects (Koppitz, 1982). The Haak Sentence Completion Test (HAAK) is designed specifically for use with elementary and secondary school students. Research on the validity and reliability of the HAAK test has yet to be presented (Worchel, 1990). The Hart Sentence Completion Test (HSCT) is a 40 item scale designed for use with children and adolescents ages 6 to 18. According to Worchel (1990) the HSCT is currently undergoing normative work and will hopefully become a sentence completion test with adequate psychometric properties.

Because young children are often reluctant to reply to direct questions, Koppitz (1982), has suggested the use of short stories or fables. The Madeleine Thomas Stories are a set of 15 stories about a hypothetical child which explore children's dreams, fantasies, attitudes, and defenses. An example story is: "A boy is at the table with his parents. Father suddenly gets angry. Why?" The stories are appropriate for children aged 5 to 11 years (Koppitz, 1982). According to Koppitz (1982) the best results have been obtained with 8 to 10 year old children. The Despert Fables are a set of 10 fables which were translated into English in 1940 by Louisa Duss (Despert). An example of a Despert fable is:
"It is the daddy's and mommy's wedding anniversary. They love each other very much and are having a beautiful party. During the party the child gets up and all alone goes to the very end of the backyard. Why?". Koppitz reports that the Despert fables appear to be most appropriate for children aged 3 to 7 years. The Munsterberg Incomplete Stories (MIS) are a set of 12 incomplete stories designed to assess the interpersonal relationships of school children with their parents, siblings, and friends. The MIS are designed for use with children aged 6 to 12 years. The completed stories are analyzed for positive and negative interpersonal attitudes between the child and his or her significant others (Koppitz, 1982).

**Visual Techniques**

According to Koppitz (1982) the Rorschach is a valuable technique in a clinic setting but is typically not suggested for use in schools. The Rorschach was developed to provide an in depth analysis of the adult psychiatric patient's personality. The Thematic Apperception Test (TAT) is widely used by school and clinical psychologists (Koppitz, 1982). The TAT, which is the oldest of the thematic apperception methods,
consists of 29 pictures with figures and scenes and one blank card. The TAT consists of cards that are appropriate for both sexes and cards that are gender specific. In the school environment the TAT is most appropriate for students age 11 or older. The *Children's Apperception Test* (CAT) is designed for use with younger children age 3 to 7. The CAT contains 10 cards with drawings of anthropomorphic animals in various situations. The CAT was designed to collect information about the children's attitudes toward parents, siblings, eating, sleeping, aggression, loneliness, and a variety of other psychological variables. There is a supplement to the CAT, the CAT-S which provides pictures of a more general nature (e.g., visit to the doctors office), and a human version of the CAT, the CAT-H. Interpretation of the CAT is similar to that of the TAT (Worchel, 1990). The *School Apperception Method* (SAM) has the distinction of being the only apperception test that deals solely with school situations. The 12 standard and 10 alternative cards are appropriate for use with children age 8 through 13. Obrzut & Boliek (1986, cf. Worchel, 1990) report that the psychometric properties of the SAM are unsubstantiated. The *Robert's Apperception Test* (RAT) is a projective test with a standardized scoring
system. Worchel (1990) indicates that the validity data for the RAT is weak when compared to more objective personality tests, but is relatively good when compared to similar projective techniques. Other apperception tests include the Michigan Picture Test (MPT), the Tasks of Emotional Development Test (TED), the Make A Picture Story Test (MAPS), and the Hands Test. A description of these tests can be found in Koppitz (1982) and Worchel (1990).

**Drawing Techniques**

The Human Figure Drawing (HFD) test requires a child to draw one whole person in the presence of the examiner. Koppitz (1982) suggests that the instruction to "draw a whole person" and not "draw a self-portrait" is especially important with adolescents because they are often reluctant to draw themselves. Koppitz (1968) has empirically identified 30 emotional indicators (EI's) on HFD protocols. The EI's were chosen if they could differentiate HFD's of groups of children with and without serious emotional disturbance, were unusual in that they did not occur in more than 15% of the HFD's of normal students, and were not related to age and developmental level. A complete description of these
indicators can be found in the *Psychological evaluation of children's human figure drawings* (Koppitz, 1968). Comprehensive reviews of the literature by Snyder & Gaston, 1970, Phil & Nimrod, 1976, Dieffenbach, 1977, & Mortensen, 1984 indicate that there is no empirical support for the Koppitz (1968) HFD scoring system.

The *Kinetic-Family-Drawing* (K-F-D) technique requires a child to draw everyone in his or her family doing something. The K-F-D's are analyzed for action, symbols, and style. Action refers to the movement of energy between people and/or objects and reflects anger, love, power, etc., symbols are interpreted according to psychoanalytic theory, and style is suppose to suggest defensiveness, e.g., compartmentalization of figures (Koppitz, 1982). The *House-Tree-Person* (H-T-P) test requires a child to draw a house, tree, and a person and is interpreted based on psychoanalytic theory through the analysis of symbols. The child is asked to draw a person of the same sex and a person of the opposite sex. The first four drawings are made in pencil and then the child is asked to produce another set of H-T-P drawings in crayon. A series of postdrawing interrogation questions are available from Buck (1966). For younger children the *Levy Animal Drawing Story* (LADS) is a useful drawing
technique. On the LADS children are asked to draw any animal they wish. After they have completed their drawing they are asked to write in the upper corner their age, sex, the kind of animal they drew, and the names of other animals they might have drawn. Under their animal drawing itself the children are asked to give it a pet name. Normative data are available to determine whether the animal drawn is common or unusual. The drawings are also analyzed for size, placement, and symbolic meaning.

Although originally designed as a test of visual-motor perception the Bender Gestalt has been modified by Koppitz (1963;1975, cf. Koppitz, 1982) to yield 12 emotional indicators. The EIs were chosen if they were able to differentiate between groups of well adjusted and emotionally disturbed youngsters, and if they were not related to age and maturation. Koppitz's scoring system is designed for use with children age 5 to 11. The test requires children to copy 9 abstract designs, one at a time, on a sheet of blank paper. Koppitz (1975) states that three or more EI's are required before a clinician can say with any degree of confidence that a child has a serious emotional problem. Furthermore, any hypotheses developed need to be checked against other psychological
data. Koppitz (1975) does not provide data on the psychometric properties of her EIs.

**An Objective Drawing Measure**

Naglieri, McNeish, & Bardos (1991) have provided an objective, standardized (n = 2,355) scoring system to be applied to the drawings of children and adolescents (Draw A Person: Screening Procedure for Emotional Disturbance - DAP:SPED). The DAP:SPED was designed as a screening measure which differentiates the drawings (man, women, and self) of normal children and adolescents from those children and adolescents with emotional or behavior problems. The DAP:SPED was developed using an actuarial approach to the identification of children and adolescents (ages 6-17) with emotional or behavioral problems and is not intended as a differential diagnostic tool (Naglieri et al., 1991).

The DAP:SPED is comprised of 55 total items. Nine of the items are measurement items (e.g., small versus large figure, figure placement, figure slanting, etc.). Fourty-six of the items are content items (e.g., crossed eyes, shading, frowning mouth, talons, etc.). The items were selected after an exhaustive review of the literature of those signs which are considered to be
emotional indicators and were retained if they occurred infrequently in the normal standardization sample and showed adequate psychometric properties. Intra and interrater reliability coefficients are reported by McNeish (1989) as .94 and .91 respectively. Internal reliability coefficients are .80 for ages 5-8, .77 for ages 9-12, and .71 for ages 11-17. Several discriminate studies with highly successful results have been reported (see test manual). The DAP:SPED is uncorrelated (median $r = -.14$) with intelligence. The DAP:SPED is a valuable assessment measure for use by psychologists because it is a recent, standardized, objective screening procedure which is quick (15 minutes maximum) and easy to administer. The DAP:SPED can be used in conjunction with the Draw A Person: Quantitative Scoring System (Naglieri, 1988) which is a nonverbal measure of intelligence.

**Objective Techniques**

According to Koppitz (1982) the major weakness of objective personality questionnaires/scales is that an adolescent student completing a questionnaire/scale can choose what he or she wants to reveal. Children are often eager to please and as such will typically answer the questions the way they believe they should be
answered (Koppitz, 1982). As Koppitz (1982) points out, personality questionnaires have been, and continue to be, developed for use in the schools. Major advantages of personality questionnaires/scales is that many of them are empirically derived and standardized and are quick and easy to administer. The following measures are discussed by Koppitz (1982) and Worchel (1990) in their reviews of personality assessment in the schools.

**Anxiety Scales**

*Taylor's Manifest Anxiety Scale* (MAS) was designed for adolescents and adults and measures for predisposition to react with anxiety in anxiety-provoking situations. The test contains 50 true-false items. The *Children's Manifest Anxiety Scale* (CMAS) is a downward extension of the MAS which includes 42 items and an 11 item lie scale. The CMAS assesses for generalized anxiety. The *Generalized Anxiety Scale for Children* (GASC) measures trait anxiety via 34 yes-no items and includes an 11 item lie scale. The *Test Anxiety Scale for Children* (TASC) was specifically designed for 1st through 6th graders to assess their feelings about classroom performance, tests, and their classmates. The TASC has been factor analyzed and consists of four
factors: 1) test anxiety, 2) somatic signs, 3) negative self-evaluation, and 4) remote school concerns. The "What I Think and Feel" (WITF) scale is a later modification of the CMAS and can be used as part of a personality assessment battery for 4th to 12th graders. The WITF is comprised of 28 true-false items and a 9 item lie scale. The State-Trait Anxiety Inventory for Children (STAIC), also referred to as the "How I Feel Questionnaire", consists of an anxiety state and an anxiety trait scale. Each scale has 20 questions. The STAIC was primarily designed for research purposes with 4th to 6th graders but can be used as a progress and outcome measure in counseling and therapy. All of the aforementioned questionnaires/scales specifically assess for anxiety in children and adolescents. In light of evidence which suggests that certain groups of students' achievement performance is influenced by their anxiety level, Koppitz (1982) has indicated that these scales can be a useful addition to the psychologist's personality assessment battery. The Early School Personality Questionnaire and the Junior-Senior High School Personality Questionnaire measure personality structure and personality adjustment in general and can also be of assistance to the school psychologist (Worchel, 1990).
Depression Scales

The Children's Depression Inventory (CDI) is a downward extension of the Beck Depression Inventory (BDI). The CDI is a 27 item self-report measure designed for use with school-aged children. Worchel (1990) reports that the CDI has adequate psychometric properties. The Children's Depression Scale (CDS) is a 66 item self-report measure designed for use with children aged 9 through 16. The Child Depression Scale (CDS) is a 30 item self-report measure designed for use with children ages 8 to 13. Both scales have adequate psychometric properties (Worchel, 1990). Finch & Saylor (1984, cf. Worchel, 1990) report that the Peer Nomination Inventory of Depression (PNID) has excellent psychometric properties. The PNID is a sociometric measure of childhood depression. For a brief review of these instruments see Worchel (1990).

