# DEVELOPMENT OF THE DISRUPTIVE STUDENT BEHAVIOR SCALE

BY

WILLIAM L. MOSES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

MAY 1986

To Billy

Who would have been proud

ACKNOWLEDGEMENTS

I wish to express appreciation to my parents for their love, understanding, and help; my committee chairperson, Dr. McDavis, for his counsel and encouragement; my committee members, Dr. Ziller for his confidence in my ability to work independently and Dr. Loesch for stepping into the breach and contributing so much so quickly while continuing his friendship and support; and my employers at Pasco-Hernando Community College for their financial support.

Special thanks go to my friend and colleague, Dr. Tom Floyd, who listened for hours and encouraged for years, and to my friends and lovers who were usually supportive, sometimes distracting, and always worth it.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial
Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

# DEVELOPMENT OF THE DISRUPTIVE STUDENT BEHAVIOR SCALE

By

William L. Moses

May 1986

Chairperson: Roderick McDavis, Ph.D.
Major Department: Counselor Education

Disruptive behavior is currently seen by both educators and the public as a major problem in American education. A procedure for quantitatively assessing disruptive behavior in schools is required to show a need for intervention programs and to select students for placement in either special education or alternative education programs. The purpose of this study was to develop and validate an instrument, the Disruptive Student Behavior Scale (DSBS). The DSBS is intended for use in assessing quantitatively the disruptive school behaviors

of middle and junior high students referred for placement in special education and alternative education programs.

This study investigated the position that disruptive school behavior (DSB) can best be described in terms of its type, frequency, and severity. The use of teachers as observers and raters of disruptive school behavior is discussed. Using teacher-generated behavioral statements from disciplinary referrals to better describe DSB is suggested. A review of various rating scale development procedures attempted by business, industry, and government is summarized.

A set of 10 constructs was selected to define DSB. Scale items were developed from referral statements on disciplinary records in a junior high school. A severity factor was incorporated into the scoring system so that behaviors rated as more detrimental to the student were given a higher DSBS rating.

The DSBS was field tested in a public middle school. Students in a norming group and a criterion, or disruptive, group were rated by their classroom teachers using the DSBS. A norm for disruptive behavior for the target school was calculated and a criterion for classifying a student as disruptive was established.

Results indicated the DSBS could identify the criterion group of disruptive students, classify individual

students as disruptive, and exclude non-disruptive students from the disruptive group. A follow-up study suggested the results were consistent over time for all DSBS ratings except those at the lowest end of the scale.

# CHAPTER ONE

# INTRODUCTION

The public school system in the United States has been assigned a major role in socializing and enculturating American youth (Filipczak, 1978). The U.S. Supreme Court in its 1954 landmark civil rights decision (Brown v. Board of Education of Topeka, 74 S.Ct. 686, 691) described education as "a principal instrument in awakening the child to cultural values, in preparing him for later professional training, and in helping him to adjust normally to his environment." The materialistic emphasis of American society and culture ordains that the educational institution at all levels be driven by the broadly defined goal of career success for its graduates (Bell, 1984; DiPrete, 1981, p. 199; National Education Association (NEA), 1975, p. 108).

Unfortunately, a significant number of students are detoured from this goal when educators describe them as displaying behaviors inappropriate to the school environment and not attributable to legally-defined mental or emotional handicaps. Suspensions, expulsions, and assignments to alternative programs are evidence of failure by the educational system to effect students' adherence to

current social norms and culturally-specified behaviors. The consequences to the schools for this failure include loss of both funds and credibility, neither of which the educational system has in sufficient quantity to squander.

Attempts to correct this failure to convey effectively norms and behaviors have included both exceptional child education and alternative schooling programs. The Education for All Handicapped Children Act of 1975 (P.L. 94-142)(Department of Health, Education, and Welfare, 1977) effectively administered the coup de grace to exceptional child education approaches in Florida by failing to include a category appropriate to disruptive behavior (Florida Department of Education, 1975, 1985). Alternative programs frequently fail to provide for selection and discharge criteria, rendering evaluation virtually impossible (Pinellas County School District, 1982). A primary reason for failure to specify behavioral criteria for alternative schooling programs is the lack of appropriate instruments for quantifying disruptive behavior (Salvia & Ysseldyke, 1981, pp. 8, 9).

Inadequacies of existing behavioral assessment instruments include failure to provide for local norming, inclusion of inappropriate items, omission of the severity factor, and inadequacy of prescriptive information (Mesinger, 1982). An instrument providing both a theoretical and a pragmatic rationale for identifying

disruptive students is a requirement for reconsidering the inclusion of this category in special education legislation and enhancing the credibility of alternative education programs (Reeves, Perkins, & Hollon, 1978).

## Statement of the Problem

Disruptive behavior in the public school system is not a new phenomenon (Garibaldi, 1979). That it remains a problem is emphasized by Robert J. Rubel in introducing a collection of papers on crime and violence in public schools:

> The issue in the 1980's no longer centers on whether or not violence in American schools is serious; the issue no longer centers on whether violence is increasing or decreasing; the issue no longer centers on technical anomalies concerning under- or over-reporting of incidents. In the debate of the 1980's, the primary issue before large proportions of our urban schools (and sizeable numbers of our suburban and even rural schools) revolves around the continued viability of American education as it existed a generation ago. (1980, p. 5)

The U. S. government has acknowledged the existence of disruptive behavior by awarding federal grants for alternative education pilot programs (Law Enforcement Assistance Administration, 1979; Moses, 1976).

Included in definitions of disruptive school behavior (DSB) are such varied activities as talking, hitting, yelling (Mayer & Butterworth, 1979); defying rules and procedures (Walker, 1979); aggressive behavior which interrupts the instructional program

(Foley, 1982); and conduct disorders (American Psychiatric Association, 1980 pp. 45-50). Forness and Cantwell (1982) and Forness, Sinclair, and Russell (1984) have identified these categories as likely to be ineligible for special education services under P.L. 94-142.

The U.S. government (Department of Health, Education, and Welfare, 1977), in implementing P.L. 94-142, specifically denied services to the "socially maladjusted." Florida law provides essentially the same restrictions (State Board of Education Rule 6A-6.3016), although Bower (1982), whose research (Bower, 1958) formed the basis for the P.L. 94-142 definition of emotionally disturbed, called this exclusion "contradictory in intent and content with . . . the research from which it came" (1982, p. 60).

The need for alternative education services for disruptive students seems supported by reports of the widespread existence of DSB. Individuals and institutions reporting on the continuing crisis in school discipline include the California Department of Education (1973), the National Education Association (1975), the U.S. Congress (Bayh, 1975; Tygart, 1980), the Michigan Department of Education (Vergon & Williams, 1978), the National Institute of Education (Feldhusen, 1978), Cross and Kohl (1978), Duke (1978), the New York State United Teachers (1979), and the National Education Association

(1980). The Safe School Study Report to Congress (National Institute of Education, 1978) indicated 5,000 teacher assaults per month occurred across the nation. The Gallup Poll on Education (Gallup, 1984) continues to report lack of student discipline as the number one concern of Americans about the public school system.

In Florida, the Governor's Task Force on Disrupted Youth (GTFDY) found 17,983 student-days lost to suspensions over a 2-year period in the 10 school districts studied (GTFDY, 1973, p. 11). An analysis of conduct code violations in Duval County, Florida, schools for 1980-1981 revealed more than 33,000 violations resulting in 13,679 days lost from school (Moses, 1981).

The aversive consequences of chronic DSB for students include lowered self-esteem and functioning level (Caliste, 1979); dropping out and underemployment (Grise, 1980; NEA, 1975; Safer, Heaton, & Parker, 1981); alienation (Garbarino, 1980; Moyer & Motta, 1982); and criminal activity (Edwards, Roundtree, Kent, & Parker, 1981; Mitchell & Rosa, 1981). Likewise, from the perspective of the school system DSB is undesirable, involving excessive teacher attention (Rubel, 1977, Chap. 1), litigation (Lufler, 1982), vandalism costs (Goldstein, Apter, & Harootunian, 1984), teacher stress (Pettegrew & Wolf, 1982), and weakened public support (Amos, 1980). Consequences for the community include criminal actions and

psychiatric referrals (Faretra, 1981). Levin (1972) estimated the expense of inadequate education to be about 6 billion dollars a year (1972 dollars) for costs associated with welfare and crime.

Researchers have identified the middle and junior high school age student as particularly prone to behavior disorder (Geiger & Turiel, 1983; Loeber, 1982; Nielsen & Gerber, 1979; Quay, 1978). These studies suggest the middle and junior high schools as a focus for identifying and remediating disruptive school behavior. Unfortunately, no adequate instruments are available specifically for this population (Mesinger, 1982). Instruments developed from clinical populations contain some items irrelevant to the non-clinical population in the public schools (Quay & Peterson, 1967). Instruments offered with norms developed from research samples and no procedure for developing local norms for disruptive behavior do not consider the placement needs of local school districts (Messick, 1980). Levels of disruptive behavior that can be managed within the regular school environment vary across settings because of differences in such factors as facilities, experience of teachers and administrators, and school board policies.

Current instruments fail to consider the widely differing consequences of specific disruptive acts (Kane & Bernardin, 1982). Some possible effects of this omission

may be to group together students whose behaviors differ widely in their severity, to encourage conceptualizing all disruptive behavior as equally deleterious, and to base placement decisions on personal judgments about the seriousness of a particular type of behavior. Neither does any available instrument provide procedures for creating a prescriptive profile of a student based on the authors' conceptual model of disruptive school behavior (Salvia & Ysseldyke, 1981). This failure may seriously limit the interpretation and application of rating scale results.

## Purpose of the Study

The purpose of this study was to develop and validate an instrument, the Disruptive Student Behavior Scale (DSBS). The DSBS would be used to assess quantitatively the disruptive school behaviors of students referred for placement in either special education or alternative education programs.

## Need for the Study

Salvia and Ysseldyke (1981, pp. 443, 444, 450) have called for norm-referenced instruments to support placement decisions, evaluate student progress, evaluate programs, provide intervention suggestions, and help parents understand their children's abilities in relation to other students. Reeves et al. (1978) called for

reliable instruments to use in placing handicapped children. Also, Camp (1981) notes that

> there is very little current, objective, research-based information in existence to help identify specific student behavior problems occurring in the schools. A need exists for research of this nature to quantitatively establish the actual, current situation with regard to student discipline problems in the public secondary schools. (p. 48)

Presumably, these calls for reliable and valid instruments apply both to special education and alternative schooling programs, as both to some degree remove the student from mainstream classroom activities. However, the Florida law (State Board of Education Rule 6A-6.3017) providing for special education programs for the socially maladjusted was repealed July 24, 1981.

"Educational alternative programs" were created in Florida in 1978 (Florida Statute 230.2315) specifically to reduce disruptive behavior and truancy. Florida Statute 229.565 provides for the evaluation of "procedures for identification and placement of students in educational alternative programs." As an example of practice, in 1982 the alternative education program in the Pinellas County School District did not require quantitative behavioral assessment prior to placement.

Studies, however, have identified problems in using subjective criteria for alternative education placement. Disagreements in ranking behaviors (Pisarra & Giblette,

1981), value systems (Messick, 1980), labels applied to students (Leyser & Abrams, 1982), teaching experience (Rubel, 1977, p.51), level of frustration (Walker & Holland, 1979), race (Arnove & Strout, 1978; Bennett & Harris, 1982; Florida DOE, 1983; Goldsmith, 1982; Mesinger, 1982), sex (Bennett & Harris, 1982), and socioeconomic status (Arnove & Strout, 1978; NEA, 1975) are variables that may confound perceptions of disruptive behavior.

One way to help neutralize these confounding variables is to use quantitative measures. A review of current literature indicates that appropriate instruments may not exist. After a major study of alternative education programs, Mesinger (1982) was unable to recommend even one instrument for use in selecting students.

Messick (1964, 1965, 1980) argued against applying to local environments behavioral norms developed elsewhere. Stott, Marston, and Neill (1975, p. 8), Wodarski and Pedi (1978, p. 480), and Quay and Peterson (1975, 1979) advised the setting of local norms. However, no instrument located in this review provides a specific procedure for determining local norms.

Another advantage of locally-developed norms is the opportunity to compute the mean DSB level for individual schools. Intervention program entry and exit criteria may be defined by the deviation of an individual student's

mean DSB score from the school mean. This may provide the type of quantitative assessment required by state (SBE Rule 6A-6.3016) and federal (P.L. 94-142) law for special education placement and may meet the need noted by Mesinger (1982) for quantitative instruments to assist in selecting students for alternative education programs.

A major need in intervention programs is prescriptive information (Lovitt, 1967 p. 238; Spivack & Swift, 1977). However, many instruments do not provide operationally-defined items which are useful in the classroom. For example, the Behavior Problem Checklist (Quay & Peterson, 1979) items used to identify conduct problem students include "restlessness," "disruptiveness," and "irresponsibility." These items originally were taken from the files of a child guidance clinic (Quay, 1977).

Defining disruptive behavior on the dimensions of type, frequency, and severity has received support from numerous sources (American Psychiatric Association, 1980, p. 45; Bernardin, LaShells, Smith, & Alvares, 1976; Camp, 1980, 1981; Grosek, 1979; Taylor, Warren, & Slocumb, 1979). Criticisms of assessment procedures not incorporating a severity factor have been made by Kane and Bernardin (1982) and Pisarra and Giblette (1981). Nevertheless, no instrument was located which specifically recommended using a severity factor in assessing disruptive school behavior.

An instrument which provides for quantifying DSB may help to protect students from placement in school programs according to inappropriate criteria. To be most effective, the instrument should include provisions for establishing locally-determined placement norms, for comparing with those norms the scores of individual students, for providing prescriptive information, and for systematically considering the type, frequency, and severity of the disruptive behaviors.

## Significance of the Study

This study investigated the theoretical position that disruptive school behavior (DSB) can best be described in terms of its type, frequency, and severity. Theoretical considerations in the use of teachers as observers and raters of disruptive school behavior were discussed. The feasibility of using teacher-generated behavioral statements from disciplinary referrals to better specify the parameters of DSB was suggested. A review of various rating scale development procedures attempted by business, industry, and government were summarized.

The instrument developed by this study will initially be most appropriate as a research tool for conducting studies of DSB. The availability of a process for establishing local norms for DSB may facilitate local research studies in evaluating the effectiveness of

disciplinary measures, in-service training, and alternative education programs. This study will likely suggest additional areas for other investigations.

The identification of disruptive students for interventions is not standardized. This instrument may assist in establishing quantitative criteria for selection, placement, and treatment of disruptive students. This, in turn, may lead to recognition of DSB as a category for exceptional student education funding.

A major premise in much of the literature concerning DSB is the role of school personnel in exacerbating disruptive behavior. It may be that an instrument which provides a behavioral profile of the disruptive student will suggest goals for in-service training programs.

## Definition of Terms

For the purposes of this study, the following definitions apply:

Alternative education program. An educational procedure which provides intervention outside the regular classroom for students exhibiting some predetermined level of disruptive or disinterested school behavior.

Disruptive school behavior (DSB). Behavior that disrupts the learning of self and/or others and is not attributable to severe emotional disturbance or other exceptional education categories.

Delinquent behavior. Behavior by persons under 18 years of age which violates laws and regulations pertaining to them.

Exceptional child (student) education programs. Programs which receive additional funding in order to better serve the needs of students meeting governmental guidelines for special assistance.

Experienced teachers. Full-time, regular classroom teachers who have held that position at least two academic years.

Expulsions. Removal from school for at least the remainder of the school year.

Locally developed norms. Criteria for comparing an individual student's DSB with the expected DSB of a specific reference population in the local school or community.

Maladaptive social behavior. Behavior not of organic origin which would be judged by impartial observers to be inappropriate for the social situation and which ultimately results in aversive consequences for the person exhibiting the behavior.

Method bias. The influence on ratings of the type of rating method used.

Non-quantitative assessment. See Qualitative assessment.

Qualitative assessment. Evaluation based on individual opinion and lacking a systematic basis.

Quantitative assessment. The use of numbers in describing behavior so that a higher number indicates a higher level of the behavior.

Severity. A prediction, stated quantitatively, of the potentially detrimental consequences a disruptive behavior would likely have for a student.

Special education programs. See Exceptional child education programs.

Suspensions. Temporary removal from the regular educational program of a school, usually involving exclusion from school facilities for a specified number of days.

## Organization of the Study

There are four remaining chapters in this dissertation. Chapter Two will present a review of the literature related to the development of an instrument to assess disruptive school behavior (DSB). Specifically, consideration will be given to disruptive behavior in the schools, existing assessment methods, rating scale development, the psychometric properties of rating scales, and the possible uses of results from a disruptive behavior rating scale.

Chapter Three will present the methodology employed in the development, validation, and field testing of the Disruptive Student Behavior Scale (DSBS). Included are

the research questions, information on the population, procedures used in developing the scale, pilot testing, data analyses, and possible limitations of the study.

Chapter Four will present the results of this study, including the data and the information inferred from the data. An explanation of the results will be given and they will be related to past research.

Chapter Five will include conclusions from this study, along with implications for theory, research, practice, and training. A summary of the entire study will be presented, followed by recommendations for additional research.

# CHAPTER TWO

# REVIEW OF LITERATURE

This study requires an investigation of the history and current status of attempts to define disruptive behavior in public schools; identification, assessment, and placement efforts directed toward disruptive students; rating scale development procedures; research into the psychometric properties of rating scales; and the use by schools of results obtained from rating scales. Accordingly, this chapter will review research and opinion covering both theoretical and applied considerations relating to these topics.

## Definition of Disruptive School Behavior (DSB)

According to Camp (1981), the major issue in student discipline in the secondary schools is how to describe quantitatively the kinds of disruptive behavior currently occurring. Summarizing a 1978 survey of state directors of special education, Hirshoren and Heller (1979) reported that while individual states define emotional disturbance consistently, there is considerable variation in the kinds of children so identified. That is, children meeting program criteria in one state appeared to be excluded in

another. Much has been written in an attempt to resolve this situation. A review of the literature suggests the emergence of five discrete perspectives: (a) empirical, (b) clinical, (c) conceptual, (d) educational, and (e) school.

The empirical approach of applying factor analysis (Cattell, 1978; Gorsuch, 1974) to a variety of items has resulted in the identification of some common behaviors associated with disruptive school behavior and has contributed to defining DSB (Achenbach, 1978; Achenbach & Edelbrock, 1978; Edelbrock, 1979; Peterson, 1961; Quay, 1964, 1978; Quay & Peterson, 1967). However, researchers utilizing the empirical approach have included a broad range of behaviors, including many which identify delinquency and personality disorders (Freemont & Wallbrown, 1979), and so the scales developed from these studies have limited application for school personnel in defining the specific category of DSB.

The classification of disorders contained in the Diagnostic and Statistical Manual of Mental Disorders, 3/e. (DSM-III) (American Psychiatric Association, 1980) and research studies incorporating these classifications and descriptions exemplify clinical efforts to define disruptive school behavior. Hewett and Forness (1982) pointed to the necessity of finding a common frame of reference between educational and psychiatric diagnoses in

order for school personnel to accurately interpret clinical reports. Forness and Cantwell (1982) concluded that the respective diagnostic systems of psychiatry and special education remain dissimilar. Likewise, other studies (Loeber, 1982; Werry, Methuen, Fitzpatrick, & Dixon, 1983) failed to find support for the use of psychiatric diagnoses to assign students to special education programs.

The conceptual approach utilizes experience, research, and opinion in formulating descriptions of what is usually referred to in this perspective as "problem behavior" (Jessor & Jessor, 1977, p. 4). Cullinan (1975), Howell (1978), and Richard Jessor (1982) are among those applying a psychosocial conceptualization of problem behavior to the study of adolescent behavior. Nevertheless, while the conceptual perspective gives support to the notion of comparing the behavior of an individual student with the behavior of peers before declaring the student to be deviant, this perspective fails to provide specific criteria for making such a comparison.

The educational perspective includes the definitions contained in federal and state statutes, guidelines proposed by governmental agencies, and district codes of student conduct. In 1977, the U.S. government, without defining the term, specifically excluded the socially maladjusted student from receiving exceptional child

education services under P.L. 94-142. The term "socially maladjusted" is not defined in the latest Florida guidelines for providing special education for exceptional students (Florida DOE, 1985). The U.S. Bureau of Education for the Handicapped has sponsored the compilation of a manual on behavior disorders (Yard, 1977). However, these items are too general for use in a quantitative instrument.

Codes of student conduct contain lists of behaviors for which punishment may be administered. Offenses listed in the codes may be violations of either school rules (e.g., inappropriate display of affection) (Duval County Public Schools, 1980) or of law (e.g., vandalism) (Pinellas County Schools, 1983). While these offenses must be considered in defining disruptive school behavior, they exclude many of the disruptive behaviors frequently occurring within the classroom. Federal, state, and local guidelines seem insufficient for operationally defining DSB specifically enough to be useful in a selection instrument.

The school perspective focuses on the interactions of students, teachers, and administrators within schools. Disruptive school behavior is seen as a product of these interactions. H. M. Walker, author of The Walker Problem Behavior Identification Checklist (1970), described the acting-out child as one who usually defies rules and

ignores classroom procedures, is difficult to manage, avoids failure by attempting little academic work, and alienates teachers and other students by behaving aversively.

Specific behaviors often include hitting, yelling, leaving seat, arguing, having temper tantrums, and provoking others and often lead to confrontations. These confrontations may be verbal, physical, or both. Acting-out behavior may occur in the classroom, in nonclassroom areas, or both. Walker (1970) proposed that acting-out children are differentiated from other students by the frequency, or quantity, of these behaviors, not by the type of behaviors. Thus, a measuring instrument must provide for a frequency component.

Camp (1981) explored the types of behavior considered to be disciplinary problems, the perceived degree of severity of these behaviors, and the frequency with which these behaviors were observed. Camp found that the types of behaviors rated most serious were rarely observed and concluded that the most serious problem may be the frequent, though mild, behaviors that undermine student and teacher morale. A study of 21 secondary school administrators' attitudes toward aggressive behavior suggested that suspensions were awarded according to the administrators' attitudes toward the referred behavior,

rather than according to a consistent standard for the school district (Pisarra & Giblette, 1981).

An evaluation of literature of the school perspective suggests that DSB can be defined in terms which students, teachers, and administrators understand; that the three factors of type, severity, and frequency need to be considered; and that measures of DSB need to be standardized. In this section five perspectives for defining disruptive school behavior were presented. Each perspective offers some assistance in differentiating this category from other behavioral categories. There appears to be support for an instrument which operationally defines types of behaviors occurring throughout the school environment, assigns a quantity to each descriptive item based on the perceived frequency of occurrence and severity, and provides for comparing the score of an individual student to a predetermined norm for that environment.

## Identification, Assessment, and Placement

"Measurement is the construction of a model of some property of the world" (Fraser, 1980, p. 27) and in education this property is often the behavior of a student. One role of the model provided by a measure is to give accurate prescriptive information for planning interventions with students (Forness, 1983). Several studies have suggested this is being performed

inadequately (Greenwood, Walker, & Hops, 1977; Schenck, 1980; Sinclair, 1980; Sinclair & Kheifets, 1982; Spivack & Swift, 1973; Strain, Cooke, & Apolloni, 1976).

Fraser (1980) acknowledged that psychological measurement has been regarded as being quantitatively and qualitatively of a lower order than physical measurement. To achieve improvement, Ysseldyke and Marston (1982) have argued for the use of direct observations of target behaviors by either teachers or trained observers. However, Jones, Reid, and Patterson (1975) found observer reliability varied inversely with the complexity of the behaviors being observed.

Attempts to improve the validity of observations have included such sophisticated approaches as Multidimensional Scaling (MDS) (Torgerson, 1958). Sanson-Fisher and Mulligan (1977), using adolescent student models, found only marginal improvement for this technique over ratings by classroom teachers. A comparison of a computer-driven program for selecting behavioral/emotional disorders with two expert psychologists' selections indicated no meaningful differences existed (McDermott & Hale, 1982). Weinrott (1979) summarized studies that indicated global ratings could be significantly influenced by expectations, while post hoc ratings of the same children by the same raters when recorded on an instrument accurately reflected discrete behavioral events. Gaynor and Gaynor (1976)

argued for instruments written to define behaviors so they may be described quantitatively by teachers.

Beltramini (1982) suggested that scale-item content is more important than other variables in obtaining reliable and valid results. A review by Albaum, Best, and Hawkins (1981) of measurement literature found evidence to support the use of from five to seven categories on Likert-type scales, with no significant losses in reliability, validity, or discrimination when compared with instruments using more intervals. Fewer intervals sometimes resulted in a loss of discriminative power and validity. It appears that teachers using instruments which operationally describe disruptive behaviors can be effective post hoc raters and are able to provide reliable and valid identification of disruptive school behavior (Edelbrock, 1979; Gresham, 1982; O'Leary & Johnson, 1979).

A review of current assessment techniques suggests the emergence of a quantitative/qualitative dichotomy, which will now be explored. In two reviews (Spivack & Swift, 1973, 1977) of instruments for measuring secondary school classroom behaviors no instrument was located which limits its focus to disruptive school behavior, uses only behaviorally-stated items, and provides for calculating local norms. Descriptions follow of representative instruments currently in use.

The Behavior Problem Checklist (BPC) (Quay & Peterson 1967, 1975, 1979) is a 55-item scale of behavioral traits developed from a review of clinical records of kindergarten through eighth grade students referred for psychiatric treatment (Quay, 1977). The items were assigned by factor analysis to four scales plus a grouping suggestive of psychosis. Epstein, Cullinan, and Rosemier (1983, p. 172) and Gresham (1982, p. 137) reported that the BPC is one of the behavior rating scales most widely used in school studies.

The BPC has been used extensively both as a research device (Eaves, 1975; Jacob, Grounds, & Haley, 1982; Kelley, 1981; Touliatos & Lindholm, 1981) and in selecting students for interventions (Algozzine, 1977; Balow, 1979; R. Bower, 1969; Gerard, 1970; Ingram, Gerard, Quay, & Levinson, 1970; McCarthy & Paraskevapoulas, 1969). Jacob et al. (1982) reported that reviews of studies utilizing the BPC suggested reliability and validity issues in need of further study. The inability of the BPC to provide other than broadband classifications has been noted (Achenbach & Edelbrock, 1978).

Comprehensive normative data are not available for the BPC for adolescents (Kelley, 1981). In an investigation of the effects of race on BPC ratings, Eaves (1975) found that white teachers consistently rated black students higher than white students on three of the

subscales. Black teachers showed no such bias. Eaves (1975) concluded this bias could have a major effect on the reported norms for the BPC. Touliatos and Lindholm (1981) found that grade level, sex, and social class had a significant effect on BPC ratings. However, differences between schools and teachers contributed more variance in the BPC ratings than grade, sex, and social class. Touliatos and Lindholm (1981) suggested that Quay and Peterson's (1967) recommendations be followed and individual assessment be based on norms calculated for particular schools and individual teachers.

Spivack and Swift (1973) concluded that the BPC was a reasonably reliable measurement tool. Potential users were cautioned, however, that most items are not specifically observable, but more like labels which imply behaviors and designate traits. Likewise, Stott (1971, p. 232) cited certain BPC items as requiring a teacher to make inferences about students' feelings (e.g., "feelings of inferiority"), being vague or ambiguous (e.g., "oddness, bizarre behavior"), and relating to behaviors unobservable by a teacher (e.g., "stays out late at night," "bed wetting"). This review has identified several areas of the BPC for which additional research has been suggested.

The Behavior Rating Profile (BRP) (Brown & Hammill, 1978) is composed of five rating scales and a sociogram. Three of the scales (60 items) are completed by the target

student, one (30 items) by the teacher, and one (30 items) by parents. The sociogram is a peer nominating technique. The student scales provide self-ratings of behaviors at home, at school, and with peers.

The BRP is based on an ecological approach which, according to the authors, recognizes that students' behaviors are dependent on the settings in which they occur. Its purposes are the identification of students with behavior problems and the differentiations among learning disabled, emotionally disturbed, and behaviorally disordered students in grades 1-12. Each of the six measures is described as independent and individually normed, allowing any scale to be used alone or in conjunction with any of the others.

The BRP manual (Brown & Hammill, 1978) reports internal consistency reliability coefficients exceeding .80. Concurrent validity was investigated by correlating the BRP with measures obtained from other rating scales. Adequate construct and content validity also are reported by the authors. Norms are provided using scale scores with means of 10 and standard deviations of 3, with scores from 7 to 13 considered to be in the normal range.

One study (Reisberg, Fudell, & Hudson, 1982) of behavior disordered students indicated that regular classroom teachers gave higher ratings than special educators ($\bar{X}$=8.85 vs. $\bar{X}$=6.87). Thus, norms may vary

according to the type of respondent (e.g., regular teacher or special education teacher). Also, students' self-ratings were inflated relative to other respondents' ratings. Other investigators have noted problems associated with attempts at multiple and self-ratings.

Lessing and her associates (Lessing & Clarke, 1982; Lessing, Williams, & Gil, 1982; Lessing, Williams, & Revelle, 1981) have reported on their unsuccessful attempts to develop parallel checklists for use by parents, teachers, and clinicians in psychiatric diagnoses. Lobitz and Johnson (1975) found low correlations between parent ratings and observed behaviors. Variables confounding self-ratings include halo effect (Holzbach, 1978), social desirability (Dunnett, Koun, & Barber, 1981; Seidman, Rappaport, Kramer, Linney, Herzberger, & Alden, 1979), and lack of self-knowledge (Beitchman & Raman, 1979).

Ledingham, Younger, Schwartzman, and Bergeron (1982) investigated teacher, peer, and self-ratings of 801 elementary school students. Self-ratings yielded the lowest ratings for deviant behavior, aggression, and withdrawal and the highest ratings for likability. Accuracy of self-evaluation has been found to be positively correlated with high intelligence, high achievement status, and internal locus of control, characteristics not usually associated with DSB (Dunnett et al., 1981).

Reported research using the Behavior Rating Profile is sparse. Additional verification of the assumptions of equivalency of norms within respondent categories and the validity of the self-report scales seems indicated.

The Bristol Social Adjustment Guides, 5/e.(BSAG) (Stott, 1972) consist of 110 behaviorally-stated items from which teachers select those descriptive of a student's behavior in the month prior to the rating. The items were originally developed in 1955 from clinical observations of children aged 6 to 14 and modified by classroom teachers (Stott & Sykes, 1956). A primary goal was to incorporate context into the behavioral descriptions (Stott, 1971).

The BSAG has been used extensively in clinical and research studies (Davis, Butler, & Goldstein, 1972; McDermott, 1980; Stott, 1978; Stott & Wilson, 1977). Reliability and validity data were obtained through extensive research (Stott et al. 1975) but are not reported in a manner that is easily abstracted. Normative data are available only for elementary school populations (Stott, 1972). More recent research (McDermott, 1980, 1981; McDermott & Hale, 1982) has questioned the specificity of the core syndromes of the BSAG and called for further investigation of construct and predictive validities (Hale & Zuckerman, 1981). At present, it

appears that not all of the core syndromes of the BSAG have the specificity required in an instrument to be used in educational placement.

The Hahneman High School Behavior Rating Scale (HHSB) is a 13-factor, 45-item scale published in 1971 (Spivack & Swift, 1971). The HHSB items were developed from observations of actual classroom behaviors, operationally stated in educational terms. The items cover both academic and interpersonal issues and can be rated by teachers in the classroom. The intent is to provide prescriptive information (Spivack & Swift, 1977). The factor scores for each student are found by adding the raw scores for the three or four items comprising each factor. These scores are then combined into a profile, which is used to classify students on the basis of their ability to adapt to total classroom demands.

According to the authors (Spivack & Swift, 1973), validity studies suggest consistent and significant relationships between factor scores and academic grades. No data are available on test-retest or interrater reliability (Spivack & Swift, 1973). Norms are available separately for suburban and urban samples. The HHSB is limited as a selection device for special education programs by lack of reliability data, use of only three or four items per factor, and overlapping among profile descriptions.

The Behavior Evaluation Scale (BES) (McCarney, Leigh, & Cornbleet, 1983) is a 52-item rating scale for use by school personnel. Each item is assigned to a subscale associated with one of the five characteristics of the Bower (1958) definition of behavior disorders used in Public Law 94-142. The BES was developed to aid in diagnosis, placement, and program planning under federal guidelines. Since federal criteria specifically exclude the "socially maladjusted" student, the BES is inappropriate for assessing DSB.

The Portland Problem Behavior Checklist (PPBC) (Waksman & Loveland, 1980) was developed to aid in assessment, evaluation, and intervention planning for school children. The 29 items cover teacher-rated behaviors for grade levels K-12. Norms are not available. Items are very generally stated (e.g., aggressive-physical, destructive) and are rated on a scale of 0 (no problem) to 5 (severe). It is not clear if this is a rating of frequency of behavior or severity of the consequences of the behavior. These features of the PPBC would seem to limit the preciseness and reduce the confidence level of quantitative scores intended to support evaluation and placement for professional services.

The Pupil Classroom Behavior Scale (PCBS) (Dayton, 1967) is a 24-item, teacher-administered rating scale intended to measure the effectiveness of special education

services for students displaying inappropriate classroom behaviors. Most items are behaviorally stated and yield a profile of three factors, achievement orientation, socio-academic creativity, and socio-cooperativeness. Dayton (1967) suggested using the scales for research on groups rather than to describe individual students. Norms are not available. Spivack and Swift (1973) concluded that the PCBS is flawed by having overlapping items in the factors and lacking data to support a relationship beween scale scores and emotional adjustment.

The 36-item Conners Teachers' Rating Scale (CTRS) (Conners, 1969) has been used primarily in clinical diagnosis of children, particularly in the area of hyperactivity (Goyette, Conners, & Ulrich, 1978). It does, however, cover a wide range of school problem behaviors (Roberts, Milich, Loney, & Caputo, 1981). There appears to be a high intercorrelation between the Conduct Problem and Hyperactivity subscales, limiting the usefulness of the CTRS in identifying DSB.

The Brief Behavior Rating Scale (BBRS) (Kahn & Ribner, 1982) was developed from the Devereux series of rating scales (Spivack, Haimes, & Spotts, 1967). A cross-validation study (Kahn & Ribner, 1982) reported that 61% of a socially maladjusted group and 27% of an emotionally handicapped group were correctly identified. These results suggest that additional development is needed

to obtain support for the discriminant validity of the BBRS.

Some of the most complete research in instrument development has been conducted in attempts to improve the diagnosis of clinical populations in the school environment. Although these efforts are not directly comparable to the intent of the present study, six instruments having potential interest to researchers working in the school setting will be summarized.

The Child Behavior Check List (CBCL) (Achenbach, 1978) contains 118 behavior problem items and 20 social competence items. Parallel forms exist for parents and teachers. A review by Achenbach and Edelbrock (1978) of empirical attempts to derive syndromes of child behavior problems concluded with the recommendation that these efforts be linked to the existing mental health system. Recent efforts by these researchers and their associates (Edelbrock & Achenbach, 1980; Reed & Edelbrock, 1983) continue to pursue this objective. At present the applicability of this instrument for educational measurement is limited.

The role of parent observations in describing children's behavior is formalized in the Louisville Behavior Check Lists (Miller, 1967, 1980). A study (Tarte, Vernon, Luke, & Clark, 1982) confirmed the validity of parent observations of clinical symptoms in their children.

The items require inferences and judgments by raters. Eight subscales were created through factor analysis and although several appear to relate to school activities (e.g., hyperactivity, antisocial), the content of individual items comprising the subscales renders them only marginally useful for school assessments.

The Children's Behaviour Questionnaire (Rutter, 1967) was developed for teachers' use in screening for psychiatric assessment large numbers of school children. Many of the 26 items are vaguely stated and some appear to require inferences by the rater. The two subscales are labeled neurotic and antisocial, terms which lack direct application to the school setting.

The Devereux Adolescent Behavior Rating Scale (Spivack et al., 1967) was developed to measure behavior requiring professional intervention. The subscales are oriented to clinical diagnosis and offer little specific information for use in placement decisions.