Children's Apperceptive Storytelling Test

One measure not mentioned by Koppitz (1982) and Worchel (1990) in their reviews is the Children's Apperceptive Storytelling Test (CAST) (Schneider, 1989). The CAST was developed out of the need for a valid, reliable, theory based, objectively scored measure which
is standardized on a nationally representative sample of children (Schneider, 1989). According to Schneider (1989), "The CAST presents contemporary, colored picture card stimuli specifically designed to elicit relevant life-task issues of childhood and early adolescence while addressing the need for psychometric rigor" (p.1). The CAST is designed to assess social, emotional, and behavioral adjustment issues experienced by children and adolescents aged 6-13. Schneider (1989) provides evidence of content, construct, and criterion validity in the test manual. Studies utilizing discriminant function analysis are presented to show that the CAST can distinguish conduct disordered, anxiety disordered, oppositional, depressed, schizoid disordered, and children with attention deficit disorder from matched control samples. Median intrarater reliability coefficients range from .84 to .98, the median test-retest reliability coefficient = .77, and the median internal consistency (split-half) reliability coefficient = .74 for the standardization sample.
The Compatibility of Projective Assessment and School Psychologists

Peterson and Batsche's Criticism

In a published debate with Howard Knoff over the merits of projective assessment in the schools, Peterson & Batsche (1983) have suggested that because the role of the school psychologist is rapidly changing, the use of projective assessment is no longer warranted in the schools. More specifically, during the 1950's and 1960's the role of the school psychologist was to identify, within a medical-type classification system, those students who exhibited pathology and to classify them by labeling their "condition". Thirty years later school psychologists now function on multidisciplinary teams whose objective is to gather valid diagnostic information to be used in the development and implementation of interventions. This data collection is multidimensional and requires observations of the child in his or her educational and social environment. According to Peterson & Batsche (1983), projective assessment does not take into account the child's social environment because of its basis in psychodynamic theory which assumes hidden variables as the target of assessment (for a definition of projective assessment which does take into account the
child's social environment see Leland, 1983). Peterson & Batsche (1983) suggest that this makes the use of projective assessment incompatible with the current role of the school psychologist. Another point offered by Peterson & Batsche (1983) in defense of their position is that projective assessment does not meet the procedural safeguards in assessment and litigation put forth by Section 504 of P.L. 94-142. These safeguards include reliability, validity, and outcome criteria of the assessment tools used by school psychologists. Peterson & Batsche (1983) state that ample evidence exists to support the claim that projective assessment techniques do not meet the aforementioned psychometric properties. The continued use of projective measures by school psychologists therefore puts them in legal jeopardy if and when they are called to testify in court (Peterson & Batsche, 1983). Finally, these authors point out that the technical inadequacy (poor to marginal reliabilities and a virtual lack of validity data) of projective techniques warrants their disuse in the schools.
Knoff's Response

In response to Peterson & Batsche's first claim that projective assessment is incompatible with the current role of the school psychologist, Knoff (1983) states that projective assessment has maintained its relevance by developing "child and adolescent procedures and norms which are based and tested in the schools" (p. 447). Furthermore, Knoff (1983) asserts that the differentiation between normal students and students with emotional problems without consideration of a differential diagnosis is in fact what P.L. 94-142 mandates. As such, Knoff (1983) suggests that projective tests do have the potential to satisfy the requirements of P.L. 94-142. According to Knoff (1983) projective tests are used more within the current zeitgeist than they are within the traditional confines of their historical roots and as a result they are capable of meeting the situational demands of the school environment. As far as the potential of litigation is concerned, Knoff (1983) notes that Peterson & Batsche (1983) "...have presented no data to support their notion that litigations involving emotional disturbance (or other) referrals are based significantly more often on the use of projective tests than any other reason or
assessment approach" (p. 449). Finally, in response to the psychometric criticisms of projective tests, Knoff (1983) states that the validity issue should be considered through decision-making theory which emphasizes the validity of the decision process (e.g., placement and programming) rather than the specific, individual tests employed in the process.

Conclusions

Naglieri, McNeish, & Bardos (1991) recently summarized the literature addressing the use of human figure drawings as a projective personality assessment technique by saying "...the research results have been characterized by conflict. Numerous case studies appear in the literature that attest to the technique's usefulness in isolated clinical situations, yet experimental research has offered little to no support for the technique's validity as an indicator of emotional or behavioral attributes" (p. 2).

Based on their review of the literature, Naglieri et al. (1991) do not suggest that the use of HFDs as projective personality assessment techniques is unwarranted, or is their use incompatible with school psychology. These authors do point out that there is a
need to provide human figure drawing systems that are objective (in terms of scoring), reliable, and that differentiate between normal and disturbed populations. Naglieri et al. (1991) have offered such a scoring system in the DAP:SPED which was described earlier in this section.

The work of Naglieri and his colleagues (1991) is an attempt to improve upon the early projective work of Machover (1949) and Koppitz (1968, 1984). Much of this work has been shown to lack empirical support (see Kahill, 1984 & Mortensen, 1984). In regard to Peterson & Batsche's (1982) position that projective assessment and school psychology are incompatible, it is important to point out that objective, standardized projective scoring systems are required to adhere to the same psychometric standards as intelligence and achievement tests are. Intelligence and achievement tests are an integral part of the school psychologist's psychological test battery. It is careless to suggest that because some projective measures are not supported by empirical data, the entire assessment practice should be abandoned.

A final but important consideration is the training of school psychologists to administer and interpret projective tests. Unlike behavior rating scales, many
projective tests require a great deal of specialized training. Not all school psychologists, especially at the masters and specialists level, have time in the course of their intense training to receive the appropriate amount of instruction and practicum experience in the interpretation of projective tests. Individual school psychologists need to be aware of their own professional limitations and practice accordingly. Another area which not all school psychologists are appropriately trained is in the use of the DSM-III-R. The final section of this chapter will look at DSM based diagnostic interviews.

DSM Structured Diagnostic Interviews

Two of the primary structured and standardized diagnostic interviews based on DSM-III criteria are the National Institute of Mental Health's (NIMH) Diagnostic Interview Schedule for Children (DISC) and the NIMH's Diagnostic Interview for Children and Adolescents (DICA). Before these two interview schedules are presented, a discussion of the reliability and validity of the DSM in the diagnosis of child psychopathology will follow.
Psychometric Properties of the DSM

Reliability

Using DSM-III field trial data, Spitzer, Forman, & Nee (1979) had pairs of clinicians make individual diagnoses for the same case and then computed the interrater reliabilities for the various disorders diagnosed. The overall interrater kappa coefficient for the Axis I disorders was .69. This breaks down to a perfect 1.0 for mental retardation and stereotyped movement disorders, .85 for pervasive developmental disorders, .66 for eating disorders, .61 for conduct disorder, .58 for attention deficit disorder, .66 for adjustment disorder, .53 for affective disorders, .25 for anxiety disorders, and .79 for other disorders. Cantwell (1988) has suggested that the low interrater reliability for the anxiety disorders is a result of a lack of consensus on whether depressive symptoms constitute a part of the diagnostic criteria for the anxiety disorders. Rutter & Shaffer (1980) have criticized the results of the DSM-III interrater reliability study because the clinicians who participated were close colleagues, and the cases assigned for diagnosis differed for each pair of raters. Other authors have conducted
independent reliability studies using the DSM-III and have produced similar results (for a review of these studies see Cantwell, 1988).

Validity

One of the major challenges facing the DSM has been a difficulty in establishing the validity of the various diagnostic categories. Cantwell (1983) has described a six-step procedure for developing valid diagnostic categories for various psychiatric disorders. Cantwell's (1983) model is an extension of an earlier model by Robins & Guze (1970, cf. Cantwell, 1988). According to Cantwell (1988), a valid diagnostic category should 1) describe the clinical picture, 2) describe any coexisting physical and neurological factors or disorders, 3) utilize laboratory studies, as well as 4) family studies, 5) natural history studies, and 6) treatment studies. Cantwell (1988) points out that very few studies exist which have employed this model and thus there is little existing empirical support for the DSM-III diagnostic categories. What child research exists has been conducted with attention deficit hyperactivity and affective disorder populations. There is some evidence to support the distinction between major depressive
disorder, dysthyemic disorder, and adjustment disorder (see Kovacs, Feinberg, Cruse-Novack, Paulauskas, & Finkelstein, 1984). According to Cantwell (1988) the research literature on attention deficit hyperactivity disorder is not adequate enough to establish diagnostic validity and is split on whether or not the disorder should be diagnosed with or without hyperactivity. With the psychometric properties of the DSM-III in mind, the two aforementioned diagnostic interviews based on DSM-III criteria, and one recently made available behavior rating scale based on DSM-III-R criteria will be presented.

**Diagnostic Interview Schedule for Children**

The Diagnostic Interview Schedule for Children (DISC) and its parallel form the Diagnostic Interview Schedule for Children-Parent Form (DISC-P) assesses the chronology, duration, and associated impairments of specific psychiatric symptoms. Both the DISC and the DISC-P produce scores for 27 hierarchically ordered symptom groups. These groups are: Behavior/Conduct, Conduct Disorder-Aggressive, Nonaggressive, Oppositionalism, Attention Deficit-Inattention, Attention Deficit-Impulsivity, Overactivity, Alcohol/Drug Abuse, Affective/Neurotic, Anxiety, Separation Anxiety,
Overanxious, Fears/Phobias, Simple Fears, Social Phobias, Obsessive-Compulsive, Schizoid/Psychotic. Affective Disorders, Mania, Depression-Affective, Cognitive, Vegetative, and Suicidal. The DISC and DISC-P also yield a Total Symptom Score. Symptoms are coded 0 ("no"), 1 ("somewhat or sometimes"), and 2 ("yes"). The interviews are scored in terms of DSM-III criteria by using an algorithm that operationalizes DSM-III criteria for each disorder. According to Costello, Edelbrock, & Costello (1985) most of the diagnoses obtained from this procedure represent a "mild" to "moderate" level which is just above diagnostic threshold and requires little or no intervention. As a result, operational rules have been written for defining "severe" disturbance which involve a number and severity of symptoms above that required by the DSM-III. Mean 2-week test-retest reliability for the DISC-P is reported to be .84 while the test-retest reliability for the DISC is reported to be .75 (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985).