The Pupil Behavior Inventory: 7-12 Grades (Vinter, Sarri, Vorwaller, & Schafer, 1966) is a 34-item, teacher-administered rating scale intended to furnish information on students referred for agency treatment. Behavioral items were collected from teachers, screened and factor-analyzed, and grouped into five factors. Lack of data on reliability, validity, and norms suggests caution in

using this instrument to select students for special services (Spivack & Swift, 1973).

The Mooney Problem Check List (MPCL) (Mooney, 1942), has been widely used by counselors to identify problems of individuals seeking counseling or to explore the problem profile of a group of students (Sundberg, 1961). However, two studies (Joshi, 1964; Stewart & Deiker, 1976) of the underlying factors of the MPCL scales have identified only a single general factor. The MPCL may be further limited by utilizing items generated from problems mentioned by high school students in 1942.

Several instruments designed for other populations include behaviors often used in descriptions of disruptive school behavior. The Adolescent Behavioral Classification Project instrument (Dreger, 1980) was developed for assessing problems of institutionalized adolescents. An analysis of the first-order factors indicates some commonalities with both the Hahnemann High School Behavior Rating Scale (Spivack & Swift, 1977) and Achenbach and Edelbrock's (1978) syndromes, but many are couched in clinical terms that have little or no relevance to the classroom setting.

Ostrov and associates (Ostrov, Marohn, Offer, Curtiss, & Feczko, 1980) developed and validated the Adolescent Antisocial Behavior Check List (AABCL) for delinquents housed in an institutional treatment setting. The authors

called for modification of the instrument for use in other settings; however, extensive rewriting of items would seem to be required.

The Jesness Inventory (Jesness, 1972) was created to measure attitude change in youthful offenders undergoing treatment. One study (Graham, 1981) found the Jesness Inventory did not have the power to discriminate between non-adjudicated and normal populations and thus would not be useful in a school setting. The Jesness Inventory appears best suited for research (Buros, 1978, pp. 876-878).

The Jesness Behavior Checklist (JBC)(Jesness, 1970) is also a measure of delinquent behavior. The reliability and validity of this instrument have been questioned and the JBC is recommended only for research purposes (Buros, 1978, pp. 873-876).

Non-quantitative assessment often uses nonsystematic observations to provide the information from which judgments will be made. Judgments about individuals are required in all assessment. Inaccurate, biased, or subjective judgments can be misleading and harmful (Salvia & Ysseldyke, 1981). The Russell Sage Foundation Conference Guidelines (Goslin, 1969) and the 1974 Family Educational Rights and Privacy Act (P.L. 93-380--the Buckley amendment) established guidelines for the proper collection, maintenance, and dissemination of data concerning students.

For data to be used in making judgments, it must be verified. For standardized tests, this verification is implicit in the psychometric qualities of the instrument. For observational data, verification requires confirmation by persons other than the original observers (Salvia & Ysseldyke, 1981). When the observation is nonsystematic, verification may be difficult to establish and support and the assessment and resulting evaluation may be open to challenge.

After a classroom teacher nominates a child for evaluation for exceptional child education services, that teacher's observation is verified by required legal procedures (P.L. 94-142). There may be no such procedures for other interventions. The Duval County, Florida, School District has used teacher and principal nominations as the criteria for admittance and dismissal from a program to intervene with students displaying inappropriate social behaviors (Duval County Public Schools, 1980). Short-term suspensions in many school districts do not require hearings and are based solely on a judgment by the school principal (Lines, 1972; Pisarra & Giblette, 1981).

Subjective assessment practices such as these may allow extraneous variables to influence judgments (Poulton, 1976). Four such variables are bias, the influence of observer expectations, inaccurate perceptions, and vagueness of the criteria for intervention.

Pupil characteristics were found by Ysseldyke and Marston (1982) to influence rater bias. Variables contributing to bias include perceived physical attractiveness (Ross & Salvia, 1975); sex, socioeconomic status, and reason for referral (Matusek & Oakland, 1979; Ysseldyke & Algozzine, 1982; Ysseldyke, Algozzine, Regan, & McGue, 1979, 1981); race (Florida Department of Education Report on Public Schools, 1983; Sikes, 1975); type of behavior displayed by the student (Algozzine, 1980); and the theoretical orientation of the observer (Messick, 1980; Salvia & Ysseldyke, 1981).

Erickson (1974) and Shuller and McNamara (1976) found naive observers' reports coincided with experimenter-induced expectancies about problem behavior. After observing decisions made by educators, Weinrott (1979); Ysseldyke, Algozzine, and Richey (1982); and Algozzine and Ysseldyke (1981) speculated that these judgments were influenced by an expectancy factor created by the situation itself. A more direct measure of expectation was reported on by Green and Brydon (1975). They found teachers' attitudes were much more favorable toward middle-income children than low-income children and that 43% of teachers' comments about black children were negative as opposed to 17% of comments about white children.

Dunlap and Dillard (1980) investigated 164 school principals' perceptions of the factors indicative of emotional disturbance in children. The factor least chosen by the principals was the one considered by the researchers most predictive of emotional disturbance.

The vagueness of criteria for suspension in one school district was investigated by Pisarra and Giblette (1981). They found the criterion to be improper conduct, which was not further defined. The researchers concluded that a student reported for fighting would be suspended, possibly suspended, or not suspended depending on the individual administrator who had jurisdiction.

A few of the possible sources of error in nonsystematic observation leading to inaccurate, biased, or subjective judgments have been presented to suggest their ubiquitous nature and the necessity of providing for systematic observations in judgments leading to educational placement decisions.

## Rating Scale Development

Designing a rating scale requires addressing four major issues: (a) what to measure (parameters), (b) how to measure (item content and format), (c) how to record (response format), and (d) how to interpret the results (statistical analysis). Literature pertaining to these issues will be reviewed in this section.

In a frequently cited longitudinal study of deviant behavior, Robins (1966) found the variables of type of behavior, frequency of occurrences, and severity of consequences to be indicators of future behavior patterns. More recent studies supporting these criteria include those of Kohn, Koretzky, and Haft (1979); Camp (1980); Forness and Cantwell (1982); Gresham (1982); Loeber (1982); and a United States Department of Justice report (1982, p. 1).

The types of behavior to be measured by a rating scale are determined by its author(s), who must consider content, sources, format, number, and order of presentation of the items to be included. Halo effects, or the tendency to rate individuals holistically (Thorndike, 1920, p. 25; Willingham & Jones, 1958), were found by Cooper (1981; 1983) to be reduced by having more specific item content. Kreitler and Kreitler (1981) demonstrated that items deemed irrelevant by raters tended to be scored neutrally, thus limiting the derived information. Nevertheless, scales for rating disruptive behavior sometimes include prosocial behavior content (Miller, 1980).

However, Deno (1979) suggested that to observe nondisruptive behavior ignores the purpose of these ratings, i.e., to determine whether inappropriate behaviors are actually excessive. Schriesheim and Hill (1981) mixed positive and negative statements on a questionnaire and

concluded that the effect was to reduce response validity. Many scales do limit their items to behaviors that focus on problem behavior (DiPrete, 1981; Duke, 1978; Governor's Task Force on Disrupted Youth, 1974; Spivack & Swift, 1966; Walker, 1979, p. 55), although not necessarily school problems. Camp (1980) suggested that only school problems directly observable by teachers and/or administrators be included in scales for rating disruptive school behavior.

Logically, items taken from the setting in which the ratings will be made best meet the criteria for relevant content. Smith and Kendall (1963) used this premise in devising Behavioral Expectation Scales (BES). Numerous examples exist of the application of this premise in education (Brown & Hammill, 1978; Camp, 1980; Duval County School Board, 1979; Ross, Lacey, & Parton, 1965; Sherry, 1979; Spivack & Swift, 1977; Stott et al., 1975), mental health (Kaufman, Swan, & Wood, 1979; Kohn et al., 1979; Lachar & Gdowski, 1979; Miller, 1980) and industry (Vance, Kuhnert, & Farr, 1978).

Item format refers to the various forms used in presenting the information to which the rater is asked to respond. It is often related to response format, which refers to the methods of collecting information from the raters. Response format literature will be presented in the section covering the frequency characteristic.

Four types of item formats are currently in use in behavioral rating scales. Behavioral Observation Scales (BOS) describe the target behavior in specific terms that require direct observation at the time the rating is made (Latham & Wexley, 1977). Behaviorally Anchored Rating Scales (BARS) provide a specific description of a behavior for each successive rating point (anchor) of an item and assess cumulative behavior over some time period (Smith & Kendall, 1963). The Mixed Standard Scale (MSS) uses several scales, with three levels of behavioral description for each trait to be measured, and randomizes the order of presentation (Blanz & Ghiselli, 1972).

Summated rating scales (Edwards, 1957), referred to as Likert scales (LT) (1932) or graphic rating scales (Waters, Reardon, & Edwards, 1982), present for each item one statement that may be specific or general. Likert scales have been used with both direct and deferred observation. BOS scales are developed using summated rating procedures (Likert, 1932), while BARS and MSS use the Thurstone (Thurstone & Chave, 1929) scale development process (Bruvold, 1969).

Conflicting conclusions have resulted from numerous investigations into the advantages and disadvantages of these scale formats. Fay and Latham (1982) found BOS to be superior to BARS in rating video-taped behavior during job interviews. However, Murphy, Martin, and Garcia

(1982) questioned the theoretical basis for BOS and found evidence to suggest that BOS tapped recall for behavior traits as well as immediate observation. Several studies (Hom, DeNisi, Kinicki, & Bannister, 1982; Ivancevich, 1980; Keaveny & McGann, 1975; Lee, Malone, & Greco, 1981) failed to find significant advantages for the BARS format over summated rating scales or other alternative methods (Jacobs, Kafry, & Zedeck, 1980; Kingstrom & Bass, 1981; Schwab, Heneman, & DeCotiis, 1975).

In opposition to MSS theory, Finley, Osburn, Dubin, and Jeanneret (1977) found evidence to suggest that an obvious scale format may be superior to a hidden continuum. Dickinson and Zellinger (1980) compared MSS, BARS, and LT formats and found MSS produced less method bias, BARS produced as much discriminant validity as MSS and provided the best feedback to ratees, and LT scales were easiest to understand and use. When Bruvold (1969) tested the application of summated scales (Likert, 1932) and successive interval scales (Edwards & Thurstone, 1952) to the same data set, no significant differences were found between the two scaling methods. According to Bernardin and Smith (1981), one explanation may be that scale constructors have deviated from the original procedures (Smith & Kendall, 1963) in developing BARS instruments.

In addition to the Thurstone and Likert scaling procedures, a third method is available. According to

Edwards (1957, p. 172), a Guttman (1944, 1945, 1947a, 1947b), or cumulative scale, requires that the construct to be measured be unidimensional. Since disruptive school behavior consists of many discrete behaviors, a Guttman scale is not suitable for the instrument developed in this study. At present, it appears that no item format is superior enough to warrant relinquishing the clarity of understanding and ease of use (Dickinson & Zellinger, 1980) of the Likert scale, which presents one descriptive item at a time to which the rater assigns a quantitative value from a given range of values.

In determining the number of items to include in a rating scale, some researchers (Quay & Peterson, 1967, 1979; Spivack & Swift, 1971; Stott, 1972) have relied on factor analysis, using an arbitrarily chosen factor score as the cut-off score. Edwards (1957) suggested an intuitive approach, utilizing 20-25 items that discriminate between the groups at the extremes of the scale. A comprehensive study (Achenbach & Edelbrock, 1978) of 18 rating scales found the range of items to be from 36 to 287 (median = 68 items; mean = 90.4 items). Of the 6 scales intended for use by teachers, 4 contained fewer than 50 items and 2 between 50 and 100 items.

In a study of preferred scale length, Meredith (1981) found half of the respondents preferred from 20 to 40 items, with 25 the median preferred length. In another

study, Meredith (1975) found a 52-item scale was judged too long. Seidman and his associates (Seidman et al., 1979) concluded their 46-item Teacher Behavior Description Form was too cumbersome and reduced it to 23 items. While item complexity is probably a factor (Meredith, 1981), this review suggests a scale using no more than 40 items would probably be acceptable to most teachers.

The ordering of items within a scale has been suggested as a possible source of leniency error, halo effect, and impaired discriminant validity (Blanz & Ghiselli, 1972). Schriesheim and DeNisi (1980) and Schriesheim (1981b) found that grouping according to constructs rather than randomizing questionnaire items resulted in impaired discriminant validity. Increased leniency response bias was also found when items were grouped (Schriesheim, 1981a).

Dickinson and Zellinger (1980) concluded that a randomized scale contributed as much discriminant validity as an ordered scale while displaying less method bias. In a comparison of randomized and grouped scales, the randomized scale engendered as much convergent and discriminant validity (Waters et al., 1982). Thus, a randomized order of presentation seems indicated.

Obtaining a meaningful measure of the frequency of target behaviors requires attention to the variables of response format, length of the observation period, and

type and number of raters. According to Tzeng (1983), four response formats are most frequently cited in the literature. They can be differentiated in terms of two psychometric criteria. First, the existence of a neutral response option defines the free choice format. Absence of a neutral rating option defines the forced choice format. Second, categorical (qualitative) ratings answer the question "Does the ratee fit this category?" while discriminatory (quantitative) ratings answer the question "To what degree does the ratee fit?"

Tzeng (1983) criticized forced choice measures for their omission of a valid response category, i.e., uncertainty or neutrality of the raters' perceptions. King, Hunter, and Schmidt (1980) concluded that a forced choice format was ineffective in reducing rater halo. Dunnette (1963, p. 96) reported that rater resistance to forced choice formats led to their abandonment.

Categorical, or qualitative, formats used in checklists cannot detect relative differences in degree between two behaviors performed by the same ratee or between the same behaviors among ratees (Tzeng, 1983). Johnson, Smith, and Tucker (1982) found less response skewness on a 5-point Likert discriminatory scale compared to a yes/?/no categorical format. A zero-based discriminatory, free choice response format seems most appropriate (Likert, 1932). The absence of a behavior can be indicated

by the 0 position or, if present, the perceived frequency can be indicated by choosing a value from the remainder of the scale (Edwards, 1957).

The number of value choices permitted to the rater is a critical issue. If few points are used some information may be lost, but the scales are less ambiguous for the rater. If there are too many points the discriminations may be too fine for the rater to make. Albaum et al. (1981) attempted to show superiority for a continuous scale format, but concluded that equivalent aggregate measurements were obtained from a 5-category, discrete rating scale.

Likewise, Bernardin et al. (1976) and Bardo and Yaeger (1982) failed to find continuous scales superior to discrete scales. The superiority of a 5-point, discrete rating scale has been suggested by Cowen, Dorr, Clarfield, Kreling, McWilliams, Pokracki, Pratt, Terrell, and Wilson (1973); Lissitz and Green (1975); McKelvie (1978); Neumann and Neumann (1981); and Broadbent, Cooper, Fitzgerald, and Parkes (1982).

Conversely, Bardo and his associates (Bardo & Yeager, 1982; Bardo, Yeager, & Klingsporn, 1982) found obtained means and variances closer to the expected values for 4-point scales over 5- and 7-point scales. These results appear contrary to most other studies. Edwards (1957, pp.

150-151) gives Likert's original statistical rationale for the use of a 5-point scale, anchored with the integers 0 through 4, and the summation of scores for individual items as a total score for each ratee. Current research provides no compelling evidence for departing from this original format.

An anchor, e.g., "always," "sometimes," "never," is usually associated with each scale point of a Likert-type summated rating scale (Pohl, 1981). While a variety of anchors has been used, the basis for the selection is often not stated (Beatty, Schneier, & Beatty, 1977; Broadbent et al., 1982; Camp, 1980; Cowen et al., 1973; Hunter, Hunter, & Lopis, 1979; Kassin & Wrightsman, 1983; Moses, 1974; Siegel, Dragovich, & Marholin, 1976; Solomon & Kendall, 1977; White, 1977).

Several studies have investigated the assumptions involved in the selection of one popular set of anchors: always, often, occasionally, seldom, and never. Parducci (1968), Chase (1969), and Pepper and Prytulak (1974) concluded that the meanings of anchor words were influenced by context. The effects of individual differences among raters on their interpretations of anchor words were demonstrated by Helson (1969) and Goocher (1965). These studies suggested that the above anchors may not define perceptually equal intervals along the rating continuum.

Four studies (Bass, Cascio, & O'Conner, 1974; Schriesheim & Shriesheim, 1974, 1978; Spector, 1976) have sought to select five anchor words that would be perceived by raters as defining equally spaced rating intervals. However, the most definitive study appears to be Pohl's (1981) partial replication of the Bass et al. (1974) and Shriesheim and Shriesheim (1974, 1978) studies. Using responses from 164 college students, Pohl (1981) calculated the means and standard deviations for 39 expressions of frequency.

Comparing these with the theoretical mean responses for a 5-point equal interval scale, Pohl (1981) derived the response set of always, quite often, sometimes, very infrequently, and none of the time. The calculated mean (26.71) for the mid-point term "sometimes" differed significantly ($p < .001$) from the theoretical mean (29.05), but nevertheless was the value closest to the optimal for a 5-point scale. The other calculated values were not significantly different from the theoretical profile. Thus, with the exception of the mid-point term, it appears that the anchors produced by the Pohl (1981) study adequately defined equal-appearing intervals on a 5-point rating scale.

The length of the period for which behaviors are to be rated has been little studied. For instance, the manual for the Behavior Problem Checklist (Quay &

Peterson, 1967, 1975, 1979) does not specify for the rater the inclusive time period to be considered in rating the listed behaviors. The authors of the Devereux Elementary School Behavior Rating Scales (Spivack & Swift, 1966) instructed their raters to "consider recent and current behavior" (p. 75). The same authors (Spivack & Swift, 1977), in developing the Hahnemann High School Behavior Rating Scale, instructed teachers to base ratings on behavior observed "over the past month" (p. 300).

A study (Hinton, Webster, & O'Neill, 1978) of hospitalized clinical patients used a 6-week time period. An investigation (Beatty et al., 1977) of performance rating in a data processing firm utilized three assessment periods of two months each for a total of six months. In a study of several response formats, Broadbent et al. (1982) used a 6-month inclusive time period. However, in none of these studies was a rationale given for selection of the time period.

Two attempts at aggregating measures over specific time periods have provided more precise instructions to the rater. Cowen et al. (1973) defined each of five rating points in terms of the inclusive time periods to be considered when aggregating occurrences of behavior. For example, the fourth anchor point, often, was defined as "you have seen this behavior more often than once a week

but less often than daily" (p. 16). Camp (1980) used the following scale:

Frequency of occurrence

0 - Never observed
1 - Once or more in semester
2 - Once or more monthly
3 - Once or more weekly
4 - Once or more daily (p. 11)

The work of Seymour Epstein (1980), in support of the stability over time of personality traits, bears directly on the issue of aggregating behavior ratings over some time period. Epstein (1980) stated that "stability can be demonstrated . . . as long as the behavior in question is averaged over a sufficient number of occurrences" (p. 791). In testing this hypothesis, Epstein conducted four studies in which he used, among other types, ratings performed in classrooms by teachers. Epstein suggested aggregating behavior over subjects, stimulus situations, time, and modes of measurement in order to establish predictive reliability and validity (p. 797).

Ratings of middle and junior high school students by their teachers in different courses would meet the conditions of subjects and situations. Epstein (1980) suggested that ratings at a single time following multiple or extended observations represent an intuitive averaging that has the "potential for producing highly replicable and valid results" (p. 802). Harrop (1979) also challenged the common assumptions (Fay & Latham, 1982;

Latham, Fay, & Saari, 1979) that coding of directly observed behaviors produced superior results to aggregating behaviors over time.

A related concern in the assessment of school-related behavior is selection of the time of year in which the ratings will be made. Several studies (Cowen et al., 1973; Epstein, et al., 1983; Larrivee & Bourque, 1980) recommend allowing student behavior and teacher perceptions to stabilize. Supporting these decisions are data from the Texas Junior High School Study (Evertson, Anderson, & Brophy, 1979).

Evertson and Veldman (1981) found a moderate but steady increase in serious misbehavior over the course of the school year and an increase in general misbehavior in April. Evertson and Veldman (1981) concluded that short-term studies should avoid ratings made either early or late in the school year. The available literature seems to suggest the feasibility of aggregating behaviors over time periods specified in the rating scale instructions and after teachers have had at least two months to observe student behavior.

Deciding on the most appropriate type of rater to use in assessing children's behavior has long been a problem. In 1965, Ross et al. recognized the potential usefulness of teacher ratings. Teacher's ratings have been found to be more accurate than peer ratings of classroom behaviors

(Bailey, Bender, & Montgomery, 1983); other school professionals' ratings (Bower & Lambert, 1971, p. 143; Freemont & Wallbrown, 1979), and institutional child care workers' ratings (Kohn et al., 1979) and to be equivalent to the ratings obtained by a multidimensional scaling technique (MDS) applied to classroom behavior (Sanson-Fisher & Mulligan, 1977).

A number of researchers have found support for teacher ratings as appropriate measures of general classroom behaviors (Solomon & Kendall, 1977), social behavior (Loranger, Lacroix, & Kaley, 1982), assertive vs. aggressive behavior (Roberts & Jenkins, 1982), acting out behavior (Walker, 1970), and behavior that would likely result in referrals for exceptional child education (Dean, 1980; Epstein et al., 1983; Horne & Larrivee, 1979; Lahey, Green, & Forehand, 1980; McKinney & Forman, 1982; Roberts et al., 1981).

Not all studies have yielded positive results. Morris and Arrant (1978) found that regular classroom teachers tended to see more behavior problems in students referred for evaluation than did school psychologists. A study (Kazdin, Esveldt-Dawson, & Loar, 1983) of psychiatric inpatient children found extra-class raters' evaluations of overt classroom behaviors to correspond more closely to direct observational data than did teachers' ratings. However, teachers were more accurate than the extra-class

raters in identifying hyperactive children using a behavior checklist. Overall, the evidence suggests strong support for the use of teachers as raters of classroom behavior.

An associated issue is the use of multiple raters to increase reliability and reduce halo effect (Epstein, 1980). Ratings of students commonly are obtained from all teachers having direct classroom contact (Linton & Chavez, 1979; Wixson, 1980). This procedure could result in as few as one or perhaps as many as seven ratings, depending on the grade level and local practice.

More recent research efforts have focused on empirically determining the most effective number of raters. Prinz and Kent (1978) increased from 1 to 4 the number of raters of parent-adolescent interactions in a clinical setting and reported increased reliabilities. Both reliability and concurrent validity of clinical judgments were shown to increase when the number of judges was increased from one to ten (Horowitz, Inouye, & Siegelman, 1979). Strahan (1980) extended the Horowitz et al. (1979) study and concluded that after using four raters, adding additional ones contributed little to measurement effectiveness. Another study (Green, Bigelow, O'Brien, Stahl, & Wyatt, 1977) of inpatient clinical behaviors found little improvement when using more than four raters.

Although in general agreement with the above studies, a cautionary note was added by Kenny and Berman (1980), who pointed out that if raters are completely unreliable, increasing their numbers will not increase reliability. The number of teachers usually available in a middle or junior high school to serve as raters would appear to be adequate to contribute to both improved reliability and concurrent validity.

Various classifications of severity have been adopted in school settings. Student conduct codes typically use some method of indicating seriousness of offenses, such as "serious misconduct" (Pinellas County Schools, 1983, p. 7) and "minor, intermediate, and major" (Duval County Public Schools, 1980, p. 16). Researchers (Pisarra & Giblette, 1981) have used categories emphasizing the targets of the behavior (e.g., offenses against persons, offenses against state laws). Teachers often focus on specific behaviors (e.g., use of drugs, striking teacher) (Camp, 1981) and administrators have used a combination of both (National School Public Relations Association, 1973).

There is little consensus on the number of levels to be used in assigning degrees of severity. Taylor et al. (1979) used levels ranging from 1 (not very severe) to 4 (extremely severe). Camp (1980) used 0 for "not concerned" through 4 for "extremely concerned." In an earlier study, Moses (1974) used three levels, 1 (mild), 2

(moderate), and 3 (severe) in asking mental health and criminal justice professionals to rate a list of problem behaviors. To use too many levels may imply a degree of confidence in discrimination not supported by the subjective nature of such ratings.

Not all rating scale authors and researchers accept the necessity for including a severity rating (Searls, Isett, & Bowders, 1981; Spivack & Swift, 1977). Even when, as in the Behavior Problem Checklist (Quay & Peterson, 1967, 1975, 1979), a severity factor is provided for, the author does not always recommend its use. However, at the practitioners' level the degree of severity of behaviors is a major concern.

Algozzine (1979), using items characteristic of several behavior rating scales, developed the Disturbing Behavior Checklist which asks teachers to rate the degree of disturbance they experience as a result of different student behaviors. This suggests a consequence to the teacher based not on the frequency of the behavior, but on the type and severity. After noting irregularities and lower reliabilities, Taylor et al. (1979) had teachers rate for severity 26 items of Part Two of the Adaptive Behavior Scale (ABS) (Nihira, Foster, Shellhaas, & Leland, 1969). Teachers were able both to categorize behaviors and rate them in terms of severity, leading Taylor et al. (1979) to conclude that this additional information would

be useful in refining the scale and adding to its clinical efficacy.

Inasmuch as the instrument developed in this study is intended to have locally developed norms, the statistical techniques used in the norming procedure and the comparing of individual scores to the derived local norms are not complex. While some more recent studies have focused on problems associated with such common procedures as the calculation of measures of central tendency (Mosteller & Tukey, 1977; Stavig, 1978, 1982), many researchers continue to rely on descriptive statistics utilizing raw scores, arithmetic means, standard deviations, and standard scores.

White (1977) compared individual student's scores on classroom behavior to the computed mean score for five classes of "Follow Through" program students in order to identify immature students. In a business setting, Fay and Latham (1982) used means and standard deviations in comparing scores obtained using two different rating methods. A study (Lyness & Cornelius, 1982) comparing judgment strategies and ratings of college instructors supported the use of a rating scale composed of discrete items, with an overall rating calculated by weighting the items and summing the weighted scores. To obtain mean sub-scores for subjects, Algozzine (1980) summed scores across the items defining each of four factors of

disturbing behaviors and used means and standard deviations in analyzing the results.

The cited studies seem to support the use of descriptive statistics in both obtaining individual scores (i.e., sum of weighted ratings) and deriving a local norm (i.e., mean) from ratings of a representative sample of a total population. Salvia and Ysseldyke (1981, chap. 4) offer definitions of common terms for descriptive statistics applied to assessment.

## Psychometric Properties of Rating Scales

Historically, rating techniques have aroused controversy over estimations of validity and reliability (Ryan, 1958). Validity is the relevance of the scale to the variables being measured. Most sources recognize three types of validity, i.e., content, criterion-related or concurrent, and construct (American Psychological Association, 1966; Cronbach, 1970; Kerlinger, 1972). Reliability is the accuracy or precision of a measuring instrument and has been usually classified as either temporal, inter-rater, or internal (Cronbach, 1970).

However, investigations (Epstein, 1980) into the effects of situations on behavior have recently introduced a fourth consideration, situational reliability, or the consistency of behavior across settings. The development of norms against which to compare results obtained from

individual administrations of rating scales is another area of active investigation (Mendelsohn & Erdwins, 1978; Messick, 1980). Research on these issues is reviewed in this section.

Content validity refers to the relevance and representativeness of the items used in construction of a scale (Epstein, 1980). Often, this is determined by obtaining judgments from experts not otherwise involved in the scale construction (DiStefano, Pryer, & Erffmeyer, 1983; Jones et al., 1975, p. 83; Lawshe, 1975; Thorne, 1978). Kreitler and Kreitler (1981) found that item content determined the rater's perception of the central theme of an instrument. Items not perceived as relevant to the central theme tended to be given neutral responses, thus limiting the information contributed by the rater.

Criterion-related validity is studied by comparing scores obtained from an instrument with one or more external criteria of the variable being measured (Kerlinger, 1972, p. 459). Criterion-related validity encompasses both concurrent and predictive qualities (Epstein, 1980). The comparison of scale results with an independent judgment or diagnosis of a subject is an example of an attempt at estimating criterion-related validity. If the judgment or diagnosis confirms the scale indications, the inference may be drawn that the scale is in agreement with the concurrent diagnosis and is

predictive that others given a similar rating would also be diagnosed similarly (Kohn et al., 1979; Mendelsohn & Erdwins, 1978).

In one validation study, Harris, Kreil, and Orpet (1977) used the school principal, guidance counselor, and two teachers as judges in selecting both disruptive and prosocial students for rating by the Behavior Coding System (Patterson, Ray, Shaw, & Cobb, 1969). In developing the Pittsburgh Adjustment Survey Scales (Ross et al., 1965), school principals were used to nominate adjusted, withdrawn, and aggressive students for rating by their teachers and scale results were compared with these nominations.

According to Kerlinger (1972, p. 461) and Cronbach (1970), the significance of construct validity is its concern with the theory behind the variable being measured. Guion (1977) argues that construct validity integrates both content and criterion considerations. Likewise, the usefulness of content and concurrent validity is questioned by Sanson-Fisher and Mulligan (1977) and construct validity is supported.

A definition of construct validity as the process of ascribing meaning to scores is offered by Stenner and Smith (1982). Messick (1980) broadens the concept of validity to include both test interpretation and test use. Messick (1980) describes construct validity as

"interpretive meaningfulness" (p. 1015) and suggests that it rests on four bases: convergent and discriminant validity, ethical interpretation, relevance and utility for the specific application, and the consequences following use of the instrument.

To be interpretable, a rating scale must be reliable. That is, a scale must produce similar results when applied to the same person over several administrations, the instrument must be relatively free of errors of measurement, and the results must closely approximate the "true" value of the variable for the person being rated (Cronbach, 1970; Kerlinger, 1972).

Typically, test-retest data are compiled for varying time periods between administrations. The correlation between the two obtained scores is used to justify estimations of temporal stability and, in the case of rating scales, intra-rater reliability. Examples of reported test-retest intervals include one week (Duval County School Board, 1979; Quay, 1977), two weeks (Mendelsohn & Erdwins, 1978; Russell, Lankford, & Grinnell, 1981) and two years (Quay, 1977). However, Masterson (1968) pointed out that low test-retest correlation coefficients may reflect the transitory nature of the measured variable and suggested high coefficients of internal consistency may be more indicative of reliability for some instruments.

Internal consistency has often been estimated by inter-item and item-total analysis (Edwards, 1957; Kerlinger, 1972). In these procedures, an individual's rating on one item is compared with the rating on all other items or with the total score from the scale or subscale to estimate the degree to which each item is similar to the other items. Item analysis may be important in reducing errors of measurement attributable to the composition of the instrument (Benson & Clark, 1982). However, internal consistency may not provide good reliability estimation for a rating scale assessing constructs comprised of many discrete behaviors (Kerlinger, 1972).

Some research (Rosenthal & Jacobson, 1968; Sulzbacher, 1973) into observer bias has suggested that beliefs about ratees may affect rater perceptions and, consequently, the reliability of the ratings. In three studies (O'Leary & Kent, 1973; Shuller & McNamara, 1976; Siegel et al., 1976) of disruptive classroom behavior, while biasing information experimentally introduced was found to influence global ratings, it had no significant effect upon results obtained from behaviorally stated scales. Siegel et al. (1976) suggested that behaviorally specific items reduce bias and improve inter-rater and intra-rater reliability.

The degrees of agreement among different raters on measures of the same subjects at the same time in the same setting have been used to indicate the inter-rater

reliability of an instrument (Cronbach, 1970). Also, the agreement among different raters of subjects in the same settings at different times has been used for the same purpose (Cronbach, 1970). In middle and junior high schools, these conditions do not usually occur naturally. Fortunately, investigations of trait consistency in subjects (Abikoff, Gittelman, & Klein, 1980; Epstein, 1980; Mischel, 1969) have encouraged the comparisons of ratings by different raters over the same elapsed time periods, but for different settings and situations, conditions which do occur naturally in the secondary school setting.

Epstein (1980) concluded that subjects do manifest trait consistency, if aggregation techniques are applied in assessing behaviors. Epstein (1980) suggested aggregation over raters (e.g., teachers), situations (e.g., classrooms), occasions (e.g., class periods), and measures (e.g., disciplinary records). Epstein further suggested that when single ratings are made after extended periods of observation, these ratings are similar to aggregated ratings in that they represent an intuitive averaging of ratings over many observations. Thus, reliability may be improved by combining different teachers' ratings of the same student over the same portion of the school year.

According to Cooper (1981), perhaps the most ubiquitous challenge to inter-rater reliability is halo error

(Thorndike, 1920) or the tendency of a rater to allow overall impressions of an individual to influence judgment of specific areas of behavior (Holzbach, 1978). Attempts (Landy, Vance, Barnes-Farrell, & Steele, 1980; Landy, Vance, & Barnes-Farrell, 1982) to statistically control for halo effects have apparently not succeeded (Harvey, 1982; Hulin, 1982; Mossholder & Giles, 1983; Murphy, 1982). One exploration of ways to reduce halo error resulted in a restatement of classic advice: do not use rating categories that are imprecise and overlapping (Cooper, 1983). In an extensive review of the literature, Cooper (1981) concluded that of nine methods currently employed to reduce halo effect, all leave residual illusory halo.

Studies of variables affecting reliability have identified several other challenges to the accuracy of school behavior ratings. The sex of the teacher was found in two studies (Levine, 1977; Silvern, 1978) to be correlated with ratings of classroom behavior, with male teachers consistently reporting lower levels of disruptive behavior. Teachers' ratings seemed to be influenced by special education labels in one study (Fogel & Nelson, 1983). In two studies (Marwit, 1982; Marwit, Marwit, & Walker, 1978), perceived unattractiveness of students has been shown to correlate with higher ratings of disruptive behavior.

While challenges to reliability from a variety of sources have been observed, several studies (Bernardin &

Pence, 1980; Fay & Latham, 1982; Latham, Wexley, & Pursell, 1975; Madle, Neisworth, & Kurtz, 1980; Pursell, Dossett, & Latham, 1980) have suggested that training in the use of rating scales may be effective in reducing errors of measurement. This review of studies of validity and reliability has identified some sources of and countermeasures for errors of measurement. Next, studies of the variables affecting the norming of rating scales will be reviewed.

Several writers have shown concern for the relationship between behavior and the context in which it occurs. The social value of a test, according to Messick (1980), is determined by its instrumental value for a particular setting. Willems (1975) stated that few phenomena have meaning independent of the context in which they occur. Likewise, researchers were cautioned by Dickinson (1978) to evaluate behavior only in an environmental context.

Epstein (1980) referred to the "extreme situational specificity of behavior" (p. 794) and warned that experiments conducted in a single situation cannot be relied on to generalize across even minor variations in stimulus conditions. Others supporting this psychosocial approach include Sherif (1954); Erickson (1963), quoted in Tinto, Paclilio, and Cullen (1978); Salvia and Ysseldyke (1981, p. 378); and Zammuto, London, and Rowland (1982).

Schools were described by Garbarino (1980) as "contexts for behavior and development" (p. 19). Some of the characteristics of schools which may influence levels and interpretations of disruptive behavior are size of enrollment (DiPrete, 1981, p. 86; Garbarino, 1980; Kowalski, Adams, & Gundlach, 1983); public or private administration (DiPrete, 1981, p. 81); control orientation (e.g., humanistic vs. custodial) (Deibert & Hoy, 1977; Gaynor & Gaynor, 1976); degree of person-environment fit (Kulka, Klingel, & Mann, 1980); traditional vs. open classrooms (Solomon & Kendall, 1975); length of faculty tenure (DiPrete, 1981, p. 107); socioeconomic level of the host community (Kowalski et al., 1983); and region of the country (DiPrete, 1981, p. xx; Kowalski et al., 1983). Researchers advocating the use of local norms for behavioral measurements include Fremont and Wallbrown (1979); Mendelsohn and Erdwins (1978); Quay and Peterson (1967); Smith (1976); Walker and Hops (1976); and Wallbrown, Wallbrown, and Blaha (1976).