Costello et al. (1985) conducted a criterion validity study using both the DISC and DISC-P. These authors used DISC interviews to evaluate 40 psychiatric referrals aged 7-11 and 40 pediatric referrals matched for age, sex, race, and socioeconomic status. The CBCL
was also given to the parents of both groups. Results indicated that the DISC-P discriminated between criterion groups better than the DISC did. The DISC-P had a sensitivity of 60%, a specificity of 95%, and an overall screening efficiency of 77.5% when one or more severe disorders was used as an index of psychopathology and referral for professional services. When mild to moderate disorders were added to the analysis, the sensitivity increased to 93%, the specificity was reduced to 30% and the overall screening efficiency dropped to 61.5%. Considering only severe disorders the DISC had a sensitivity of 45%, a specificity of 80%, and a screening efficiency of 60%. When mild, moderate, and severe disorders were analyzed together the sensitivity increased to 95%, the specificity dropped to 25%, and the overall screening efficiency remained at 60%. In both groups mild oppositional behavior and fears were often reported by parents while children commonly reported mild separation anxiety, fears, and dysthymia. Attention deficit disorder, conduct disorder, and affective disorders were significantly more common among the psychiatric referral group. Costello and her colleagues (1985) reported a significant association between abnormally high CBCL scores and diagnoses derived from
the DISC-P. There was not a significant association between the CBCL and DISC. Costello et al. (1985) concluded that their data provides preliminary evidence of criterion validity for both the DISC and DISC-P. They note that when mild to moderate diagnoses are used as the criterion a resulting high level of sensitivity occurs at the expense of specificity.

**Diagnostic Interview for Children and Adolescents**

The Diagnostic Interview for Children and Adolescents (DICA) is a structured interview divided into 18 parts which correspond with the 18 DSM-III diagnostic categories. The DICA includes two separate interviews, the DICA-C for children and adolescents and a parallel parent interview (DICA-P). The DICA is designed so that one or more questions assess each symptom for every DSM-III disorder. The present DICA is a revised version of an earlier form which was based on the International Classification of Psychiatric Disorders. The present form is patterned after the National Institute of Mental Health's Diagnostic Interview Schedule (DIS).

Welner, Reich, Herjanic, Jung, & Amado (1987) report that the DICA-C has good inter-interview reliability. These authors determined this by interviewing 27
psychiatric inpatients aged 7 to 17 and comparing their
DICA-C interviews with the discharge diagnosis of the
treating clinician. The highest agreement is seen in the
diagnostic category of affective disorders with a kappa
coefficient of .52. The lowest agreement is seen in the
anxiety diagnostic category with a kappa coefficient of
.03 with higher rates of these diagnoses being indicated
by the DICA-C than by the clinician. A comparison of the
DICA-C diagnoses with the hospital discharge diagnoses
for validation of the DICA-C indicated that in 81.5% of
the cases, the treating clinicians were in agreement with
the DICA-C diagnoses. Welner et al. (1987) conducted an
analysis of mother-child agreement on the DICA for any
diagnostic group in which they had 10 or more subjects.
Five groups were obtained with an N of 84. The highest
agreement was obtained for the Conduct Disorder
diagnostic group with a kappa coefficient of .80.
Attention Deficit Disorder yielded a kappa of .66,
Affective Disorder .63, Oppositional Disorder .52, and
Enuresis .49. Based on their data Welner and his
colleagues (1987) suggest that the DICA-C is a reliable
and valid instrument for either clinical or research use.
Disruptive Behavior Disorders Rating Scale

Pelham, Evans, Gnagy, & Greensdale (1992) have recently produced a behavior rating scale which is based on DSM-III-R diagnostic criteria for the disruptive behavior disorders triad of Attention Deficit Hyperactivity Disorder, Oppositional Disorder, and Conduct Disorder. These authors administered their scale, named the Disruptive Behavior Rating Scale (DBD), to 364 boys in special education classrooms in elementary schools in North America. In their initial report Pelham, et al., (1992) report good psychometric properties. Factor analysis of teacher ratings of the 364 boys revealed four factors. Factor 1 contains all of the ODD diagnostic criteria, 4 of the CD criteria, and two of the ADHD criteria. The second factor is comprised only of those ADHD criteria which reflect inattention while the third factor is comprised of those ADHD criteria describing impulsive behavior. The final factor consists of 2 CD criteria. Of interest to diagnosticians and researchers alike was the finding that several of the hallmark symptoms of ADHD ("fidgeting" and "distractible" behavior) were poor predictors of the presence of the disorder in Pelham et al.'s (1992) male, special education sample. "Interrupting" and "difficulty
"waiting turn" were the best at predicting ADHD. 55% of the children with ADHD exhibited concurrent difficulties with oppositional or conduct disorder. Most of the overlap was between ODD and ADHD, not CD and ADHD. Pelham and his colleagues (1992) note that the results of their factor analysis are consistent with the suggested modification of DSM-IV diagnostic criteria for ADHD which proposes that symptoms be present from two clusters: inattention and impulsivity/overactivity. While more psychometric data is needed on this new instrument, initial results suggest that the DBD will be of great use to psychologists assessing for disruptive behavior disorders. Studies need to be conducted that employ special education populations which are different from that utilized by Pelham and his colleagues (1992) so that generalizability of the DBD is enhanced. Of special importance to school psychologists is that the DBD allows for a special education classification which is commensurate with a DSM psychiatric diagnosis. Despite the apparent utility of the DSM as part of a comprehensive assessment for emotional disturbance in school aged children and adolescents, there is evidence to suggest that not all school psychologists are incorporating the DSM-III into their practice.
**DSM in the Schools**

In a national survey of school psychologists conducted by Prout (1983) 18.1% of the respondents indicated that they were "essentially unfamiliar" with the DSM-III, 43.9% said they were "somewhat familiar", 17.5% reported that they were "very familiar", and 20.5% indicated that they have "clinical working knowledge" of the DSM-III. According to Sattler (1983), thorough knowledge of the DSM can be helpful to psychologists in the schools for four reasons: 1) the DSM provides valuable guidelines for understanding mental disorders, 2) knowledge of the DSM can improve communication with the psychiatric community as well as improve the mental health network with the community, 3) the multiaxial system of the DSM can be a potentially valuable method for following individual cases over time, and 4) the relationship between the DSM-III categories and P.L. 94-142 may provide greater understanding of referred and classified children. Sattler (1983) states that school psychologists are in a key position to obtain objective information that can be used to classify children according to DSM criteria. The diagnoses of Developmental Reading Disorders, Developmental Arithmetic Disorders, and Mental Retardation require the
administration of individual, standardized intelligence
and achievement measures. Psychologists in the schools
are in a primary position to make these diagnoses.
Sattler (1983) suggests that all psychologists who
practice in the schools should be thoroughly familiar
with the DSM-III.

Tharinger, Laurent, & Best (1986) investigated the
utility of the DSM-III in the school setting by comparing
it with two other major classification systems, P.L. 94-
142 Seriously Emotionally Disturbed (SED) guidelines, and
the Child Behavior Checklist (CBCL). Thirty eight boys
(mean age = 7 1/2 years) who were referred because of
suspected emotional or behavioral problems were evaluated
by a psychological diagnostic staff of a school district
and classified, if appropriate, according to P.L. 94-142
SED guidelines. The files of these boys were then
reviewed by two graduate students trained in the use of
the DSM-III who made Axis I diagnoses if warranted.
Interrater agreement for the 38 cases was 52%. For
purposes of the study the remaining 48% of the cases were
discussed between the two raters and a mutual diagnosis
reached. Each boy was then classified using the CBCL
system as having an internalizing, externalizing, mixed,
or no behavior disorder.
Results indicated that 82% of the sample was assigned a DSM-III Axis I diagnosis, 66% met CBCL criteria for a significant behavior disorder (3% internalizing, 26% externalizing, and 37% mixed), and 53% were classified as having a serious emotional disturbance. Considering the agreement between the systems, the total agreement between P.L. 94-142 SED criteria and DSM-III was 55%. The total agreement between P.L. 94-142 SED criteria and the CBCL system was also 55% while the total agreement between DSM-III and the CBCL system was 63%. A three way comparison between P.L. 94-142 criteria, DSM-III, and the CBCL system yielded a total agreement of 29%. In an effort to establish the validity of the DSM-III in the schools the authors looked at the differences in the number and types of DSM-III diagnoses when compared with CBCL classifications. No differences were apparent in the DSM-III diagnoses assigned to the boys classified as having a significant behavior disorder and those classified as having no behavior disorder by the CBCL system. Tharinger and her colleagues (1986) conclude that this finding questions the validity of the child and adolescent disorders of DSM-III and as a result further research is needed on this classification system.
Conclusions

The validity of the DSM-III-R in the diagnosis of child and adolescent psychopathology has yet to be adequately established. Achenbach (1980) has reported that while there is some overlap with empirically derived syndromes and DSM-III diagnostic categories, some DSM-III diagnoses have not been found in empirical analyses and some empirically derived syndromes have no clear counterparts in the DSM-III. According to Achenbach (1980), few empirical studies provide evidence to support the DSM-III syndromes of attention deficit disorder without hyperactivity (inattention does however appear in other syndromes without hyperactivity), undersocialized conduct disorders without aggression, separation or avoidant disorders (with the exception of generalized withdrawal and anxiety), oppositional disorders (other than aggressive), identity disorders, and pervasive developmental disorders (other than schizoid).

It should be noted that little research exists which has utilized the DSM-III-R with school aged children and adolescents, however the differences between the DSM-III and DSM-III-R are minimal. The American Psychiatric Association is currently in the process of producing the DSM-IV which hopefully will address some of the
criticisms of the DSM-III-R. The issue of the validity of the DSM does not all together preclude the use of the DISC and DICA by psychologists, although it does indicate that these interviews should be interpreted in light of the existing research on the validity of the DSM-III.