The effects of sex, age, race, and socioeconomic status on ratings of disruptive behavior have been frequently studied. The types of disruptive behavior displayed in both educational and clinical settings have not been found to be significantly different for the variables of sex (Behar & Stewart, 1984; Epstein et al., 1983; Morris & Arrant, 1978; Stott et al., 1975, p. 166),

age (Behar & Stewart, 1984; Ghodsian, Fogelman, Lambert, & Tibbenham, 1980; Stott et al., 1975, p. 83), race (Gajar & Hale, 1982), or socioeconomic status (Behar & Stewart, 1984; Stott et al., 1975, p. 97). Thus, providing for separate norms for these variables seems unnecessary in any scale rating only disruptive behaviors.

## Uses of Behavior Rating Scales

Bailey et al. (1983) supported the use of rating scales in program planning and evaluation. Likewise, the lack of effective measurement devices was seen by Hirshoren and Heller (1979) as limiting the evaluation of program effectiveness. Mesinger (1982) called for the use of appropriate measurement devices in providing services for deviant youth within the public school setting. Cooper (1983), Peed and Pinsker (1978), and Beatty et al. (1977) have suggested providing rating scale results to ratees to influence behavior changes. Using rating scales to provide a standardized description of behavioral problems has been suggested (Edelbrock & Achenbach 1978).

In a study comparing resource room delivery models, Wixson (1980) used a behavior rating scale in developing and evaluating intervention programs for various categories of handicapped children. Morton Bortner (Buros, 1978, p. 493), reviewing the AAMD Adaptive Behavior Scale, pointed out its usefulness for evaluating the progress of

individuals and evaluating program goals. The Duval County School Board (1979) used a locally constructed behavior checklist to evaluate their grant-funded program for disruptive students.

Several programs which retained students in their regular classrooms have used behavior scales for evaluation purposes. Walker and Holland (1979) and Linton and Chavez (1979) developed and used rating scales for this purpose in elementary and junior high schools, respectively. The Hahnemann High School Behavior Rating Scale (Spivack & Swift, 1977) was intended to provide teachers with a practical means of describing disruptive classroom behavior to parents and other school personnel. In a study of junior high school truants, Nielsen and Gerber (1979) used a behavior rating scale to match school interventions with student needs.

A quantitative measure of disruptive behavior was developed by Mendelsohn and Erdwins (1978) to assist community agencies in devising programs for expelled students. Haskell (1979) developed a method of quantifying clinical behavior in institutional settings to provide a basis for planning individual programs and evaluating results. McSweeney and Trout (1979) used the Jesness Behavior Checklist (Jesness, 1970) to evaluate the social progress of deviant children in a wilderness camp program. Five reasons for obtaining measures of students

are offered by Salvia and Ysseldyke (1981): "Screening, placement, program planning, program evaluation, and assessment of individual programs" (p. 14). Behavior rating scales have been used to obtain measures for each of these needs.

## Summary

This review has identified five approaches in the literature to define disruptive school behavior (DSB). A conceptualization of DSB based on the interactions of students, teachers, and administrators within the school setting was suggested as most relevant for the development of an instrument to quantify DSB.

Psychometric challenges to the use of rating scales for identifying behavioral characteristics were considered. Research was cited to suggest that teachers using reliable and valid scales could accurately identify DSB. Nineteen instruments available for assessing problem behaviors were reviewed. None appeared to meet the psychometric criteria required for educational placement decisions. Possible sources of error in nonsystematic observations were presented with the suggestion that inaccurate, biased, or subjective judgments may result.

Type, frequency, and severity of behaviors were related to item content, item format, and response format. Support was found for the inclusion of these

measurement parameters in assessing DSB. The use of descriptive statistics in current research for obtaining individual behavior ratings and deriving local norms was demonstrated.

A review of the sources of error in measurement was conducted and counter-measures for improving validity and reliability estimations were suggested. A number of variables affecting the norming of rating scales were investigated. Research evidence rejected separate norms based on gender, race, or socioeconomic status. The effective use of behavioral instruments in a variety of settings was documented, suggesting the suitability of such a device for describing students who display DSB.

# CHAPTER THREE

# METHODOLOGY

The purpose of this study was to develop and validate an instrument, the Disruptive Student Behavior Scale (DSBS). The DSBS is intended to be used to assess quantitatively the disruptive school behaviors of students referred for placement in either special education or alternative education programs. This chapter presents the research questions, defines the target population, presents a plan for constructing the scale, describes procedures for a pilot study, details statistical tests and procedures for the data analyses, and discusses possible limitations of the study.

## Research Questions

1. Does the content of the DSBS represent behaviors recognized and accepted by educators as occurring in and disruptive to the school environment?
2. In the judgment of experts, does the DSBS contain an equitable distribution of items descriptive of the underlying theoretical constructs that identify disruptive students and discriminate them from non-disruptive students?

3. To what degree does the DSBS demonstrate criterion, convergent, and discriminant validity?
4. To what degree does the DSBS provide ratings which are stable over time?

## Construction of the DSBS

The following plan is a modification of a suggested procedure (Benson & Clark, 1982) for rating scale construction. A review of disruptive school behavior (DSB) literature provided a research base for defining the constructs comprising DSB. A total of 303 descriptive items and 22 categories were found in 36 studies. After eliminating duplications and items not pertaining directly to DSB, 56 items remained. Combining similar categories resulted in a total of 13 potential categories of behaviors associated with DSB.

In a project conducted by the Research Committee of the Psychological Services Department of the Duval County, Florida School District, the 56 items and 13 categories were presented to 16 teachers of middle school students enrolled in a behavior management program for disruptive students. The rating group was composed of 10 females and 6 males, and all had at least two years full-time teaching experience. Group members were instructed to assign each item to one or more of the 13 categories. Instructions and results are reproduced in Appendix A.

The judges' ratings and comments resulted in the retention of 10 categories, which were considered to be one set of constructs which could be used in identifying DSB. A tentative definition of each construct was formulated using the descriptive items assigned by the teachers. A verification was attempted of the inclusiveness of these derived constructs. A frequency distribution was prepared for all of the conduct code violations reported for chronic violators in a sample of Duval County, Florida, elementary, middle/junior high, high, and alternative schools (Moses, 1981). Of 7717 behavior violations, 7686, or 99.6%, were included within the definitions of the proposed constructs.

The items, as taken from the studies and used in developing the constructs, were not considered specific enough for use in a quantitative rating scale. However, a readily available pool of potential items was located in the disciplinary referral records of an inner-city junior high school in a metropolitan Florida school district. Verbatim transcriptions were made of the reasons recorded on the referral forms by teachers when sending students to the deans. All active folders for the 1980-1981 school year were reviewed. A total of 395 items, including duplications, were recorded without regard for gender, age, race, or grade level. Combining obvious duplications and

similarities resulted in 66 items (Appendix B) to be considered for inclusion in a scale for rating DSB.

All of the 66 items were then presented individually to six male and five female volunteers, experienced secondary school regular classroom teachers from suburban Florida middle and junior high schools. Instructions are reproduced in Appendix C. These teachers were asked to verify the specificity of the items and edit those considered ambiguous. This review yielded 40 items for possible use on an instrument. These items were then stated in the past tense to reflect the intention to measure students' past behavior (Appendix D). This preliminary study indicated the feasibility of using research-based constructs and teacher-generated items as the basis for a rating scale for disruptive school behavior.

In order to reduce halo and leniency errors, it has been suggested (Blanz & Ghiselli, 1972) that a scale be arranged so that items from the same construct will not be contiguous. Accordingly, items were initially randomly ordered, then inspected and rearranged to meet this criterion. (Appendix M). Research studies previously cited suggested that in addition to specifying the type, a quantitative measure of disruptive behavior must provide for rating both frequency and severity. Frequency rating was provided for by the choice of response format selected for the instrument. The literature review suggested the

suitabilty of a 5-point, equal interval, summated rating scale (Likert, 1932) using the following anchors:

0 - None of the time

1 - Very infrequently

2 - Sometimes

3 - Quite often

4 - Always (Pohl, 1981, p. 239)

The rating scale (Appendix F) utilized this response format.

The severity rating for each scale item was established with assistance from the faculty, staff, and administrators of two alternative schools located in two metropolitan Florida school districts. From their experience with disruptive students, these educators were particularly aware of the consequences for students who display DSB. Respondents were selected from volunteers, including the principal and assistant principal, school psychologist, social worker, educational evaluator, and faculty members. This group contained both males and females in approximately equal numbers. All had more than two years' experience working with disruptive students.

The school experience may be conceptualized as influencing the social, personal, and academic domains of a student's life. Each of these domains may be subdivided to facilitate closer study of the consequences of the school experience (See Table 1). One way for educators to

assign a severity factor to a disruptive activity is to have them estimate which domains of student life would likely be affected adversely by that particular behavior. Instructions for this procedure are reproduced in Appendix G. The number of adverse consequences assigned by at least 50% of the raters, divided by a constant of three to keep the numbers small and with fractions rounded up to the next whole number, gave a severity rating of 1, 2, or 3 to each of the items on the rating scale. Results are reported in Chapter Four.

A scoring template incorporating the severity factor was prepared for the DSBS (Appendix H). This template has five holes, one corresponding to each possible frequency rating (i.e., 0, 1, 2, 3, 4) for each rating scale item. Through the holes are read the rater's mark (X) indicating the frequency rating assigned. Above each hole is printed a number which is the product of that frequency rating and the previously determined severity factor for that item. Thus, the weighted score for that item may be read by the scorer directly from the scoring template and recorded on the DSBS rating form beside each item.

These item scores were then added to give the page score and form score (see Appendix F) and recorded onto a summary sheet (Appendix I). The Summary of Teacher Ratings form (Appendix I) contains for each student the DSBS rating; the deviation, in z-scores, from the local

Table 1. Domains of Student Life Influenced by the School Experience

| Social | Personal | Academic |
|---|---|---|
| 1. Interpersonal relations<br>a. school personnel<br>b. peers<br>2. Vocational opportunities | 1. Cognitive development<br>2. Affective development<br>3. Physical development | 1. Learning vs. ignorance<br>2. Passing vs. failing<br>3. Attending vs. suspension or expulsion |

norm; a comparison of ratings by each teacher; and the basis for constructing a DSBS profile for prescriptive use (Appendix J). These data are intended to provide local school authorities with criteria for estimating the deviation of any student's rating from the local DSB norm and are intended to assist in determining a student's need for an intervention program. The DSBS is normed locally within each school district. Norms from this study are reported in Chapter Four for information, but are not to be used as criteria for judgments about students in other settings.

## Validation of the DSBS

To assure content validity, the 40 items and the 10 constructs developed from this preliminary study were presented to a group of 24 teachers with instructions to assign each item to a construct category or to no category. The instructions are reproduced in Appendix E. Each judge had at least two years of regular classroom teaching experience in a middle or junior high school. Thirteen male and 11 female teachers participated. The judges were also asked to verify the specificity of the retained items and reword those considered ambiguous. Revisions were made as suggested and confirmed by a follow-up study using another group of eight similarly-qualified teachers.

As described in the field study section, at a Florida middle school a criterion group of disruptive students was selected by nomination by seven non-teaching school personnel, including two deans, three guidance counselors, and two administrators. Students in the disruptive group were ranked numerically on a continuum from non- to severely disruptive, based on subjective ratings from all the nominating personnel. DSBS ratings from teachers were compared to these subjective ratings to determine how well high DSBS teacher ratings correlated with high levels of disruptiveness as perceived by non-teaching school personnel.

To estimate how well the DSBS identified the disruptive group, the mean of DSBS ratings for the disruptive group students were compared with the mean of DSBS ratings for a norming group representing a sample, stratified by grade, of the school population. If the DSBS demonstrated agreement with the concurrent judgments of disruptiveness made by non-teaching school officials, there would be made a prima facie case for predicting that students in other settings identified by the DSBS as disruptive would also be judged disruptive by non-teaching school officials.

Messick (1980) described construct validity as based on convergent and discriminant validity, ethical interpretation, relevance and utility for the specific application, and the consequences following use of the

instrument. Convergent validity requires that the DSBS be able to identify all students who are considered excessively disruptive. To demonstrate satisfactory convergent validity, the DSBS ratings of 100% of the students in the disruptive group would have to be significantly above the local DSB norm. The disruptive group ratings are reported in Chapter Four.

Discriminant validity requires that the DSBS be able to reject those students who are not considered excessively disruptive. To demonstrate satisfactory discriminant validity, the DSBS ratings of only those students in the disruptive group, or eligible for inclusion, could be significantly above the local DSB norm. Ratings of the norming group are reported in Chapter Four.

Ethical interpretation of DSBS ratings requires an understanding of both the theoretical and practical concepts underlying development of this instrument. Therefore, a manual will be prepared before the DSBS is offered for research use. Relevance was supported by the theoretical basis on which the 10 constructs were chosen to define DSB for this study. Utility was provided by the procedures used to select appropriate items, score the forms, interpret the ratings, and present the results. The consequences of using the DSBS cannot be predicted until it is thoroughly researched. The intent is to improve the validity of the selection process for programs assisting disruptive students.

## Reliability of the DSBS

The DSBS rating for each student is an aggregate of scores from at least four teachers. A test-retest measure compared two DSB ratings obtained from individual teachers. Fourteen days after the receipt of teacher ratings, a follow-up rating by the same teachers of approximately 10% of both the norming and disruptive groups was made. These results are reported in Chapter Four.

The internal consistency of the DSBS was protected by choosing only items previously used by teachers to describe DSB. Item analysis is not an effective technique for establishing reliability of individual administrations of the DSBS. Patterns of disruptive behavior are often narrow and stereotypical, while the DSBS contains items descriptive of a broad range of possible behaviors. Thus, item scores were not likely to correlate with each other.

No attempt was made to assess interrater reliability. Classroom settings are conceptualized as discrete environments, whose norms for behavior are determined by the personality of the teacher. The behavior of interest is the interaction of students with their teachers totally, not individually.

## Field Study

The purpose of the field study was to identify and correct any problems, actual or potential, with item

content, response format, or administration and scoring procedures of the DSBS. Following a successful field study, the instrument may be offered to the profession for further research and development (Benson & Clark, 1980). Accordingly, the operational goal of this present effort was to conduct a field study to determine the readiness of the DSBS for use as a research instrument.

The target population consisted of students enrolled in grades six through nine (i.e., middle and junior high school grades) in public schools anywhere in the United States. No restrictions were placed on age, gender, race or socioeconomic status. The selection criteria for the host school were a heterogeneous ethnic population, an urban or suburban location, public middle (grades 6, 7, 8) or junior high (grades 7, 8, 9) school status, random assignment of students to basic courses, and an average daily attendance figure of at least 500 students. Special schools, such as alternative schools and special education centers, were not considered.

A public middle school meeting these criteria was located in a predominately urban school district on the west coast of Florida. The student enrollment was approximately 76% white, 22% black, and 2% Asian- and Hispanic-American, with an average daily attendance of 733. Socioeconomic status was said by the principal to be primarily upper-lower class and lower-middle class.

For the norming group, a sample consisting of 90 students was selected using one English and one mathematics class, with randomly-assigned enrollments, at each grade level. A total of six classes containing 203 students and ranging from 32 through 35 students each were sampled. The numbers 1 through 35 were written on individual slips of paper and 15 numbers drawn randomly using the replacement procedure. For each class, the students whose class roll numbers matched the 15 randomly selected numbers were included in the norming groups.

The disruptive group was selected by nomination by non-teaching school personnel, who were asked to list the names of all of the excessively disruptive students encountered during the current school year. It was thus possible for a student's name to be included in both the norming and the disruptive groups. The nominating process initially produced a group of 64 students. After a conference among the raters, this group was reduced to 36 students.

All students finally nominated into the disruptive group were assigned to one of four levels of disruptiveness (none, mild, moderate, or severe) by each nominating person working independently. Nominated students were assigned a numerical rating according to the following scale:

| Level of Disruptiveness | Rating |
|---|---|
| None | 0 |
| Mild | 1 |
| Moderate | 2 |
| Severe | 3 |

Students were ranked according to the average of these ratings. This ranking permitted the correlation, reported in Chapter Four, of levels of disruptiveness between the DSBS results and the qualitative assessments by school personnel for each disruptive group student.

Schedules for the sample students were obtained from school records. No contact was made with any student. Training of all participating teachers took place in a meeting at which a DSBS form for each period of a sample student's current schedule was distributed. Appendix K contains these instructions. The purpose of the study was explained and a date and procedure for returning the forms agreed upon. Emphasis was placed on the need to respond to only the behaviors actually mentioned on the instrument and to perform the ratings independently of other teachers. Provision was made for a faculty member to either answer or refer questions that might arise during the rating period.

Teachers not submitting all their DSBS forms by the agreed upon date were contacted and reminded of the importance of their participation. Upon receipt of at

least four completed DSBS forms for each student, the DSBS rating for that student was calculated. At least four scorable forms, totaling 622, were received for 108 students, 76 in the norming group and 32 in the disruptive group. The scoring template (Appendix H) provided for calculating item scores weighted for severity.

The item scores were totaled to produce a form score, which was entered on the Summary of Teacher Ratings form (Appendix I). This summary form contains spaces for the student's name, grade, age, and sex; school name; evaluator's name and title; individual form scores; each rater's name, subject, and class period; and calculation of the student's DSBS rating and z-score. Each sample student's form scores were summed to give a total score. The total score was divided by the number of raters to yield the average score, which is the student's DSBS rating.

After the DSBS ratings for all the norming group students were calculated, the mean DSBS rating and standard deviation for the group were obtained. This mean of the means is the local DSBS rating, or norm, for the target school. The local DSBS norm was subtracted from the students' DSBS rating, giving their deviation from the local norm.

Dividing this deviation by the local standard deviation gave the number of standard deviation units, or z-scores, the student's DSBS rating differed from the

local DSBS norm. The criterion of two standard deviation units above the local DSBS norm translates to a disruptiveness score higher than approximately 98% of the predicted scores from the school population. The distribution of scores obtained from the norming group was inspected to assure the existence of sufficient variance to make the z-scores meaningful.