The DISC and DICA are especially appropriate for those school psychologists not trained in the use of the DSM. These interviews can be administered and scored by persons not familiar with the DSM system and are useful for classification purposes. All psychologists utilizing the DSM need to be aware of its limitations which largely stem from its lack of adequate validity data.
Chapter III

Methodology

Subject Selection

Clinical Group

The subjects comprising the sample for the present study are 154 children and adolescents referred to 1 of 4 Children's Hospital Guidance Centers in central Ohio between January and November, 1992. These children and adolescents were referred for psychological assessment by their parents. Parental permission was obtained from the parents of the subjects included in this study. The age range of the subjects utilized in this study was predetermined. Subjects were selected if their age fell within the age ranges of the assessment instruments employed in the present investigation. Age and sex distribution were not predetermined. All age appropriate children and adolescents who were referred, and whose parents agreed to participate, were involved in the present investigation.
**Contrast Group**

Data for the contrast (control) group was randomly generated through the Random Number Generation option in the Analysis function of Excel 4.0. Both standard scores with a mean of 100 and a standard deviation of 15, and a mean of 50 and a standard deviation of 10 were generated for comparison with the scores of the subjects in the clinical group. This was done for two reasons: 1) the objective of this study was to examine the psychometric properties of the DAP:SPED, DSP, and DSF, not group characteristics, and 2) clinical research has shown that when a clinical group is compared to a normal control group on a psychological measure of behavioral and/or emotional problems, the control group approximates the normal distribution with respect to their test scores (see Naglieri, McNeish, & Bardos, 1990; Naglieri, LeBuffe, & Pfeiffer, 1993; Naglieri, LeBuffe, & Pfeiffer, in press; Naglieri & Gottling, 1992; Naglieri & Pfeiffer, 1992. It is important to point out that this is not a Monte Carlo experiment. In a Monte Carlo experiment, scores are constructed to have the same mean but different shapes and variances (Keppel, 1982). In the present study, the scores were generated to approximate the scores of a normal sample.
Procedure

All subjects were administered the Draw A Person: Screening Procedure for Emotional Disturbance (DAP:SPED) (Naglieri, McNeish, & Bardos, 1991), Child Behavior Checklist (CBCL) (Achenbach, 1991), Conners' Parent Rating Scale (CPRS) (Conners, 1990), Devereux School Form (DSF) (Naglieri, LeBuffe, & Pfeiffer, in press), and Devereux Scales of Psychopathology (DSP) (Naglieri, LeBuffe, & Pfeiffer, in press) by Children's Hospital Guidance Center mental health clinicians who were assigned to the subjects' cases. Each clinician scored his/her own test protocols.

The clinicians who participated in this study were 3 licensed Ph.D. psychologists, 2 licensed M.A. social workers (LSW), 1 licensed professional clinical counselor (LPCC), 6 licensed M.S.W. social workers (LSW), 12 M.S.W. licensed independent social workers (LISW), and 1 predoctoral psychology intern.
Assessment Measures

A brief description of the three psychological measures utilized in the present study is presented below.

Draw A Person: Screening Procedure for Emotional Disturbance

The Draw A Person: Screening Procedure For Emotional Disturbance (DAP:SPED) was designed as a screening measure which differentiates the drawings (man, women, and self) of normal children and adolescents from those children and adolescents with emotional or behavior problems. It is an objective, standardized projective scoring system which uses an actuarial approach to the identification of children and adolescents (ages 6-17) with emotional or behavioral problems and is not intended as a differential diagnostic tool (Naglieri et al., 1991).

Devereux Behavior Rating Scale - School Form

The Devereux Behavior Rating Scale - School Form (DSF) (Naglieri, LeBuffe & Pfeiffer, in press) is a behavior rating scale which measures behaviors corresponding to the federal definition of Severe
Emotional Disturbance. The DSF consists of four scales (10 items each) which correspond to the four areas of the federal definition: Interpersonal Problems, Inappropriate Behaviors/Feelings, Depression, and Physical Symptoms/Fears. The sum of the 40 items yields a Total Test standard score with a mean of 100 and a standard deviation of 15.

**Devereux Scales of Psychopathology**

The Devereux scales of psychopathology (DSP) (Naglieri, LeBuffe, & Pfeiffer, in press) is a parent-report behavior rating scale designed to assist in the identification of behavior signs related to particular categories, e.g. P.L. 94-142 & DSM-III-R, of psychopathology (Naglieri, et al., in press). The DSP consists of three levels. The first level is the Scale level. For the 5-12 year old version the scales are: Conduct, Attention, Withdrawal/Depression, Anxiety, Acute Problems, and Autistic. For the 13-18 year old version all the scales are the same except for the Attention scale. A Delinquency scale replaces the Attention scale on the 13-18 year old version. On both versions, the scales are divided into three Composite areas: Externalizing, Internalizing, and Psychotic. The three
composites are added together to yield a Total Test Score. Naglieri, LeBuffe, & Pfeiffer (in press) indicate in their test manual that the DSP was developed to assist in the identification of behavior signs related to particular categories (P.L. 94-142 & DSM-III-R) of psychopathology.

**Statistical Transformations of Raw Scores to Standard Scores**

The DSP and DSF standard scores presented in the next chapter were converted from raw scores to standard scores by the author. The following formula was used to compute the 6 DSP Factor Scores, 3 DSP Composite Scores, and 1 Total Score:

\[
\text{Formula A: } (\text{RS2 - RSMean/RSsd}) \times 10 + 50
\]

where RS is the raw score total for the standard score of the Factor being computed, e.g. Conduct, RS mean is the mean of the standardization sample, and RS sd is the standard deviation of the standardization sample. The raw score totals obtained on the 6 DSP Factors, 3 Composites, and 1 Total Score were transformed into
standard scores with a mean of 50 and a standard deviation of 10.

The following formula was used to compute the 4 DSF Scale Scores and 1 DSF Total Score:

Formula B: \((RS2 - RS\text{Mean}/RS\text{sd}) \times 15 + 100\)

where RS is the raw score total for the standard score of the Scale being computed, e.g. Interpersonal Problems, RS mean is the mean of the standardization sample, and RS sd is the standard deviation of the standardization sample. The raw score totals obtained on the 4 DSF Factors and 1 Total Score were transformed into standard scores with a mean of 100 and a standard deviation of 15.

Data Analysis

The data were analyzed in accordance with the research questions outlined in chapter 1. The statistical analyses used to answer these questions are presented below.
1. Will there be a significant difference between scores earned by children in a clinical group and a contrast group on DAP:SPED, DSP, and DSF Total Scores?

This question was answered through a multivariate analysis of variance (MANOVA). The independent variable for this analysis was group (clinical and nonclinical). The dependent variables were the groups' total scores on the DAP:SPED, DSP, and DSF.

2. What is the degree of discrimination of the DSF 4 Scale Scores, the DSP 6 Factor Scores, and the DSP 3 Composite Scores for a sample of children in clinical group and a contrast group?

This question was answered through a discriminant-function analysis. The dependent measures for this analysis were the subjects' DSF 4 Scale Scores, DSP 6 Factor Scores, and DSP 3 Composite Scores. The relative contributions of the dependent variables to the group differences were determined through the interpretation of discriminant function standardized coefficients. A
classification analysis was used to evaluate the
diagnostic utility of the discriminations.

3. What are the percentages of true positives and
true negatives for the DAP:SPED, DSF, and
DSP Total Scores at various cut-off scores, for a
sample of children in a clinical group and a contrast
group?

a. What are the percentages of true positives and
true negatives for the DSF 4 Scale Scores, at
various cut-off scores, for a sample of
children in a clinical group and a contrast
group?

b. What are the percentages of true positives and
true negatives for the DSP 6 Factor Scores, at
various cut-off scores, for a sample of
children in a clinical group and a contrast
group?

c. What are the percentages of true positives and
true negatives for the DSP 3 Composite Scores,
at various cut-off scores, for a sample of
children in a clinical group and a contrast
group?
These questions were answered through an analysis of sensitivity (the proportion of clinical subjects correctly identified or true positives) and specificity (the proportion of nonclinical subjects correctly identified or true negatives). The sensitivity and specificity results for the various scores from each behavior rating scale were plotted.
Chapter IV

Results

This chapter is organized into the following sections: 1) summary of descriptive statistics, 2) summary of multivariate analysis, 3) description of discriminant analysis, 4) summary of discriminant analysis, 5) description of analysis of sensitivity and specificity, 6) summary of analysis of sensitivity and specificity, and 6) summary of results.

Summary of Descriptive Statistics

Data describing the distribution of the sample by age, sex, and race are presented in Table 1. The clinical group consisted of 153 children and adolescents: 71 males and 41 females between the ages of 5-12, and 19 males and 22 females between the ages of 13-18. As indicated in Table 1, the sample was predominantly white with African Americans being most represented in the 5-12 year old female group (13.7%), and least represented in the 13-18 year old female group (1%).
Table 1
Age, Race, and Gender Characteristics of the Clinical Group

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>sd</th>
<th>Race</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>White</td>
<td>Afr. Am.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-12</td>
<td>71</td>
<td>9.16</td>
<td>2.08</td>
<td>94.1%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-12</td>
<td>41</td>
<td>8.90</td>
<td>2.23</td>
<td>86.3%</td>
<td>13.7%</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-18</td>
<td>19</td>
<td>13.37</td>
<td>0.60</td>
<td>86.2%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-18</td>
<td>22</td>
<td>15.41</td>
<td>1.22</td>
<td>99.0%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

* Age reported in years

Table 2 shows the total test means, standard deviations, and minimum/maximum scores by age and sex for the clinical group and the contrast group. A comparison of these two tables indicates that the clinical groups' overall test scores are higher than those of the contrast group. The minimum scores of the clinical group are approximately 2/3 of a standard deviation lower than
those of the contrast group. The maximum scores of the clinical group are approximately 2/3 of a standard deviation higher than those of the contrast group.

Table 2

Total Test Means, Standard Deviations, and Minimum/Maximum Scores for the Clinical and Contrast Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min./Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>DAP:SPED</td>
<td>47.84</td>
<td>10.38</td>
<td>29-80</td>
</tr>
<tr>
<td></td>
<td>DSP</td>
<td>84.26</td>
<td>7.55</td>
<td>27-90</td>
</tr>
<tr>
<td></td>
<td>DSF</td>
<td>143.24</td>
<td>15.53</td>
<td>64-160</td>
</tr>
<tr>
<td>Contrast</td>
<td>DAP:SPED</td>
<td>50.81</td>
<td>9.67</td>
<td>29-69</td>
</tr>
<tr>
<td></td>
<td>DSP</td>
<td>50.40</td>
<td>8.90</td>
<td>65-90</td>
</tr>
<tr>
<td></td>
<td>DSF</td>
<td>100.02</td>
<td>14.45</td>
<td>113-160</td>
</tr>
</tbody>
</table>

*Note.* DAP:SPED = Draw A Person: Screening Procedure for Emotional Disturbance; Devereux Scales of Psychopathology; DSF = Devereux Behavior Rating Scale - School Form.

*Note.* DAP:SPED and DSF have a normative mean of 100 and a sd of 15. DSP has a normative mean of 50 and a sd of 10.
Table 3 provides a breakdown of the DSM-III-R diagnoses of the subjects in the clinical group. Adjustment disorders represented the biggest percentage of the group (42%). 33% of the subjects had an attention deficit disorder diagnosis.

Table 3

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment Disorder</td>
<td>59</td>
<td>42%</td>
</tr>
<tr>
<td>Attention Deficit Disorder</td>
<td>47</td>
<td>33%</td>
</tr>
<tr>
<td>Developmental Disorder</td>
<td>6</td>
<td>4%</td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>5</td>
<td>3%</td>
</tr>
<tr>
<td>Functional Enuresis</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>23</td>
<td>16%</td>
</tr>
</tbody>
</table>

Note. Adjustment disorders includes all subtypes. Attention Deficit Disorder includes undifferentiated type.
Table 4 provides the Pearson $r$ correlations between the DAP:SPED, DSP, and DSF Factor, Composite, and Total scores for the total sample. The DAP:SPED correlated low with the DSP and DSF (lowest correlation $= -.005$ with the DSF Total Score; highest correlation $= .08$ with the DSP Attention Score).