A reliability check was performed. Fourteen days after all the rating forms were collected, approximately 10% of the students from both the norming and disruptive groups were selected to be representative of the range of scores. New forms were submitted to the original raters for rerating the same students and the results compared. These results are reported in Chapter Four. After completion of the data analyses, all participants were invited to a meeting to discuss the results, offer comments, and receive appreciation for their participation.

## Data Analyses

### Validity

To establish content validity, 24 expert judges assigned proposed DSBS items to construct categories. Results of the judges assignments were totaled for each item. An item was dropped if not assigned to at least one category by each judge. If this content validation procedure had resulted either in fewer than 30 items being

assigned to at least one construct or in having a construct with fewer than three items assigned by 80% of the respondents, enough items would have been constructed and validated to meet these criteria. The judges' item assignments are reported in Chapter Four.

To ascertain how well the DSBS identified the disruptive group, the t-test was used to estimate the significance of the difference between the means of the norming group and disruptive group. An obtained probability level of .05 or less was considered evidence of statistical significance. The magnitude of the difference between the means was used to evaluate the practical significance of the instrument and its potential for identifying disruptive students. These results are be reported and discussed in Chapter Four.

To estimate convergent validity for the DSBS, the DSBS rating for each disruptive group member was compared with the mean DSBS rating of the norming group. For the purposes of this study, a DSBS rating of at least two z-scores above the norming group mean was accepted as evidence that the DSBS had correctly identified a disruptive group member. The standard error of the mean was used to include students when evaluating borderline cases. The criterion for satisfactory convergent validity was the correct identification of 100% of the disruptive group.

Discriminant validity also was estimated by using ratings, means, and z-scores. The DSBS rating for each norming group member was compared with the mean of that group. Any norming group member whose DSBS rating exceeded the mean by at least two z-scores was considered identified by the DSBS as excessively disruptive. Identified cases, not members of or eligible for the disruptive group, were considered challenges to the discriminant validity of the DSBS. All cases not meeting the construct validity criteria were investigated. Construct validity results are reported and discussed in Chapter Four.

## Reliability

The Pearson product-moment correlation statistic was used to compare the original ratings on approximately 10% of the completed forms with follow-up ratings made after 14 days. Individual coefficients of at least .80 were set arbitrarily to establish an acceptable level of test-retest reliability.

## Limitations

1. The school for the field study was selected based on the willingness to cooperate by both the school and the faculty. This may have mitigated problems that would occur in a less favorable environment.

2. Teacher resistance and/or concerns about this type of research may have biased or limited their participation.
3. The study was limited to exploration and the results are not intended to generalize beyond the administration and scoring procedures. Specifically, the calculated DSB norm is valid only for this school.
4. No provision was made to assess the possible effects of grade and sex on DSB norms. Studies have indicated the influences are not significant, but at some point this should be investigated.
5. The disruptive sample group was likely composed of students who had been referred to the dean. The same teachers who referred these students to the dean may have rated their behaviors, with bias a possibility.
6. The use of expert judges in the validation procedures may have introduced personal bias into the items used on the instrument.

# CHAPTER FOUR

# RESULTS AND DISCUSSION

The purpose of this study was to develop and validate an instrument, the Disruptive Student Behavior Scale (DSBS). The study focused on identifying components of disruptive school behavior as perceived by middle and junior high school teachers and constructing an instrument to quantify these behaviors. To accomplish this, an instrument was constructed using behaviors taken from disciplinary referrals and field tested on a representative sample of students from a Florida middle school. Teacher ratings for a norm group and a disruptive group were collected and analyzed as outlined in Chapter Three. These results are reported in this chapter.

## Results

### The Severity Factor

Results of the assignment of potential adverse consequences resulting from DSBS behaviors are reported in Table 2. Twenty packets containing 40 DSBS items and an instruction sheet were distributed and 16 were returned. At least 50%, or 8, of the raters had to assign a DSBS item to a particular domain before that domain was

Table 2. Potential Adverse Consequences of DSBS Behaviors

| Old Item Number | Domains: Social 1a | Social 1b | 2 | Personal 1 | Personal 2 | Personal 3 | Academic 1 | Academic 2 | Academic 3 | Total Domains | Severity Rating Factor | DSBS Item Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 16 | 8 | 5 | 3 | 2 | 0 | 10 | 8 | 5 | 4 | 2 | 1 |
| 2. | 13 | 0 | 3 | 2 | 7 | 0 | 6 | 5 | 3 | 1 | 1 | 13 |
| 3. | 12 | 15 | 7 | 6 | 3 | 0 | 15 | 11 | 8 | 5 | 2 | 3 |
| 4. | 13 | 5 | 2 | 4 | 8 | 0 | 3 | 2 | 4 | 2 | 1 | 4 |
| 5. | 10 | 13 | 2 | 2 | 5 | 1 | 1 | 2 | 13 | 3 | 1 | 2 |
| 6. | 9 | 12 | 4 | 2 | 8 | 0 | 2 | 1 | 14 | 4 | 2 | 36 |
| 7. | 12 | 1 | 5 | 4 | 8 | 0 | 9 | 14 | 8 | 5 | 2 | 25 |
| 8. | 15 | 3 | 6 | 1 | 9 | 0 | 0 | 1 | 9 | 3 | 1 | 5 |
| 9. | 9 | 14 | 0 | 10 | 0 | 0 | 0 | 0 | 2 | 3 | 1 | 28 |
| 10. | 12 | 2 | 4 | 2 | 8 | 0 | 8 | 9 | 8 | 5 | 2 | 16 |
| 11. | 13 | 3 | 0 | 1 | 8 | 0 | 2 | 5 | 11 | 3 | 1 | 26 |
| 12. | 12 | 9 | 0 | 9 | 0 | 0 | 0 | 0 | 2 | 3 | 1 | 19 |
| 13. | 11 | 11 | 3 | 8 | 9 | 0 | 2 | 1 | 0 | 4 | 2 | 15 |
| 14. | 8 | 8 | 0 | 8 | 11 | 2 | 0 | 0 | 1 | 4 | 2 | 37 |
| 15. | 13 | 14 | 1 | 2 | 8 | 0 | 1 | 1 | 13 | 4 | 2 | 17 |
| 16. | 12 | 3 | 8 | 11 | 3 | 0 | 13 | 10 | 1 | 5 | 2 | 18 |
| 17. | 12 | 0 | 4 | 2 | 0 | 1 | 1 | 1 | 10 | 2 | 1 | 20 |
| 18. | 9 | 10 | 5 | 8 | 8 | 0 | 0 | 0 | 1 | 4 | 2 | 29 |
| 19. | 11 | 15 | 3 | 1 | 8 | 1 | 2 | 1 | 8 | 4 | 2 | 12 |
| 20. | 13 | 3 | 3 | 2 | 3 | 0 | 12 | 11 | 9 | 4 | 2 | 21 |
| 21. | 16 | 7 | 3 | 2 | 8 | 0 | 0 | 0 | 2 | 2 | 1 | 27 |
| 22. | 10 | 8 | 2 | 1 | 2 | 0 | 2 | 2 | 8 | 3 | 1 | -- |
| 23. | 10 | 9 | 3 | 8 | 15 | 0 | 0 | 0 | 8 | 5 | 2 | 22 |

Table 2.

(Continued)

| Old Item Number | Domains: Social 1a | Social 1b | Social 2 | Personal 1 | Personal 2 | Personal 3 | Academic 1 | Academic 2 | Academic 3 | Total Domains | Severity Rating Factor | DSBS Item Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24. | 12 | 2 | 2 | 9 | 5 | 0 | 0 | 0 | 9 | 3 | 1 | 11 |
| 25. | 15 | 5 | 4 | 10 | 8 | 0 | 3 | 2 | 12 | 4 | 2 | 23 |
| 26. | 12 | 1 | 1 | 8 | 2 | 0 | 10 | 16 | 2 | 4 | 2 | 10 |
| 27. | 11 | 8 | 3 | 2 | 10 | 0 | 0 | 3 | 0 | 3 | 1 | 35 |
| 28. | 14 | 8 | 2 | 11 | 13 | 0 | 0 | 1 | 1 | 4 | 2 | 7 |
| 29. | 12 | 1 | 1 | 1 | 0 | 0 | 11 | 15 | 1 | 3 | 1 | 9 |
| 30. | 13 | 10 | 2 | 8 | 8 | 0 | 0 | 0 | 11 | 5 | 2 | 6 |
| 31. | 12 | 2 | 1 | 1 | 8 | 0 | 2 | 2 | 8 | 3 | 1 | 31 |
| 32. | 14 | 15 | 1 | 8 | 8 | 0 | 9 | 8 | 9 | 7 | 3 | 34 |
| 33. | 11 | 8 | 1 | 11 | 11 | 0 | 1 | 1 | 9 | 5 | 2 | 8 |
| 34. | 12 | 3 | 2 | 9 | 8 | 0 | 2 | 2 | 15 | 4 | 2 | 30 |
| 35. | 15 | 2 | 2 | 10 | 9 | 0 | 2 | 2 | 16 | 4 | 2 | 32 |
| 36. | 10 | 3 | 8 | 8 | 10 | 9 | 9 | 8 | 12 | 8 | 3 | -- |
| 37. | 12 | 2 | 9 | 8 | 11 | 8 | 8 | 8 | 14 | 8 | 3 | 14 |
| 38. | 14 | 10 | 8 | 8 | 8 | 0 | 0 | 0 | 11 | 6 | 2 | 24 |
| 39. | 12 | 11 | 11 | 2 | 9 | 1 | 2 | 2 | 15 | 5 | 2 | 33 |
| 40. | 13 | 10 | 8 | 8 | 8 | 1 | 2 | 2 | 16 | 6 | 2 | 38 |

included in the severity factor calculation. Table 2 includes the DSBS item numbers as presented to the raters, the item numbers as randomly assigned for the final revision of the DSBS, the assignment to domains of the individual items, the number of domains assigned by at least 50% of the raters, and the calculated severity factor for each item. Severity factors of 1, 2, and 3 were obtained for 14, 23, and 3 items, respectively. Items 38, 39, and 40 received ratings producing a severity factor of 2. Each of these is a law violation and thus was arbitrarily upgraded to a severity factor of 3 to recognize the seriousness of the behavior.

A reliability check was performed by having 8 of the original raters rerate all 40 items. The Pearson product-moment correlation statistic was used to compare the number of domains assigned to each item, resulting in correlations ranging from .89 to 1.00, with an overall correlation coefficient of .92.

## The Samples

A public middle school (grades 6, 7, 8) in a predominantly urban school district on the west coast of Florida was selected as the host school for the study. For the norming group, a sample consisting of 90 students was selected using one English and one mathematics class, with randomly assigned enrollments, at each grade level.

At least four scorable rating forms, totaling 424, were received for 76 students. For each student, an average of 5.6 ratings was received. Table 3 shows the distribution of the rating forms by grade, age, and gender. The higher grades (7 & 8) and middle ages (13 & 14) accounted for 72% and 61%, respectively, of the cases. The gender distribution was approximately equal, i.e., 36 females and 40 males. Overall, the coverage of grade, age, and gender is typical for a middle school.

The disruptive group was selected by nomination by non-teaching school personnel, who were asked to list the names of all of the excessively disruptive students encountered during the current school year. Thirty-six students were nominated for this group. At least four scorable rating forms, totaling 198, were received for 32 students. For each student, an average of 6.2 ratings was received. Table 4 shows the distribution of the rating forms by grade, age, and gender. The sample is unevenly distributed, with grade 8, ages 14 and older, and males accounting for 63%, 78%, and 78%, respectively, of the cases.

Research Question One

Does the content of the DSBS represent behaviors recognized and accepted by educators as occurring in and disruptive to the school environment?

Table 3. Rating Form Distribution by Demographic Categories--Norming Group

| | Sample N(76) | % Total | Forms N(424) | % Norm |
|---|---|---|---|---|
| Grade | | | | |
| 6 | 21 | 28 | 107 | 25 |
| 7 | 26 | 34 | 153 | 36 |
| 8 | 29 | 38 | 164 | 39 |
| Age | | | | |
| 12 & below | 18 | 24 | 92 | 22 |
| 13 | 21 | 28 | 120 | 28 |
| 14 | 25 | 33 | 152 | 36 |
| 15 & above | 12 | 16 | 60 | 14 |
| Gender | | | | |
| Female | 36 | 47 | 207 | 49 |
| Male | 40 | 53 | 217 | 51 |

Table 4. Rating Form Distribution by Demographic Categories--Disruptive Group

| | Sample N(32) | % Total | Forms N(198) | % Total |
|---|---|---|---|---|
| Grade | | | | |
| 6 | 2 | 6 | 12 | 6 |
| 7 | 10 | 31 | 61 | 31 |
| 8 | 20 | 63 | 125 | 63 |
| Age | | | | |
| 12 & below | 1 | 3 | 6 | 3 |
| 13 | 6 | 19 | 35 | 18 |
| 14 | 11 | 34 | 68 | 34 |
| 15 & above | 14 | 44 | 89 | 45 |
| Gender | | | | |
| Female | 7 | 22 | 44 | 22 |
| Male | 25 | 78 | 154 | 78 |

All items on the DSBS were developed using behavioral statements taken from the disciplinary records of a junior high school (Appendix B). The item development procedure is described in detail in Chapter Three and the 40 proposed items are listed in Appendix D. In the current study, these 40 items were presented for a final review to 24 experienced middle and junior high school teachers.

These judges were asked specifically to identify any items not immediately recognizable as potentially occurring in and disruptive to the school environment. Seven of the 40 items were so identified. Items 2, 7, 23, and 33 were reworded for clarification. Items 19 and 22 were dropped as requiring inferences from the rater. Item 36 was combined with item 16. These revisions resulted in the retention of 37 items. These final items were accepted unanimously by a similarly qualified group of eight teachers. Thus the item content of the DSBS does appear to represent behaviors accepted by educators as descriptive of disruptive school behavior.

To investigate the recognition of behaviors described by the DSBS items, a frequency count was made of the responses to individual items by five raters, all of whom had rated the same 10 students, 8 from the norming group and 2 from the disruptive group. Ratings averaged across the five raters ranged from .2 for the lowest to 86.6 for the highest rated student. Table 5 summarizes these

Table 5. Frequency of Observed DSBS Behaviors by Constructs

| | Construct Number | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Frequency of Observation | 33(1) | 38(3) | 3(8) | 6(4) | 12(2) | 24(10) | 18(5) | 11(7) | 22(9) | 1(6) |
| | 10(13) | 15(19) | 13(15) | 8(11) | 23(17) | 23(18) | 5(20) | 6(12) | 2(16) | 1(14) |
| | 12(26) | 19(27) | 17(22) | 1(30) | 1(23) | 13(25) | 24(29) | 15(28) | 3(21) | 1(24) |
| | 10(31) | 31(34) | | 6(37) | 1(32) | | 7(35) | | | 2(33) |
| | | | | | 3(38) | | | | | 13(36) |
| Total Observations Per Construct | 65 | 103 | 33 | 21 | 40 | 60 | 54 | 32 | 27 | 18 |
| % of Total Observations | 14 | 23 | 7 | 5 | 9 | 13 | 12 | 7 | 6 | 4 |

( ) = DSBS Item Number

responses. Response frequencies for individual items ranged from 1 (for 6 items) to 38 (for 1 item). These results indicate that all of the items were recognized at least once. Twenty-one of the items were recognized 10 or more times. This suggests that the content included in the DSBS items is recognized by educators frequently enough to justify the inclusion of each item on the scale and to provide for the rating of a broad sprectrum of disruptive behavior occurring in the school environment.

Research Question Two

In the judgment of experts, does the DSBS contain an equitable distribution of items descriptive of the underlying theoretical constructs that identify disruptive students and discriminate them from non-disruptive students? For the purposes of this study, experts were defined as educators with at least two years' teaching experience in an appropriate setting. The constructs used as criteria for selecting DSBS items were derived from a review of disruptive school behavior literature. This procedure is described in detail in Chapter Three and Appendix A.

Table 6 contains a list of the final 10 constructs. Appendix D contains the 40 proposed items. In the current study, these constructs were presented individually to 24 experienced middle and junior high school teachers with

Table 6. DSBS Constructs by Number

| Construct | Number |
|---|---|
| Disobedience | 1 |
| Disruptiveness | 2 |
| Impulsiveness | 3 |
| Destructiveness | 4 |
| Aggression | 5 |
| Academic irresponsibility | 6 |
| Social/personal irresponsibility | 7 |
| Ineffective interpersonal relationships | 8 |
| Attendance violations | 9 |
| Law violations | 10 |

instructions to match the proposed DSBS items to the constructs. Table 7 contains these results.

Eight of the constructs met the criterion of having at least three items assigned by 80% of the raters. Only two items met the criterion for the two constructs labeled "Ineffective Interpersonal Relations" and "Impulsiveness." To remedy these deficiencies, one additional item for the "Ineffective Interpersonal Relations" construct was prepared in consultation with the raters and incorporated into the scale as item 12. The final revision of the DSBS thus contains 38 items. Item 33 was rewritten to be more indicative of "Impulsiveness."

The final 38 items and 10 constructs were presented to eight experienced junior high school teachers. Each construct had at least three items assigned by seven, or 88%, of the raters. Table 8 reports these results. Four constructs are represented by three items, four constructs are represented by four items, and two constructs are represented by five items. This distribution of items to constructs appears equitable and thus this research question is answered in the affirmative.