The DSP Total Score correlated highly with the DSF Total Score ($r = .82$). Furthermore, Table 4 indicates that the DSP and DSF Factor Scores are highly correlated. The lowest correlation between these factors was $-.01$ (DSP Acute Problems and DSF Inappropriate Behaviors/Feelings). The highest correlation between these factors was $.90$ (DSP Depression and DSF Depression).

The intercorrelations of the Factors comprising both the DSP and DSF are moderate to high. The lowest intercorrelation on the DSP existed between the Autistic and Acute Problems Factors ($r = -.15$). The highest intercorrelation existed between the DSP Depression and Anxiety Factors ($r = .71$). The lowest intercorrelation on the DSF was found between the Inappropriate Behaviors/Feelings and Physical Symptoms/Fears Factors ($r = .70$). The highest intercorrelation was found
between the Inappropriate Behaviors/Feelings and Interpersonal Problems Factors ($r = .79$).
### Table 4

Obtained Pearson r Correlations Between DAP:SPED, DSP, and DSF Factor, Scale, Composite, and Total Scores for the Clinical Group (N = 153)

|       | SPED | DSPCT | DSPANT | DSPAVT | DSPATT |DSPAYT |DSPMTT | DSPTRT | DSPFTT | DSPISS | DSPFSS | DSPIFS | DSFSS | DSPSSS | DSPFTSS |
|-------|------|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| SPED  | 1    |       |        |        |        |       |       |       |       |       |       |       |       |       |         |
| DSPCT | 0.002| 1     |        |        |        |       |       |       |       |       |       |       |       |       |         |
| DSPANT| -0.02| 0.02  | 1      |        |        |       |       |       |       |       |       |       |       |       |         |
| DSPAVT| 0.62 | 0.91  | 0.77   | 1      |        |       |       |       |       |       |       |       |       |       |         |
| DSPATT| 0.56 | 0.65  | 0.56   | 0.30   | 1      |       |       |       |       |       |       |       |       |       |         |
| DSPAYT| 0.95 | 0.62  | 0.65   | 0.56   | 0.56   | 1     |       |       |       |       |       |       |       |       |         |
| DSPMTT| 0.86 | 0.63  | 0.65   | 0.56   | 0.56   | 0.54  | 1     |       |       |       |       |       |       |       |         |
| DSPTRT| 0.73 | 0.63  | 0.65   | 0.56   | 0.56   | 0.54  | 0.54 | 1     |       |       |       |       |       |       |         |
| DSPFTT| 0.62 | 0.63  | 0.65   | 0.56   | 0.56   | 0.54  | 0.54 | 0.54 | 1     |       |       |       |       |       |         |
| DSPISS| 0.02 | 0.02  | 0.02   | 0.02   | 0.02   | 0.02  | 0.02 | 0.02 | 0.02 | 1     |       |       |       |       |         |
| DSPFSS| 0.02 | 0.02  | 0.02   | 0.02   | 0.02   | 0.02  | 0.02 | 0.02 | 0.02 | 0.02 | 1     |       |       |       |         |
| DSFSS | 0.02 | 0.02  | 0.02   | 0.02   | 0.02   | 0.02  | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 1     |       |       |         |

Note. SPED = DAP:SPED Total T Score; DSPCT = DSP Conduct T Score; DSPANT = DSP Anxiety T Score; DSPAVT = DSP Depression T Score; DSPATT = DSP Atypical T Score; DSPAYT = DSP Acute Problems T Score; DSPMTT = DSP Attention T Score; DSPTRT = DSP Externally Oriented T Score; DSPISS = DSP Psychotic T Score; DSFSS = DSP Total Y Score; DSPFSS = DSP Interpersonal Problems Standard Score (mean = 100, sd = 15); DSFSS = DSP Inappropriate Behaviors/Feelings SS; DSFSS = DSP Depression SS; DSFSS = DSP Physical Symptoms/Feats SS; DSFSS = DSP Total Standard Score.
Summary of Multivariate Analysis

Table 5 provides a summary of the multivariate analysis of the DAP:SPED, DSP, and DSF Total Test Scores by group. This analysis indicated that the independent variable group (clinical and contrast) can predict subjects' scores on the DAP:SPED: $F(1, 289), p < .05$; the DSP: $F(1, 289), p < .0001$; and the DSF: $F(1, 289), p < .0001$). The omnibus multivariate test statistic, $= .1737: F(3, 287), p < .0001$, indicated that the dependent variables (DAP:SPED, DSP, and DSF overall scores) can predict group (clinical and contrast) membership.

Table 5

Summary of Multivariate Analysis of DAP:SPED, DSP, and DSF Total Test Scores by Group

<table>
<thead>
<tr>
<th>Source</th>
<th>NDF</th>
<th>DDF</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAP:SPED</td>
<td>1</td>
<td>289</td>
<td>5.7521</td>
<td>.0162</td>
</tr>
<tr>
<td>DSP</td>
<td>1</td>
<td>289</td>
<td>1242.9250</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>DSF</td>
<td>1</td>
<td>289</td>
<td>585.3177</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Description of Discriminant Analysis

According to Stevens (1992), "Discriminant analysis is used for two purposes: (1) describing major differences among the groups in MANOVA, and (2) classifying subjects into groups on the basis of a battery of measurements" (p. 273).

Stevens (1992) states that there are two possible methods in use for interpreting a discriminant function analysis. The first is to examine the standardized coefficients (obtained by multiplying the raw coefficient for each variable by the standard deviation for that variable). The second is to examine the discriminant function-variable correlations (the correlations between each discriminant function and each of the original variables). Because partial correlations between the DAP:SPED, DSP, and DSF Factor, Composite, and Total Scores for the total sample have already been presented (see Table 4), and only one function was generated (the number of discriminant functions = the minimum of p and (k-1) where k is the number of groups and p is the number of dependent variables), the standardized coefficients were used to analyze the discriminant function results in the present study. "The coefficients are partial
coefficients, with the effects of the other variables removed" (Stevens, 1992, p. 276).

Standardized coefficients allow for a determination of redundancy between a given variable and other variables in the function (Stevens, 1992). The standardized coefficient indicates how much each function variable (DSP and DSF test scores) contributes, and in what direction to the differentiation between groups (clinical and nonclinical). For clinical purposes, variables in the present study whose standardized coefficients are more than one half of the largest standardized coefficient in the function are considered significant (see Fiumera, 1993).

In order to assure that variables selected as most important in interpreting discriminant functions would show up again in subsequent studies, the $N$ (total sample size)/$p$ (number of variables) ratio needs to be around 20:1 (Stevens, 1992). In the present study, the $N/p$ ratio = 50:1 for the measure with the most variables (DSP).
Summary of Discriminant Analysis of the DSF 4 Scale Scores

Table 6 presents unweighted marginal means from the one-way MANOVA of the DSF 4 Scale Scores. As expected, the mean scores of the clinical group (IP = 137.89; IBF = 138.38; DEP = 141.39; PSF = 137.59) exceed those of the contrast group (IP = 99.76; IBF = 99.12; DEP = 99.82; PSF = 100.12).

Table 6

Unweighted Marginal Means From One-Way MANOVA of DSF 4 Scale Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Clinical Means (n=147)</th>
<th>Contrast Means (n=153)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>137.89</td>
<td>99.76</td>
</tr>
<tr>
<td>IBF</td>
<td>138.38</td>
<td>99.12</td>
</tr>
<tr>
<td>DEP</td>
<td>141.39</td>
<td>99.82</td>
</tr>
<tr>
<td>PSF</td>
<td>137.59</td>
<td>100.12</td>
</tr>
</tbody>
</table>

Note. The unweighted discriminant function marginal mean for the clinical group = 12.10. The unweighted discriminant function marginal mean for the contrast group = 8.67.
The 4 Scale Scores were subjected initially to a two-group discriminant function analysis with an accompanying one-way MANOVA. These analyses were performed by General ANOVA in the MacSS statistical program and MANOVA in the JMP (SAS for Macintosh) program.

The omnibus multivariate test statistic, $F(4,295)$, $p < .0001$, indicated that the clinical and nonclinical groups were not homogeneously distributed with respect to the principal underlying discriminant function. The overall ability of the resultant function to discriminate between the clinical and contrast groups can be inferred from the overall hit-miss ratio (the percentage of correct group classification should the function be used for purposes of prediction) that was obtained in an ancillary analysis (from JMP MANOVA). Table 7 shows the accuracy of classification of subjects based on discriminant function scores for the DSF 4 Scale Scores. The DSF 4 Scale Scores accurately classified 100% of the subjects in the clinical group (hit-miss ratio of 131/131) and 90% of the subjects in the nonclinical group (hit-miss ratio = 153/169).
Table 7

Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSF Scale Scores

<table>
<thead>
<tr>
<th>Actual Group</th>
<th>Predicted Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Clinical</td>
<td>131</td>
<td>0</td>
</tr>
<tr>
<td>Contrast</td>
<td>16</td>
<td>153</td>
</tr>
</tbody>
</table>

|              | 147 | 153 | 300 |

The relative ability of the function to account for the overall discriminatory power, given by the Pillai-Bartlett Trace, was observed to be .75 (75%). A summary of results given by the multivariate analysis, along with results provided by univariate ANOVAs on the two groups, can be found in Table 8.

An examination of the standardized weights, shown in Table 8, suggested that the discriminant function separated the groups on the basis of the DSF 4 Scale Scores. The standardized weights indicate that the DSF Depression Scale was the best discriminator between
groups (standardized weight = .43) The Inappropriate Behaviors/Feelings (.35) and Physical Symptoms/Fears (.30), and the Interpersonal Problems Factor (.27) were also significant predictors.

Table 8
Summary of Univariate and Multivariate Analyses of DSF 4 Scale Scores

<table>
<thead>
<tr>
<th>Variable*</th>
<th>NDF</th>
<th>DDF</th>
<th>p**</th>
<th>Raw</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0169</td>
<td>.2668**</td>
</tr>
<tr>
<td>IBF</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0221</td>
<td>.3456**</td>
</tr>
<tr>
<td>DEP</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0285</td>
<td>.4334**</td>
</tr>
<tr>
<td>PSF</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.3044</td>
<td>.3044**</td>
</tr>
</tbody>
</table>

* IP = Interpersonal Problems; IBF = Inappropriate Behaviors/Feelings; DEP = Depression; PSF = Physical Symptoms/Fears.