Research Question Three

To what degree does the DSBS demonstrate criterion, convergent and discriminant validity? Ninety students were selected for the norming group and 36 students for

Table 7. Assignment of Proposed Scale Items to Constructs for Content Validation

| Item Number | Constructs 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | % Agreement | Action Taken |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 23 | 1 | | | | | | | | | 96 | |
| 2. | 18 | | | | 2 | | | 4 | | | 75 | Revised |
| 3. | | 22 | 2 | | | | | | | | 92 | |
| 4. | | | | 23 | | | | 1 | | | 96 | |
| 5. | | | | | 24 | | | | | | 100 | |
| 6. | | 1 | 1 | | 20 | 1 | | 1 | | | 83 | |
| 7. | | | 1 | | | 23 | | | | | 96 | Revised |
| 8. | | | | | | 1 | 23 | | | | 96 | |
| 9. | | | | | | | 1 | 23 | | | 96 | |
| 10. | 2 | | | | | | 1 | 1 | 20 | | 83 | |
| 11. | 22 | | | | | | 1 | 1 | | | 92 | |
| 12. | 1 | 20 | 1 | | | | 1 | 1 | | | 83 | |
| 13. | | | 21 | | | | 1 | 2 | | | 88 | |
| 14. | | | | 24 | | | | | | | 100 | |
| 15. | | | | 1 | 21 | | | 1 | | 1 | 88 | |
| 16. | | 1 | | | | 22 | 1 | | | | 92 | |
| 17. | 2 | | 1 | | | 1 | 20 | | | | 83 | |
| 18. | | | | | | 1 | 23 | | | | 96 | |
| 19. | | | | 4 | 12 | | 2 | 6 | | | 50 | Dropped |
| 20. | 1 | | | | | | 1 | | 22 | | 92 | |
| 21. | 2 | 20 | 2 | | | | | | | | 83 | |
| 22. | 4 | 4 | | | | | 2 | 14 | | | 58 | Dropped |
| 23. | | | 23 | 1 | | | | | | | 96 | Revised |

Table 7.

(Continued)

| Item Number | Constructs 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | % Agreement | Action Taken |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24. | | | | 24 | | | | | | | 100 | |
| 25. | 1 | | | | 21 | | | 1 | | 1 | 88 | |
| 26. | | | | | | 24 | | | | | 100 | |
| 27. | | | | | | 2 | 22 | | | | 92 | |
| 28. | | | | | | | | 24 | | | 100 | |
| 29. | | | | | | | | | 24 | | 100 | |
| 30. | | | | | | | | | | 24 | 100 | |
| 31. | 20 | | | | 3 | | | 1 | | | 83 | |
| 32. | | 24 | | | | | | | | | 100 | |
| 33. | 3 | 7 | 14 | | | | | | | | 58 | Revised |
| 34. | | | | 22 | | | | | | 2 | 92 | |
| 35. | 1 | | 1 | | 20 | | | 1 | | 1 | 83 | |
| 36. | | | 2 | | 4 | 7 | 4 | | | 7 | 29 | Combined w/ #16 |
| 37. | | | 1 | | | | | | | 23 | 96 | |
| 38. | | | 1 | | | | | | | 23 | 96 | |
| 39. | | | 1 | | | | | | | 23 | 96 | |
| 40. | | 1 | | | | | | | | 23 | 96 | |
| Totals | 3 | 4 | 2 | 4 | 5 | 3 | 4 | 2 | 3 | 5 | | |

Table 8. Follow-up Study for Assignment of Proposed Scale Items to Constructs

| Item Number | Constructs 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | % Agreement | Action Taken |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 8 | | | | | | | | | | 100 | |
| 2. | 7 | | | | | | | 1 | | | 88 | |
| 3. | | 8 | | | | | | | | | 100 | |
| 4. | | | | 8 | | | | | | | 100 | |
| 5. | | | | | 8 | | | | | | 100 | |
| 6. | | | | | 7 | | | 1 | | | 88 | |
| 7. | | | | | | 8 | | | | | 100 | |
| 8. | | | | | | | 8 | | | | 100 | |
| 9. | | | | | | | | 8 | | | 100 | |
| 10. | 1 | | | | | | | | 7 | | 88 | |
| 11. | 8 | | | | | | | | | | 100 | |
| 12. | 1 | 7 | | | | | | | | | 88 | |
| 13. | | | 7 | | | | | 1 | | | 88 | |
| 14. | | | | 8 | | | | | | | 100 | |
| 15. | | | | | 7 | | | 1 | | | 88 | |
| 16. | | | | | | 8 | | | | | 100 | |
| 17. | 1 | | | | | | 7 | | | | 88 | |
| 18. | | | | | | | 8 | | | | 100 | |
| 19. | | | | | | | | | | | | Dropped |
| 20. | | | | | | | | | 8 | | 100 | |
| 21. | 1 | 7 | | | | | | | | | 88 | |
| 22. | | | | | | | | | | | | Dropped |
| 23. | | | 8 | | | | | | | | 100 | |

Table 8.

(Continued)

| Item Number | Constructs 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | % Agreement | Action Taken |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24. | | | | 8 | | | | | | | 100 | |
| 25. | | | | | 7 | | | | | 1 | 88 | |
| 26. | | | | | | 8 | | | | | 100 | |
| 27. | | | | | | | 8 | | | | 100 | |
| 28. | | | | | | | | 8 | | | 100 | |
| 29. | | | | | | | | | 8 | | 100 | |
| 30. | | | | | | | | | | 8 | 100 | |
| 31. | 7 | | | | 1 | | | | | | 88 | |
| 32. | | 8 | | | | | | | | | 100 | |
| 33. | | | 8 | | | | | | | | 100 | |
| 34. | | | | 8 | | | | | | | 100 | |
| 35. | | | | | 7 | | | | | 1 | 88 | |
| 36. | | | | | | | | | | | | Combined w/ #16 |
| 37. | | | | | | | | | | 8 | 100 | |
| 38. | | | | | | | | | | 8 | 100 | |
| 39. | | | | | | | | | | 8 | 100 | |
| 40. | | | | | | | | | | 8 | 100 | |
| Totals | 4 | 4 | 3 | 4 | 5 | 3 | 4 | 2 | 3 | 5 | | |
| | | | | | | New Item | | 1 | | | | |
| | | | | | | | | 3 | | | | |

the disruptive group. DSBS forms totaling 882 were distributed to 39 teachers. At least four scorable forms were received for 76, or 84%, of the norming group and 32, or 89%, of the disruptive group. A total of 622 scorable forms was used in the validity studies of the DSBS. Ratings ranged from .2 to 85.5 and 55.1 to 86.9 for the norming and disruptive groups, respectively, indicating the DSBS provides for collecting data representative of a wide-ranging population.

Criterion validity was estimated by comparing subjective ratings by non-teaching personnel with teachers' DSBS ratings of the disruptive group students. Table 9 presents these ratings. A Pearson product-moment correlation significant at the $p \leq .01$ ($r = .47$, $df = 30$) was obtained, indicating a positive relationship.

Using the t-test to compare means (Tables 10 and 11) of the disruptive and norming groups produced a difference significant at the $p \leq .01$ level ($t = 18.4$ , $df = 106$). A comparison of the values of the means, 64.9 and 19.9, respectively for the two groups, suggests the difference is meaningful and that the DSBS was able to identify a criterion group of disruptive students.

Convergent validity was estimated by comparing the DSBS ratings, converted to z-scores, of each disruptive group student to the mean DSBS rating for the norming group. Tables 10 and 11 contain the DSBS ratings and

Table 9. Comparison of Disruptiveness Ratings by Teachers and Non-teaching Personnel

| Student Number | Teachers' DSBS Rating | Non-Teaching Personnel Subjective Rating |
|---|---|---|
| 1 | 55.1 | 4 |
| 2 | 55.1 | 3 |
| 3 | 55.2 | 3 |
| 4 | 55.9 | 3 |
| 5 | 56.4 | 3 |
| 6 | 56.9 | 4 |
| 7 | 57.3 | 3 |
| 8 | 57.8 | 4 |
| 9 | 58.2 | 3 |
| 10 | 58.6 | 3 |
| 11 | 59.1 | 3 |
| 12 | 59.6 | 3 |
| 13 | 60.1 | 4 |
| 14 | 61.2 | 4 |
| 15 | 62.3 | 4 |
| 16 | 62.8 | 4 |
| 17 | 63.3 | 4 |
| 18 | 63.8 | 3 |
| 19 | 64.2 | 4 |
| 20 | 64.7 | 4 |
| 21 | 65.1 | 3 |
| 22 | 65.6 | 3 |
| 23 | 66.1 | 4 |
| 24 | 66.5 | 4 |
| 25 | 66.8 | 4 |
| 26 | 66.9 | 4 |
| 27 | 67.0 | 4 |
| 28 | 81.1 | 4 |
| 29 | 84.3 | 4 |
| 30 | 85.5 | 4 |
| 31 | 86.0 | 4 |
| 32 | 86.9 | 4 |
| | $r = .47^{**}$ | |

**$P \leq .01$

Table 10. DSBS Ratings and z-scores for the Disruptive Group

| Student Number | DSBS Rating | z-score |
|---|---|---|
| 101 | 55.1 | 2.3 |
| 102 | 55.1 | 2.3 |
| 103 | 55.2 | 2.3 |
| 104 | 55.9 | 2.3 |
| 105 | 56.4 | 2.4 |
| 106 | 56.9 | 2.4 |
| 107 | 57.3 | 2.4 |
| 108 | 57.8 | 2.5 |
| 109 | 58.2 | 2.5 |
| 110 | 58.6 | 2.5 |
| 111 | 59.1 | 2.6 |
| 112 | 59.6 | 2.6 |
| 113 | 60.1 | 2.6 |
| 114 | 61.2 | 2.7 |
| 115 | 62.3 | 2.8 |
| 116 | 62.8 | 2.8 |
| 117 | 63.3 | 2.8 |
| 118 | 63.8 | 2.9 |
| 119 | 64.2 | 2.9 |
| 120 | 64.7 | 2.9 |
| 121 | 65.1 | 2.9 |
| 122 | 65.6 | 3.0 |
| 123 | 66.1 | 3.0 |
| 124 | 66.5 | 3.0 |
| 125 | 66.8 | 3.0 |
| 126 | 66.9 | 3.0 |
| 127 | 67.0 | 3.1 |
| 128 | 81.1 | 4.0 |
| 129 | 84.3 | 4.2 |
| 130 | 85.5 | 4.3 |
| 131 | 86.0 | 4.3 |
| 132 | 86.9 | 4.4 |

Mean = 64.9
Standard Deviation = 9.5
Standard Error = 1.7

Table 11. DSBS Ratings and z-scores for the Norming Group

| Student Number | DSBS Rating | z-score |
|---|---|---|
| 1 | .2 | -1.3 |
| 2 | 1.1 | -1.2 |
| 3 | 1.2 | -1.2 |
| 4 | 1.6 | -1.2 |
| 5 | 2.1 | -1.2 |
| 6 | 2.3 | -1.1 |
| 7 | 4.2 | -1.0 |
| 8 | 5.3 | -1.0 |
| 9 | 5.9 | -0.9 |
| 10 | 6.4 | -0.9 |
| 11 | 6.7 | -0.9 |
| 12 | 7.5 | -0.8 |
| 13 | 7.8 | -0.8 |
| 14 | 8.6 | -0.7 |
| 15 | 10.1 | -0.6 |
| 16 | 11.5 | -0.6 |
| 17 | 13.1 | -0.4 |
| 18 | 14.2 | -0.4 |
| 19 | 14.8 | -0.3 |
| 20 | 15.3 | -0.3 |
| 21 | 15.5 | -0.3 |
| 22 | 15.9 | -0.3 |
| 23 | 16.1 | -0.2 |
| 24 | 16.3 | -0.2 |
| 25 | 16.4 | -0.2 |
| 26 | 16.6 | -0.2 |
| 27 | 16.7 | -0.2 |
| 28 | 16.8 | -0.2 |
| 29 | 16.8 | -0.2 |
| 30 | 16.9 | -0.2 |
| 31 | 17.0 | -0.2 |
| 32 | 17.2 | -0.2 |
| 33 | 17.2 | -0.2 |
| 34 | 17.3 | -0.2 |
| 35 | 17.4 | -0.2 |
| 36 | 17.4 | -0.2 |
| 37 | 17.5 | -0.2 |
| 38 | 17.5 | -0.2 |
| 39 | 17.5 | -0.2 |
| 40 | 17.6 | -0.2 |

Table 11

(Continued)

| Student Number | DSBS Rating | z-score |
|---|---|---|
| 41 | 17.6 | -0.2 |
| 42 | 17.7 | -0.1 |
| 43 | 17.8 | -0.1 |
| 44 | 18.0 | -0.1 |
| 45 | 18.2 | -0.1 |
| 46 | 18.3 | -0.1 |
| 47 | 18.4 | -0.1 |
| 48 | 18.5 | -0.1 |
| 49 | 18.9 | -0.1 |
| 50 | 19.1 | -0.1 |
| 51 | 19.6 | 0.0 |
| 52 | 19.6 | 0.0 |
| 53 | 19.7 | 0.0 |
| 54 | 20.2 | 0.0 |
| 55 | 20.3 | 0.0 |
| 56 | 20.5 | 0.0 |
| 57 | 21.1 | 0.1 |
| 58 | 21.6 | 0.1 |
| 59 | 22.0 | 0.1 |
| 60 | 22.1 | 0.1 |
| 61 | 23.6 | 0.2 |
| 62 | 24.1 | 0.3 |
| 63 | 24.5 | 0.3 |
| 64 | 25.3 | 0.4 |
| 65 | 26.1 | 0.4 |
| 66 | 26.4 | 0.4 |
| 67 | 27.7 | 0.5 |
| 68 | 27.8 | 0.5 |
| 69 | 28.0 | 0.5 |
| 70 | 28.3 | 0.6 |
| 71 | 30.6 | 0.7 |
| 72 | 60.5 | 2.6 |
| 73 | 61.2 | 2.7 |
| 74 | 63.3 | 2.8 |
| 75 | 81.1 | 4.0 |
| 76 | 85.5 | 4.3 |

Mean = 19.9
Standard Deviation = 15.4
Standard Error = 1.77

z-scores for the disruptive and norming groups, respectively. Using the criterion for significance of two standard deviations, or two z-scores, above the mean, which will exclude approximately 98% of a normally distributed population, a DSBS rating of 30.8 was required for a student to be identified as disruptive. The standard error was used to include cases. The DSBS so identified 32, or 100%, of the students in the disruptive group. Thus, the DSBS demonstrated satisfactory convergent validity.

Discriminant validity was estimated by comparing each norming group student's DSBS rating with the criterion rating (30.8 or z-score $\geq 2$) for inclusion in the disruptive group. The scores of five norming group students met the criterion for their identification as disruptive. Two of these students also were in the disruptive group. The remaining three students had only one disciplinary referral each in the current school year and were not viewed by the deans as disruptive. The DSBS correctly excluded 96% of the norming group and appears to have demonstrated satisfactory discriminant validity.

Research Question Four

To what degree does the DSBS provide ratings which are stable over time? Fourteen days after the initial rating period, four students from the disruptive group and eight

from the norming group were rated again by their teachers (Table 12). Twenty-six forms were received for the disruptive group students. Pearson product-moment correlations ranging from .90 to .98 were obtained, with an overall correlation of r = .94. All correlations were significant at $p \leq .05$ or better.

For the norming group, 47 forms were received. Correlations ranging from .72 to .97 were obtained, with an overall correlation of .92. With the exception of the one correlation of .72, which was not significant, all correlations were significant at $p \leq .05$ or better. The DSBS provided ratings that appear consistent over a period of time. Temporal reliability for the DSBS thus appeared satisfactory.

Summary

The data seem to suggest that the content of the items selected for the final version of the DSBS was acceptable to teachers as being descriptive of the disruptive behaviors usually observed in middle and junior high schools. In addition, all of the behaviors described by the DSBS items were actually observed in the field study. The data on actual observations also support the conclusion that the DSBS items equitably represent underlying theoretical constructs of disruptive behavior as presented in the research literature.

Table 12. Test-Retest Correlations

| Disruptive Group | | Norming Group | |
|---|---|---|---|
| Student Number | Correlation Coefficient | Student Number | Correlation Coefficient |
| 107 | .98** | 8 | .72 |
| 114 | .95** | 16 | .92** |
| 121 | .90* | 24 | .90** |
| 128 | .94** | 32 | .92* |
| | | 40 | .97** |
| | | 48 | .94* |
| | | 56 | .91* |
| | | 64 | .97** |

* $p \leq .05$
** $p \leq .01$

Although the DSBS did not satisfactorily predict the subjective ratings by non-teaching personnel of individual disruptive students, the DSBS did identify the criterion group as being composed of disruptive students. With a high degree of accuracy, the DSBS identified individual disruptive students and excluded non-disruptive students. The reliability data suggest that the obtained DSBS ratings would be consistent over at least a 14-day period. Thus, the DSBS appeared to meet the proposed major criteria for content relevance and representativeness, validity, and reliability.

## Discussion

A major task of this study was choosing behaviors that would represent disruptiveness across a wide range of school environments. Using teachers' reports for an entire year of behaviors that had resulted in disciplinary referrals created a large number and variety of potential items. This procedure provided evidence of content validity for the items refined from this pool and utilized on the instrument.

Each proposed item had been evaluated several times in previous content validation studies in which the basis for judgment was the raters' past experience. Nevertheless, difficulties were experienced by teachers in applying to the school environment seven of these prior-approved

items. These teachers, with minimal assistance from the researcher, were able to make revisions to the items that resulted in unanimous acceptance of these items by another group of raters. This experience reaffirmed the hypothesis (Smith & Kendall, 1963) that involving experts in the content validation procedure and providing for testing items in the target environment would eliminate some problems encountered by instrument developers relying more on clinical descriptions and statistical determinations. No problems with applying any DSBS items to the study samples were reported.

Numerous studies, cited in the review of literature, have explored broad-band classifications of disruptive behavior. These investigative techniques were based primarily on statistical analysis or clinical classification systems, neither of which seemed to produce items descriptive of the most prevalent disruptive school behaviors. A project preliminary to this study utilized experienced teachers in selecting categories of disruptive behavior from those identified in these previous research efforts. Thus, the scope, or underlying construct base, of the DSBS was defined by expert judges utilizing research-based categories of disruptive behavior. Equitable coverage of the 10 constructs so selected was evaluated by having experienced teachers assign each proposed DSBS item to one of the constructs. After item

refinements, a final review by expert judges resulted in the assignment of from three to five items to each construct. This was accepted as demonstrating equitable distribution of items among the constructs defining the scope of the DSBS.

Reporting the distribution of DSBS items by constructs (Table 8) is not intended to imply subscale characteristics for the constructs. Too few items are included to assure adequate convergent and discriminant validity and no attempt was made to create mutually exclusive categories. However, by examining the rating forms for a student referred for excessive disruptiveness and extracting key descriptors from the items selected by his or her teachers, a qualitative behavioral profile for prescriptive use can be proposed.