** Indicates Significance of the Variable as a Predictor.
Summary of Discriminant Analysis of the DSP 6 Factor Scores

Table 9 presents unweighted marginal means from the one-way MANOVA of the DSP 6 Factor Scores. As expected, the mean scores of the clinical group (COND = 74.99; ANX = 76.35; DEP = 77.97; AUT = 77.85; ACP = 87.20; ATTN = 74.99) exceed those of the contrast group (COND = 50.08; ANX = 50.09; DEP = 50.19; AUT = 49.20; ACP = 49.59; ATTN = 49.59).
Table 9
Unweighted Marginal Means From One-Way MANOVA of DSP 6 Factor Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Clinical Means (n=147)</th>
<th>Contrast Means (n=153)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COND</td>
<td>74.99</td>
<td>50.08</td>
</tr>
<tr>
<td>ANX</td>
<td>76.35</td>
<td>50.09</td>
</tr>
<tr>
<td>DEP</td>
<td>77.97</td>
<td>50.19</td>
</tr>
<tr>
<td>AUT</td>
<td>77.85</td>
<td>49.20</td>
</tr>
<tr>
<td>ACP</td>
<td>87.20</td>
<td>49.59</td>
</tr>
<tr>
<td>ATTN</td>
<td>74.99</td>
<td>49.59</td>
</tr>
</tbody>
</table>

**Note.** The unweighted discriminant function marginal mean for the clinical group = 14.55. The unweighted functional marginal mean for the contrast group = 8.81.

The 6 Factor Scores were subjected initially to a two-group discriminant function analysis with an accompanying one-way MANOVA. These analyses were performed by General ANOVA in the MacSS statistical program and MANOVA in the JMP (SAS for Macintosh) program.
The omnibus multivariate test statistic, $F(6,293), p < .0001$, indicated that the clinical and contrast groups were not homogeneously distributed with respect to the principal underlying discriminant function. The overall ability of the resultant function to discriminate between the clinical and contrast groups can be inferred from the overall hit-miss ratio (the percentage of correct group classification should the function be used for purposes of prediction) that was obtained in an ancillary analysis (from JMP MANOVA).

Table 10 shows the accuracy of classification of subjects based on discriminant function scores for the 6 DSP Factor Scores. The DSP 6 Factor Scores accurately classified 100% of the subjects in the clinical group (hit-miss ratio = 146/146) and 99% of the subjects in the contrast group (hit-miss ratio = 153/154).
Table 10
Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSP Factor Scores

<table>
<thead>
<tr>
<th>Actual Group</th>
<th>Predicted Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>146</td>
<td>0</td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>Contrast</td>
<td>1</td>
<td>153</td>
<td></td>
<td>154</td>
</tr>
</tbody>
</table>

| 147 | 153 | 300 |

The relative ability of the function to account for the overall discriminatory power, given by the Pillai-Bartlett Trace, was observed to be .90 (90%). A summary of results given by the multivariate analysis, along with results provided by univariate ANOVAs on the two groups, can be found in Table 11.

An examination of the standardized weights, shown in Table 11, suggested that the discriminant function separated the groups on the basis of the DSF 6 Factor Scores. The standardized weights indicate that the DSP Acute Problems Factor was the best discriminator between groups (standardized weight = .90). The Autistic Factor
was also significant (standardized weight = .37). The Conduct, Anxiety, Depression, and Attention factors were not significant.

Table 11
Summary of Univariate and Multivariate Analyses of DSP 6 Factor Scores

<table>
<thead>
<tr>
<th>Variable*</th>
<th>NDF</th>
<th>DDF</th>
<th>p**</th>
<th>Raw</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>COND</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0011</td>
<td>.0120</td>
</tr>
<tr>
<td>ANX</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0168</td>
<td>.0179</td>
</tr>
<tr>
<td>DEP</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0195</td>
<td>.1969</td>
</tr>
<tr>
<td>AUT</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0377</td>
<td>.3665**</td>
</tr>
<tr>
<td>ACP</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0863</td>
<td>.7396**</td>
</tr>
<tr>
<td>ATTN</td>
<td>1</td>
<td>298</td>
<td>&lt;.0001</td>
<td>.0161</td>
<td>.1716</td>
</tr>
</tbody>
</table>

* COND = Conduct; ANX = Anxiety; DEP = Depression, AUT = Autistic; ACP = Acute Problems; ATTN = Attention.

** Indicates Significance of the Variable as a Predictor.
Summary of Discriminant Analysis of the DSP 3 Composite Scores

Table 12 presents unweighted marginal means from the one-way MANOVA of the DSP 3 Composite Scores. As expected, the mean scores of the clinical group (EXT = 76.83; ANX = 79.11; PSY = 87.24) exceed those of the contrast group (EXT = 50.84; ANX = 49.54; PSY = 49.41).

Table 12
Unweighted Marginal Means From One-Way MANOVA of DSP 3 Composite Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Clinical Means (n=148)</th>
<th>Contrast Means (n=153)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT</td>
<td>76.83</td>
<td>50.84</td>
</tr>
<tr>
<td>INT</td>
<td>79.11</td>
<td>49.54</td>
</tr>
<tr>
<td>PSY</td>
<td>87.24</td>
<td>49.41</td>
</tr>
</tbody>
</table>

Note. The unweighted discriminant function marginal mean for the clinical group = 13.21. The unweighted discriminant function marginal mean for the contrast group = 7.78.
The 3 Composite Scores were subjected initially to a two-group discriminant function analysis with an accompanying one-way MANOVA. These analyses were performed by General ANOVA in the MacSS statistical program and MANOVA in the JMP (SAS for Macintosh) program.

The omnibus multivariate test statistic, $F(3,297)$, $p < .0001$, indicated that the clinical and contrast groups were not homogeneously distributed with respect to the principal underlying discriminant function. The overall ability of the resultant function to discriminate between the clinical and contrast groups can be inferred from the overall hit-miss ratio (the percentage of correct group classification should the function be used for purposes of prediction) that was obtained in an ancillary analysis (from JMP MANOVA). Table 13 shows the accuracy of classification of subjects based on discriminant function scores for the 3 DSP Composite Scores. The DSP 3 Composite Scores accurately classified 99% of the subjects in the clinical group (hit-miss ratio = 145/146) and 98% of the subjects in the contrast group (hit-miss ratio = 152/155).
Table 13

Accuracy of Classification of Subjects Based on Discriminant Function Scores From DSP Composite Scores

<table>
<thead>
<tr>
<th>Actual Group</th>
<th>Predicted Group</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>145</td>
<td>1</td>
<td>146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>3</td>
<td>152</td>
<td>155</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>148</td>
<td>153</td>
<td>301</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The relative ability of the function to account for the overall discriminatory power, given by the Pillai-Bartlett Trace, was observed to be .88 (88%). A summary of results given by the multivariate analysis, along with results provided by univariate ANOVAs on the two groups, can be found in Table 14.

An examination of the standardized weights, shown in Table 14, suggested that the discriminant function separated the groups on the basis of the DSP 3 Composite Scores. The standardized weights indicate that the DSP Psychotic Factor was the best discriminator between groups (standardized weight = .81). The Internalizing
Composite approached significance (standardized weight = .39). The externalizing composite was not significant.

Table 14
Summary of Univariate and Multivariate Analyses of DSP 3 Composite Scores

<table>
<thead>
<tr>
<th>Variable*</th>
<th>NDF</th>
<th>DDF</th>
<th>p**</th>
<th>Raw</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT</td>
<td>1</td>
<td>299</td>
<td>&lt;.0001</td>
<td>.0164</td>
<td>.1804</td>
</tr>
<tr>
<td>INT</td>
<td>1</td>
<td>299</td>
<td>&lt;.0001</td>
<td>.0168</td>
<td>.3931</td>
</tr>
<tr>
<td>PSY</td>
<td>1</td>
<td>299</td>
<td>&lt;.0001</td>
<td>.0102</td>
<td>.8100**</td>
</tr>
</tbody>
</table>

* EXT = Externalizing; INT = Internalizing; PSY = Psychotic.

** Indicates Significance of the Variable as a Predictor.
Description of Analysis of Sensitivity and Specificity

Verhulst & Koot (1992) have stated that "The criterion-referenced validity of a test for assessing general psychopathology in children can be measured by its ability to do what it is supposed to do, that is, correctly identify children who can be regarded disordered (cases) as "test-positive" and those who can be regarded normal (noncases) as "test-negative" " (p. 48). Test-positive refers to the number of cases for whom the test is positive. Expressed as a proportion (number of test positive cases/n) this value refers to sensitivity. Test-negative refers to the number of noncases for whom the test is negative. Expressed as a proportion (number of test negative cases/n) this value refers to specificity. As Verhulst & Koot (1992) point out, "...sensitivity and specificity do not reflect an intrinsic quality of a test. They are not absolute values. Sensitivity and specificity will vary with the samples on which they were based and with the critical values chosen" (p. 49).
Summary of Analysis of Sensitivity and Specificity

Analyses of the sensitivity and specificity of DSP Factor, Composite, and Total Scores, DSF Scale and Total Scores, and DAP:SPED Total Scores are presented in Figures 1-16. For all three measures, the lower the score used to identify a subject, the lower the specificity. As the scores increased, so did the specificity values. Conversely, the higher the cut-off score, the lower the sensitivity. Results of individual analyses of sensitivity and specificity are presented next by test and type of score analyzed.

DSP 6 Factors

Figure 1

Figure 1 shows the sensitivity and specificity for the DSP Conduct Factor for the clinical and contrast groups using various cut-off scores. Figure 1 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around three-quarters of a standard deviation. The DSP Conduct Factor correctly classified 87% of the subjects in the clinical group and 93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the
mean) is used, the DSP Conduct Factor correctly classified 62% of the clinical group and 100% of the contrast group.

Figure 2

Figure 2 shows the sensitivity and specificity for the DSP Anxiety Factor for the clinical and contrast groups using various cut-off scores. Figure 2 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around three-quarters of a standard deviation. The DSP Anxiety Factor correctly classified 84% of the subjects in the clinical group and 92% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Anxiety Factor correctly classified 55% of the clinical group and 100% of the contrast group.
Figure 1

Sensitivity and Specificity for the DSP Conduct Factor for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 2

Sensitivity and Specificity for the DSP Anxiety Factor for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 3

Figure 3 shows the sensitivity and specificity for the DSP Depression Factor for the clinical and contrast groups using various cut-off scores. Figure 3 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one standard deviation. The DSP Depression Factor correctly classified 89% of the subjects in the clinical group and 95% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Depression Factor correctly classified 58% of the clinical group and 99% of the contrast group.

Figure 4

Figure 4 shows the sensitivity and specificity for the DSP Autistic Factor for the clinical and contrast groups using various cut-off scores. Figure 4 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one standard deviation. The DSP Autistic Factor correctly classified 91% of the subjects in the clinical group and 95% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score
of 70 (2 standard deviations above the mean) is used, the DSP Autistic Factor correctly classified 58% of the clinical group and 100% of the contrast group.

**Figure 5**

Figure 5 shows the sensitivity and specificity for the DSP Acute Problems Factor for the clinical and contrast groups using various cut-off scores. Figure 5 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one and a half standard deviations. The DSP Acute Problems Factor correctly classified 99% of the subjects in the clinical group and 93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Acute Problems Factor correctly classified 95% of the clinical group and 99% of the contrast group.
Figure 3

Sensitivity and Specificity for the DSP Depression Factor for Contrast and Clinical Groups Using Various cut-off Scores
Figure 4

Figure 5

Sensitivity and Specificity for the DSP Acute Problems Factor for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 6

Figure 6 shows the sensitivity and specificity for the DSP Attention Factor for the clinical and contrast groups using various cut-off scores. Figure 6 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one half of a standard deviation. The DSP Attention Factor correctly classified 77% of the subjects in the clinical group and 92% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Attention Factor correctly classified 52% of the clinical group and 100% of the contrast group.
Figure 6

Sensitivity and Specificity for the DSP Attention Factor for Contrast and Clinical Groups Using Various Cut-off Scores
DSP 3 Composites

Figure 7

Figure 7 shows the sensitivity and specificity for the DSP Externalizing Composite for the clinical and contrast groups using various cut-off scores. Figure 7 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one standard deviation. The DSP Externalizing Composite correctly classified 84% of the subjects in the clinical group and 88% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Externalizing Composite correctly classified 55% of the clinical group and 99% of the contrast group.

Figure 8

Figure 8 shows the sensitivity and specificity for the DSP Internalizing Composite for the nonclinical and clinical groups using various cut-off scores. Figure 8 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one standard deviation. The DSP Internalizing Composite correctly classified 88% of the
subjects in the clinical group and 93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Internalizing Composite correctly classified 68% of the clinical group and 99% of the contrast group.

Figure 9

Figure 9 shows the sensitivity and specificity for the DSP Psychotic Composite for the nonclinical and clinical groups using various cut-off scores. Figure 9 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around two standard deviations. The DSP Psychotic Composite correctly classified 99% of the subjects in the clinical group and 93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Psychotic Composite correctly classified 98% of the clinical group and 99% of the contrast group.
Figure 7
Sensitivity and Specificity for the DSP Externalizing Composite for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 8

Sensitivity and Specificity for the DSP Internalizing Composite for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 9

Sensitivity and Specificity for the DSP Psychotic Composite for Contrast and Clinical Groups Using Various Cut-off Scores
DSP Total Score

Figure 10

Figure 10 shows the sensitivity and specificity for the DSP Total Score for the clinical and contrast groups using various cut-off scores. Figure 10 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one and a half standard deviations. The DSP Total Score correctly classified 99% of the subjects in the clinical group and 92% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSP Total Score correctly classified 87% of the clinical group and 99% of the contrast group.
Figure 10

Sensitivity and Specificity for the DSP Total Score for Contrast and Clinical Groups Using Various Cut-off Scores
DSF 4 Scales

Figure 11

Figure 11 shows the sensitivity and specificity for the DSF Interpersonal Scale for the clinical and contrast groups using various cut-off scores. Figure 11 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one half of a standard deviation. The DSF Interpersonal Scale correctly classified 76% of the subjects in the clinical group and 92% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSF Interpersonal Scale correctly classified 52% of the clinical group and 100% of the contrast group.

Figure 12

Figure 12 shows the sensitivity and specificity for the DSF Inappropriate Behaviors/Feelings Scale for the clinical and contrast groups using various cut-off scores. Figure 12 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one half of a standard deviation. The DSF Inappropriate Behaviors/Feelings Scale correctly
classified 80% of the subjects in the clinical group and 97% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSF Inappropriate Behaviors/Feelings Scale correctly classified 53% of the clinical group and 100% of the contrast group.

Figure 13

Figure 13 shows the sensitivity and specificity for the DSF Depression Scale for the clinical and contrast groups using various cut-off scores. Figure 13 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around three quarters of a standard deviation. The DSF Depression Scale correctly classified 87% of the subjects in the clinical group and 93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSF Depression Scale correctly classified 60% of the clinical group and 99% of the contrast group.
Figure 11

Sensitivity and Specificity for the DSF Interpersonal Problems Scale for Contrast and Clinical Groups Using Various Cut-off Scores
Sensitivity and Specificity for the DSF Inappropriate Behaviors/Feelings Scale for Contrast and Clinical Groups Using Various Cut-off Scores

Figure 12
Figure 13

Sensitivity and Specificity for the DSF Depression Scale for Contrast and Clinical Groups Using Various Cut-off Scores
Figure 14

Figure 14 shows the sensitivity and specificity for the DSF Physical Symptoms/Fears Scale for the clinical and contrast groups using various cut-off scores. Figure 14 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around three quarters of a standard deviation. The DSF Physical Symptoms/Fears Scale correctly classified 77% of the subjects in the clinical group and 94% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSF Depression Scale correctly classified 49% of the clinical group and 99% of the contrast group.

DSF Total Score

Figure 15

Figure 15 shows the sensitivity and specificity for the DSF Total Score for the clinical and contrast groups using various cut-off scores. Figure 15 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around one standard deviation. The DSF Total Score correctly classified 88% of the subjects in the clinical group and
93% of the contrast group at a cut-off score of 60 (1 standard deviation above the mean). If a cut-off score of 70 (2 standard deviations above the mean) is used, the DSF Total Score correctly classified 65% of the clinical group and 99% of the contrast group.

DAP:SPED Total Score

Figure 16

Figure 16 shows the sensitivity and specificity for the DAP:SPED Total Score for the clinical and contrast groups using various cut-off scores. Figure 16 indicates that the total percent of correct classification (true positives and true negatives combined) peaked around a negative one half of a standard deviation. The DAP:SPED Total Score correctly classified 35% of the subjects in the clinical group and 46% of the contrast group at a cut-off score of 55 (1 standard deviation above the mean). If a cut-off score of 65 (1.5 standard deviations above the mean) is used, the DAP:SPED Total Score correctly classified 3% of the clinical group and 95% of the contrast group.
Figure 14

Sensitivity and Specificity for the DSF Physical Symptoms/Fears Scale for Contrast and Clinical Groups Using Various Cut-off Scores
Sensitivity and Specificity for the DSF Total Score for Contrast and Clinical Groups Using Various Cut-off Scores

Figure 15
Figure 16

Sensitivity and Specificity for the DAP:SPED for Contrast and Clinical Groups Using Various Cut-off Scores
Summary of Results

The results presented in this section are summarized according to the research questions presented in Chapter I.

1. Will there be a significant difference between the means scores earned by children in a clinical group and a contrast group on DAP:SPED, DSF, and DSP Total Scores?

Results from the multivariate analysis of variance (MANOVA) indicated that there was a difference between the mean scores earned by the subjects in the clinical group and those of the contrast group. The mean scores of the subjects in the clinical group were significantly (p < .0001) higher than those of the contrast group.

2. What is the degree of discrimination of the DSF 4 Scale Scores, the DSP 6 Factor Scores, and the DSP 3 Composite Scores for a sample of children in a clinical group and a contrast group?
**DSF 4 Scales**

Results from the discriminant function analysis indicated that the two groups (clinical and contrast) were not homogeneous with respect to the DSF 4 Scale Scores. The Depression Scale was the best predictor of the discriminant function (standardized weight = .43). The Interpersonal Problems, Inappropriate Behaviors/Feelings, and Physical Symptoms/Fears Scales were also significant predictors of the discriminant function.

**DSP 6 Factors**

Results from the discriminant function analysis indicated that the two groups (clinical and contrast) were not homogeneous with respect to the DSP 6 Factor Scores. The Acute Problems Factor was the best predictor of the discriminant function (standardized weight = .74). The Autistic Factor was also a significant predictor of the discriminant function. The Conduct, Anxiety, Depression, and Attention Factors did not significantly predict the discriminant function.
DSP 3 Composites

Results from the discriminant function analysis indicated that the two groups (clinical and contrast) were not homogeneous with respect to the DSP 3 Composite Scores. The Psychotic Factor was the best predictor of the discriminant function (standardized weight = .81). The Internalizing Composite approached significance (standardized weight = .40). The Externalizing Composite did not significantly predict the discriminant function. Table 15 provides a summary of the variables which significantly predicted the discriminant function.

Table 15

Significant Predictors of Discriminant Function

<table>
<thead>
<tr>
<th>Measure</th>
<th>Variable</th>
<th>Standardized Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSP 6 Factors</td>
<td>ACP</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>AUT</td>
<td>.40</td>
</tr>
<tr>
<td>DSF 4 Factors</td>
<td>DEP</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>IBF</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>PSF</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>IP</td>
<td>.27</td>
</tr>
<tr>
<td>DSP 3 Composites</td>
<td>PSY</td>
<td>.81</td>
</tr>
</tbody>
</table>

Note. ACP = Acute Problems; AUT = Autistic; DEP = Depression; IBF = Inappropriate Behaviors/Feelings; PSF = Physical Symptoms/Fears; IP = Interpersonal Problems; PSY = Psychotic
3. What are the percentages of true positives and true negatives for the DAP:SPED, DSP, and DSF Total Scores, at various cut-off scores, for a sample of children in clinical and contrast groups?

Results from the analysis of sensitivity and specificity indicated the DAP:SPEDs sensitivity was highest at one standard deviation above the mean (57%). Its specificity was highest at 2 standard deviations and above (100%). The DSP Total Score identified the most amount of subjects in the clinical group (sensitivity) (99%), and the most amount of nonclinical cases (specificity) (92%) at one standard deviation above the mean. The DSF Total Score also identified the most amount of subjects in the clinical and nonclinical groups at one standard deviation above the mean (sensitivity = 88%; specificity = 93%).
a. What are the percentages of true positives and true negatives for the DSP 6 Factor Scores, at various cut-off scores, for a sample of children in a clinical group and a contrast group?

Results from the analysis of sensitivity and specificity indicated that all 6 factors sensitivity was highest at a cut-off score of around 1/2 of a standard deviation above the mean; their specificity was highest at 1 1/2 standard deviations above the mean. The total percent of accuracy peaked at about 3/4 of a standard deviation above the mean for all the factors except Acute Problems. The total p
Results from the analysis of sensitivity and specificity indicated that for all 6 factors sensitivity was highest at a cut-off score of around 1/2 of a standard deviation above the mean; specificity was highest at 1 1/2 standard deviations above the mean. The total percent of accuracy peaked at about 3/4 of a standard deviation above the mean for all 4 scales except Acute Problems.

c. What are the percentages of true positives and true negatives for the DSP 3 Composite Scores, at various cut-off scores, for a sample of children in a clinical group and a contrast group?

Results from the analysis of sensitivity and specificity indicated that all three Composites' sensitivity was highest at 1/2 of a standard deviation above the mean; their specificity was highest at 2 1/2 standard deviations above the mean. The total percent of accuracy peaked around 1 standard deviation above the mean for the Externalizing and Internalizing Composites. The total percent of accuracy peaked around 2 standard deviations above the mean for the Psychotic Composite.
Chapter V
Discussion

A summary of the present study will be presented in this chapter. Specifically, the ability of the DAP:SPED, DSP, and DSF to discriminate between clinical and contrast groups will be addressed. In addition, the DAP:SPED, DSP, and DSF will be discussed in terms of how they relate to current measures of behavioral and emotional problems in children and adolescents. Finally, a discussion of the limitations of the present study as well as directions for future research will conclude this chapter.

Summary of the Study

The present study was designed to investigate the diagnostic validity of the DAP:SPED, DSP, and DSF for a sample of clinic referred children and adolescents through the examination of the aforementioned measures' ability to discriminate between clinical and nonclinical groups. It is unique in that it represents a
comprehensive investigation of the psychometric properties of one recently made available screening procedure for emotional disturbance (DAP:SPED - 1992) and two behavior rating scales, which, at the time of this writing, were in press (DSP and DSF). This study is an attempt to: 1) support the diagnostic validity of the DAP:SPED, DSP, and DSF, and 2) to provide a comprehensive research study on three unique psychological measures of behavioral and emotional problems in children and adolescents, which can serve as a catalyst for future research.

The subjects in the present study were 71 males between the ages 5-12, 41 females between the ages 5-12, 19 males between the ages 13-18, and 22 females between the ages 13-18. These children and adolescents were referred to 1 of 4 Children's Hospital Guidance Centers in central Ohio for psychological assessment by their parents. All subjects were administered the DAP:SPED (Naglieri, McNeish, and Bardos, 1992) DSP (Naglieri, LeBuffe, & Pfeiffer, in press) and DSF (Naglieri, LeBuffe, & Pfeiffer, in press).
The beginning stages of data manipulation involved the transformation of DSP and DSF raw scores into standard scores. This was followed by a summary of the correlations between the DAP:SPED, DSP, and DSF. These correlations indicated that the DAP:SPED correlated low with the DSP and DSF. The DSP and DSF exhibited a moderate to high degree of correlation. In addition, the factors and scales comprising the DSP and DSF respectively, showed a high degree of intercorrelation.

A multivariate analysis (MANOVA) of the DAP:SPED, DSP, and DSF was conducted to determine if the mean scores of the subjects in the clinical group differed from the mean scores of the contrast group. The MANOVA indicated that the mean scores of the subjects in the clinical group were significantly higher than those of the contrast group. In an effort to identify the source of this discrimination between groups, a discriminate function analysis was conducted. This analysis indicated that all four scales (Interpersonal Problems; Inappropriate Behaviors/Feelings; Depression; Physical Symptoms/Fears) accounted for the significant difference between the two groups among the DSF 4 Scales. Two factors (Acute Problems and Autistic) accounted for the significant difference between the two groups among the
DSP 6 Factor Scores. Among the DSP 3 Composites, 1 Composite (Psychotic) accounted for the significant difference between the groups.

**Within Group Differences of DSF 4 Scale Scores for the Clinical Group**

Within group comparisons of the DSF 4 Scale Scores obtained by the subjects in the clinical group showed that their mean score on the Depression Scale was significantly higher than their mean scores on the Interpersonal Problems, Inappropriate Behaviors/Feelings, and Physical Symptoms/Fears Factors.

It is likely that the significantly higher Depression Scale Score is, in part, attributable to the fact that 42% (59 out of 142) of the subjects in the clinical group had an adjustment disorder diagnosis. 4 of the subjects had a diagnosis of adjustment disorder with depressed mood (DSM-III-R: 309.00) and 17 had a diagnosis of adjustment disorder with mixed emotional features (DSM-III-R: 309.40). Depressed mood is part of the mixed emotional features subtype of the adjustment disorder with mixed emotional features diagnosis. According to Sarason & Sarason (1984) "Depression...disrupted sleep patterns, deterioration in
performance at work or school, and social withdrawal are typical problems in adjustment disorders" (p. 122).

**Within Group Differences of DSP 6 Factor Scores for the Clinical Group**

Within group comparisons of the DSP 6 Factor Scores obtained by the subjects in the clinical group showed that their mean scores on the Depression, Autistic, and Acute Problems Factors were significantly higher than their mean scores on the Conduct, Anxiety, and Attention Factors. As with the significantly higher DSF Depression Score, the significantly higher DSP Depression Factor Score is best explained by the percentage of the subjects in the clinical group with an adjustment disorder diagnosis. Consistent with the aforementioned adjustment disorder-depression hypothesis is the significantly higher Acute Problems Factor Score. Many of the items comprising the DSP Acute Problems Factor measure behaviors which exemplify part B(2) of the DSM-III-R diagnostic criteria for adjustment disorder: "...symptoms that are in excess of a normal and expectable reaction to the stressor(s)" (APA, 1987, p.330). Examples of the items comprising the DSP Acute Problem Factor are: "threaten or attempt suicide", "talk about suicide or
death", "engage in compulsive acts or rituals", and "hit, bite, or otherwise injure himself/herself". Why the Autistic Factor is significantly higher is not clear. However, given the subjects' DSM-III-R diagnoses it is likely that the subjects bizarre behaviors (as indicated on the Autistic Scale) are reactive, and not process based (see Davison & Neale, p. 378, 1990).

Within Group Differences of DSP 3 Composite Scores for the Clinical Group

Within group comparisons of the 3 DSP Composite Scores for the subjects in the clinical group showed that their mean scores on the Internalizing and Psychotic Composites were significantly higher than their mean score on the Externalizing Composite. The Internalizing Composite is comprised of the Depression and Anxiety Factors. The elevated Depression Score is undoubtedly contributing to the significantly higher Internalizing Composite. The Psychotic Composite is comprised of the Acute Problems and Autism Factors. Both of these factor scores are elevated; this accounts for the significantly higher Psychotic Composite.
Sensitivity and Specificity.

DSP and DSF

The highest degree of sensitivity was consistently found at one standard deviation above the mean for the DSP Factor, Composite, and Total Scores, and for the DSF Scale and Total Scores. Achenbach & Edelbrock (1986) recommend a cut score of 70 (2 standard deviations above the mean) when interpreting results from the CBCL (Child Behavior Checklist). Results from the present study clearly indicate that 1 standard deviation above the mean is a more sensitive cut-off score for the DSP and DSF, with respect to the present sample, than is a cut-off score of 2 standard deviations. These findings are consistent with those of Naglieri & Gottling (1992). These authors compared the DSF and TRF (Teacher Report Form) with learning disabled and emotionally disturbed students. As Naglieri & Gottling (1992) report, "For both rating scales, the lower the score used to identify an individual, the lower the specificity, and as the scores increased so did the specificity values... Inversely, the higher the cut-off score, the lower the sensitivity rate..." (p. 8).
Speaking about tests which measure the presence of psychiatric disorder, Verhulst & Koot (1992) state that "Such tests usually detect severe cases more readily than mild ones" (p. 49). Based upon the standard scores of the clinical subjects on the DSP and DSF, it is suggested that the subjects in the clinical group represent severe cases of adjustment and attention deficit disorders. It is likely that the severity of these cases accounted for the high sensitivity rating at the 1 standard deviation cut-off score.

**DAP:SPED**

The sensitivity and specificity of the DAP:SPED was lower than that of the DSP and DSF. The DAP:SPED's highest sensitivity rating was at -1 standard deviation below the mean (57%); its highest specificity rating was at 2 1/2 standard deviations above the mean. When using the recommended cut-off score of >55 ("further evaluation indicated") the DAP:SPED's sensitivity was 35% and its specificity 46%. When using the recommended cut-off score of 65-69 ("further evaluation strongly indicated") the DAP:SPED's sensitivity was 3% and its specificity was 95%. While the DAP:SPED's sensitivity and specificity are significantly lower than those of the DSP and DSF, it
is important to keep in mind that the DAP:SPED is a standardized projective measure. The DSP and DSF are behavior rating scales. Historically, behavior rating scales have exhibited adequate psychometric properties, while projective measures have not. The DAP:SPED is the first standardized projective measure to have adequate psychometric properties (see Chapter II). Traditional projective measures, e.g. Rorschach and TAT, have failed to show criterion validity (see Gittelman, 1980). The DAP:SPED has shown criterion validity (see McNeish, 1989; McNeish & Naglieri (in press); Naglieri & Pfeiffer, 1992). Results from the present study further support the criterion validity of the DAP:SPED.

**Limitations of the Present Study**

One limitation of the present study relates to the external validity of the study. Although the randomly generated, normally distributed standard scores allowed for the examination of the psychometric properties, i.e. diagnostic validity, of the DAP:SPED, DSP, and DSF, these results are not specific to any given characteristics of a normal (contrast) control group. Given this, it can be said that the DAP:SPED, DSP, and DSF have good diagnostic validity with respect to a normal distribution of
standard scores. It is likely that some control groups obtained randomly from the population (children and adolescents ages 5-12 and 13-18) will not approximate the normal distribution. In these instances, the results of the present study do not generalize.

A second, and related, limitation is the slightly restricted range of scores generated for the contrast group. On the right side of the randomly generated normal distribution the highest scores were around 2/3 of a standard deviation lower than the highest scores obtained by the subjects in the clinical group. This restriction of range may have enhanced the consistent levels of statistical significance (p < .0001) described in this study. While the direction of the effects are undoubtedly correct (clinical group scoring higher than the contrast group), the magnitude of this direction may be artificially elevated.

A final limitation relates to the grouping of all clinic referred children and adolescents with a DSM-III-R diagnosis into one group (the clinical group). Given the fact that 42% of the subjects in the clinical group had an adjustment disorder diagnosis, and 33% had an attention deficit disorder diagnosis, the results of this
study can only be generalized to children and adolescents who have one of these two diagnoses or subtypes.

**Suggestions for Future Research**

The findings of the present study need to be cross-validated with children from various geographic and demographic regions across the United States. Replication of this study with randomly obtained control group (contrast) subjects matched to the experimental group (clinical) by age, sex, and SES is needed.

Future research needs to include DSM-III-R diagnostic categories other than adjustment disorder and attention deficit disorder. Information on the DAP:SPED, DSP, and DSF profiles obtained by children and adolescents can provide valuable diagnostic information to psychologists.

Finally, research utilizing the present methodological design with other measures of behavior and/or emotional problems in children and adolescents in conjunction with the DAP:SPED, DSP, and DSF is needed. Research of this nature can provide information on how well certain measures discriminate between different diagnostic groups. Information of this nature can assist
the psychologist working with children and adolescents in test battery selection.
References