Criterion validity was estimated by comparing teachers' ratings from the DSBS with non-teaching personnel's subjective ratings of the 32 students in the disruptive group. The emergence of a statistically significant correlation between teachers' and non-teaching personnel's ratings of the disruptive group is not surprising, as the group members were being evaluated on a common variable, disruptiveness. The weak nature of the correlation may be a statistical artifact resulting from the restricted range of the ratings by non-teaching personnel. The results suggest that disruptiveness ratings by non-teaching personnel

are poor predictors of DSBS ratings and that the weak correlation is a true indication of the relationship between the two ratings systems.

Since no objective criteria exist, it is not possible to state conclusively which rating system is superior in ranking students within the disruptive group. Except for borderline cases, this is not a significant limitation. Membership in the disruptive group would be prima facie evidence of the need for alternative education services, as the criterion has been set at two z-scores, or a disruptiveness rating above approximately 98% of the target school population.

The influence of environmental conditions on behavior may limit the period of time over which consistent behavioral measures can be obtained. Nevertheless, an instrument must demonstrate some degree of temporal consistency to be considered reliable enough for use in placement decisions. Since the basis for rating students on the DSBS is the teacher's past experience with a student, consistency over a 14-day period should be obtainable.

In the follow-up study to estimate temporal reliability, the test-retest correlations were significant for 11 of the 12 cases. For one student, while the scores from each rater varied significantly, the variances tended to cancel out and the overall effect was negligible. The

absolute values were at the very low end of the range and placement decisions would not be made on students with similar scores. Teachers reported some difficulty in rating students who displayed disruptive behavior very infrequently, as the behavior tended to be forgotten by the teachers with even a short passage of time. This is not considered a meaningful challenge to reliability or validity. The overall results of this reliability check suggest that the items are specific enough to describe observed disruptive behaviors and inclusive enough not to be affected by minor variations in rating performance.

# CHAPTER FIVE

# CONCLUSIONS, IMPLICATIONS, SUMMARY, AND RECOMMENDATIONS

## Conclusions

Based on the results of this study, the following conclusions were drawn:

1. Disruptive behaviors described by the items on the DSBS are accepted by educators as occurring in and disruptive to the school environment. These behaviors are recognizable by classroom teachers in a school setting.
2. In the opinion of experienced educators, the DSBS items are equitably distributed across the theoretical constructs which define disruptive behavior for the purposes of this study.
3. The DSBS demonstrates satisfactory criterion, convergent, and discriminant validity. The DSBS can correctly identify a criterion group of disruptive students, correctly classify individual students as disruptive, and with a high degree of accuracy exclude non-disruptive students from the disruptive group.
4. Except for ratings of students displaying minimal DSB, the DSBS provides consistent ratings over time.

## Implications

One implication of this study for existent theory (Sanson-Fisher & Mulligan, 1977; Weinrott, 1979) is support for the use of teachers as raters of classroom behavior. Likewise, the study results appear to confirm Epstein's (1980) suggestion that behaviors could be aggregated over situations and time to produce valid ratings. The study also demonstrated that disruptive school behavior is a definable category of behavior that can be measured empirically (Edelbrock, 1979; Gresham, 1982; O'Leary & Johnson, 1979).

The contention (Dickinson & Zellinger, 1980) that Likert scales remain viable because they are easy to understand and use was borne out. Support is offered for Messick's (1980) suggestion that tests of validity include interpretation, relevance, utility, and consequences of use. The research of Smith and Kendall (1963) in taking items from settings in which ratings will be made and involving raters in scale development was supported. The need for considering type, severity, and frequency as criteria for disruptive behavior also was supported (Robins, 1966).

Specifically, this study emphasized the necessity of using severity factors to determine the difference between normal and pathological levels of disruptiveness. In addition to supporting prior attempts to develop a theory

for defining and assessing disruptive behavior, this study specifically developed a set of underlying constructs that may prove useful in future studies of disruptive behavior. In proposing a method of establishing local norms, this study provides a procedure for evaluating behavior in a relevant context (Dickinson, 1978; Messick, 1980; Willems, 1975).

An important implication for future research is the provision in this study of an alternative to the approaches based on factor analysis, clinical descriptions, theoretical conceptions, and legal guidelines in describing disruptive behavior. Providing a set of underlying constructs for disruptive school behavior may help to focus future research activities. Emphasizing the importance of perceived severity of disruptive school behavior may influence future researchers to incorporate this factor into their designs. The DSBS might be used as a survey tool in research to identify and describe at-risk groups in the middle and junior high school population. This might also lead to a revising of gender and ethnic stereotypes of disruptive students.

Colleges of education could utilize the DSBS items as descriptors of typical disruptive school behavior and offer training in resolution techniques. School psychologists could benefit from being trained in assessing disruptive school behavior and making recommendations for

intervention strategies. Procedures used in validating the DSBS have suggested that descriptive statistics are adequate in establishing validity for research instruments to be used in school settings.

The DSBS provides a means of assessing needs for alternative programs and for in-service training. The DSBS provides a quantitative basis for admittance to and discharge from alternative education programs, allowing accountability, and perhaps attracting additional funding. The DSBS can provide prescriptive information for student program development. The ease of administration and scoring will permit school counselors to assess disruptive school behavior and make recommendations for interventions. The DSBS provides for assessment of disruptive school behavior by teachers, the persons in the best position to observe DSB.

## Summary

The purpose of the study was to develop and validate an instrument, the Disruptive Student Behavior Scale (DSBS). The DSBS is to be used to assess quantitatively the disruptive school behaviors of students referred for placement in either special education or alternative education programs. The study investigated previous attempts to define disruptive behavior; identification, assessment, and placement efforts directed toward

disruptive students; rating scale development procedures; research into the psychometric properties of rating scales; and the use by schools of results obtained from rating scales.

In an urban Florida middle school, consisting of grades six through nine, a norming group was chosen from a sample stratified by grade. A criterion, or disruptive, group was selected by nomination by non-teaching personnel. Classroom teachers rated both groups using the DSBS. A local norm for the school was calculated and the DSBS rating of each sample group student was compared to this norm. A rating of two or more standard deviation units, or z-scores, above the norm resulted in the student being classified as disruptive. A follow-up rating was performed after 14 days to test for temporal reliability.

Disruptive behaviors described by the items on the DSBS were accepted by educators as occurring in and disruptive to the school environment. The DSBS items were equitably distributed across the theoretical constructs which define disruptive behavior. The DSBS correctly identified a criterion group of disruptive students, correctly classified individual students as disruptive, and with a high degree of accuracy excluded non-disruptive students from the disruptive group. Except for ratings at the lowest end of the scale, the DSBS provided consistent ratings over time.

## Recommendations

There are several follow-up studies that need to be conducted. A study could be conducted to investigate the interactive effects on DSBS ratings of such demographic variables as race, gender, and socioeconomic status of both students and teachers. A study is needed to test specifically for the school setting the assumption (Epstein, 1980) of situational reliability, which permits the collapsing of scores across raters to improve reliability. The usefulness of the DSBS in identifying groups of students in need of remediation and prevention programs needs to be studied. An evaluation of DSBS ratings as a source of placement and discharge data for alternative education programs would be a useful addition to the literature.

APPENDIX A
CONSTRUCT DEVELOPMENT STUDY

Instructions for Matching Items with Categories

1. In each envelope there are 64 slips of paper (56 containing items and 8 blank) and a list of categories.

2. Please read each item and assign it to one or more of the categories from the list provided. Write the number of the category(s) in the upper left corner of the slip containing the item. Unless otherwise specified, think of these in terms of academic and conduct behaviors within the classroom or on school property.

3. If any items appear not to fit any category, please write "None" on the slip.

4. Reexamine each slip to verify your choice of category. For those items assigned to more than one category, please try again to determine <u>one</u> category which you think <u>best</u> represents the intent of the item. Indicate by <u>circling</u> your final choice.

5. Use the blank slips to write down any other categories or items which you think should be considered when determining a student's need for the program.

6. Replace all of the slips in the brown envelopes and return to me.

7. Please retain the list of categories and discuss these with your school contacts during the next two weeks. Indicate additions, deletions, suggestions, etc. on the list. Arrangements for getting this feedback from you will be made later.

8. Thank you all very much for your participation and assistance.

Bill Moses
3/6/81

Categories of Constructs Defining
Maladaptive Social Behavior

1. Disobedience (personal confrontation)
2. Aggression (verbal or physical, e.g., pushing, hitting)
3. Disruptiveness (in violation of rules)
4. Peer relationship problems (excluding fighting)
5. Law violations
6. Attendance violations
7. Fighting (initiates)
8. Academic irresponsibility (fails to complete assignments)
9. Destructiveness
10. Impulsiveness
11. Poor social relationships with school personnel
12. Denial of responsibility for actions
13. Inappropriate sexually-oriented behavior

TO: ACT Program Personnel

FROM: Bill Moses

SUBJECT: Analysis of Your Input Into Defining Maladaptive Social Behavior in the Schools

DATE: 3/18/81

Fifty-six items were to be assigned to thirteen categories. This assignment resulted in the following distribution:

1. Disobedience
   - defiance
   - defiance of adult authority
   - disrespect/defiance
   - ignoring the teacher
   - non-compliance
   - rebellion
   - resisting authority

2. Aggression
   - aggression
   - anger-defiance
   - hostility
   - inappropriate gross motor behaviors
   - inappropriate verbalizations
   - social aggression
   - using an object in interfering with another

3. Disruptiveness
   - classroom disturbance
   - conduct problems
   - distracting behaviors
   - general unruliness
   - making inappropriate noises
   - out of seat behavior

4. Peer relationships
   - inappropriate group behavior
   - interpersonal alienation
   - interpersonal relationships problems
   - peer relations difficulties
   - socialization difficulties
   - unfriendliness

5. Law violations
   none (see note below)

6. Attendance violations
   attendance/truancy

7. Fighting
   none (see Aggression)

8. Academic irresponsibility
   low productivity
   off-task behavior
   task avoidance

9. Destructiveness
   destructiveness

10. Impulsiveness
    impulsivity
    inability to delay gratification

11. Poor social relations with school personnel
    none (see Disobedience)

12. Denial of responsibility for actions
    externalization of blame
    irresponsibility
    unreliability

13. Inappropriate sexually oriented behavior
    none (see Peer relationships, Disobedience, or Disruptiveness, depending on the type and consequences of the behavior)

To be included under a category, an item had to be assigned to that category by at least one-half of the raters, or 8 people. The following items did not meet the criterion:

Failure to conform to social rules
Failure to function independently
Inappropriate physical contacts
Inattention
Inconsiderateness
Irrelevant responses
Lack of anger control
Lack of social control
Proneness to emotional upset
Seeking attention inappropriately
Sensation-seeking
Uncooperative
Unethical behavior

These items were either too ambiguous or they fit into too many categories, so were dropped. By consensus, six items were combined with other similar items, reducing the number of items to 50.

Note: Inspection of the list of items suggests few law violation items were included. Also, two items, drug-related behavior and stealing, were suggested as additional items. Recognizing the validity of these items requires continuing the category of law violations and ensuring that behaviors in this category are included in the final checklist.

Your ratings resulted in identifying the following ten constructs as defining maladaptive social behavior in the school. In order to operationalize these constructs, I will ask teachers to assign behaviorally stated items to each one. To assist them in making their assignments, I have included suggested definitions for each construct. These definitions also were developed from your input.

Final Constructs

1. Disobedience
   a. defying or challenging legitimate authority
   b. disrespect to teachers and staff

2. Disruptiveness
   a. classroom behavior which interferes with learning by others and/or the teachers' attempts to teach
   b. behavior outside the classroom which interferes with the orderly operation of the school

3. Impulsiveness
   a. reacting immediately without concern for consequences
   b. failure to delay gratification

4. Destructiveness
   a. intentional destroying of school or personal property
   b. careless, inattentive behavior resulting in the destruction of property

5. Aggression
   a. verbal or physical threats or attacks to a person
   b. passive acts that result in harm or hurt feelings

6. Academic irresponsibility
   a. failure to complete assignments
   b. task avoidance

7. Social/personal irresponsibility
   a. denial of responsibility
   b. blaming others
   c. unreliability

8. Ineffective interpersonal relationships
   a. alienation of peers resulting in avoidance by them
   b. lack of regard for others

9. Attendance violations
   a. truancy
   b. excessive tardiness or absences

10. Law violations
   a. any behavior which violates criminal laws, e.g., drug possession, arson

I would appreciate two kinds of additional feedback from you:

(1) Your opinion of the definitions of the constructs on this final list of ten.

(2) The names of teachers who would be willing to give some time to assigning specific behaviors, from a list, to these constructs. I will visit the schools to explain the project to each teacher.

Thank you again for your participation and assistance.

APPENDIX B
BEHAVIORS COLLECTED FROM DISCIPLINARY RECORDS

1. Leaving classroom without permission
2. Lying to the teacher
3. Inappropriate display of affection
4. Dress code violation--backless shoes
5. Denying responsibility for behavior
6. Sleeping in class
7. Smoking
8. Walking around auditorium
9. Possession of cigarettes
10. Abuse of hall pass privilege
11. Leaving school grounds with permission
12. Bringing vulgar materials to class
13. Having extraneous material in class
14. Walking away from other students in line
15. Shooting a bird at girl in class
16. Use of obscene language to other students
17. Kicking classroom door in anger
18. Skipping school
19. Skipping classes
20. Excessive tardiness that is unexcused
21. Possessing a stolen lock from a locker

22. Possession of concealed weapon (razor blade)
23. Stealing ($1.00)
24. Hitting a student in class
25. Throwing things at other students
26. Fighting in class
27. Threatened me "leave me alone or else"
28. Verbal threats to students
29. Failure to return report card
30. Interfering in another student's discipline
31. Refuse to complete sentences
32. Rude and disrespectful to teacher
33. Making obscene gestures to teacher
34. Verbally abusive to teacher
35. Silly and impudent acting
36. Defiance of authority--refused to follow direction
37. Running from teacher
38. Refused to give me the comic book he was looking at when class was reviewing homework
39. Refusal to obey teacher
40. Disturbing class, playing with electronic games
41. Passing wallets back and forth in class
42. Passing notes in class
43. Asking for pass from class excessively
44. Does not follow class standards and procedures
45. Out of seat without permission
46. Talking excessively loud

47. Arrogant (has to have last word) in classroom discussions with teacher

48. Constantly disrupting class by comments that add nothing to class discussion

49. Defiance--warned not to talk anymore or he would receive a referral--said "No, wanted written up"

50. Using vulgar language in class

51. Calling each other names

52. Disrupting testing

53. Inciting other students to misbehavior

54. Horse play in the halls

55. Misbehavior in lunch period (cafeteria)

56. Taking material off the bulletin board

57. Talking during fire drill

58. Singing (in class or hall)

59. Vandalism--tearing up another's clothes

60. Marking on desk

61. Writing obscenities on wall (malicious mischief)

62. Defiance--does nothing when class is assigned work--just sits

63. Failure to dress out for P.E.

64. Took another student's test paper and put their name on it

65. No books, paper, pencils

66. Writing notes to friends in class instead of doing assignments

## APPENDIX C
## ORAL INSTRUCTIONS FOR THE EDITING STUDY

"On these cards, one to a card, are written behaviors that are often considered disruptive when they occur in school. Please read each card carefully and answer the question: 'Would I recognize this behavior if I saw it occur?' If the answer is yes, please place the card in the pile to the left beside the 'yes' card. If not, place the card in the pile to the right beside the 'no' card. In a task such as this, your first reaction is often the most accurate, but you may take as much time as you need. If really undecided, place the card in the 'no' pile."

(After finishing) "Thank you. Now please look at the cards in the 'yes' pile again one more time, just to be sure. Feel free to make any other comments that occur to you. (After finishing) Thank you. Now, please look at each card in the 'no' pile and edit or rewrite the item in specific terms on one of these blank cards." (Clip the two cards together as S completes task). (After finishing) "Thank you very much for your cooperation."

APPENDIX D
ITEMS DEVELOPED FROM CONTENT VALIDATION STUDY

1. Has refused to follow my disciplinary instructions (e.g., to sit down, be quiet).

2. Has broken rules primarily to provoke a confrontation with me.

3. Has disrupted class by inappropriate activities (e.g., throwing objects, passing notes, hitting, walking around).

4. Has contributed to the destruction of personal or school property through carelessness, inattention, or neglect (e.g., using equipment improperly, failing to report unsafe conditions).

5. Has threatened to harm other students.

6. Has fought with other students.

7. Has been observed cheating on assignments (e.g., using notes, copying).

8. Has falsely denied being involved in a disruptive activity (e.g., breaking rules, cheating, fighting).

9. Has alienated other students (e.g., using obscene language or gestures, name calling).

10. Has left class without permission.

11. Has refused to accept disciplinary interventions from me (e.g., sentences, detention, referral to dean).

12. Has been seen outside of class being disruptive (e.g., running, shouting, singing, playing in the halls; throwing food in cafeteria).

13. Has displayed impatience (e.g., demanded immediate attention, refused to wait turn).

14. Has intentionally damaged or destroyed personal property (e.g., clothes, book, bicycle).

15. Has hit or thrown objects at other students.

16. Has avoided classroom assignments through passive activities (e.g., sleeping, not bringing materials).

17. Has broken rules primarily for personal convenience (e.g., smoking, eating, dress code violations).

18. Has blamed others for his or her actions.

19. Has been careless or malicious, resulting in harm to another person.

20. Has been present at homeroom but absent from class without permission.

21. Has interfered when I was disciplining another student.

22. Has encouraged other students to break school or classroom rules.

23. Has acted impulsively, without seeming to care about the consequences.

24. Has defaced school property (e.g., writing on walls or sidewalks, carving on desk).

25. Has threatened to harm me.

26. Has failed to complete classroom assignments satisfactorily from lack of interest (e.g., oral or written work, projects, dressing out).

27. Has been unreliable (e.g., lying, failing to return report cards or borrowed articles).

28. Has attempted to manipulate me (e.g., using friendship or flattery, exaggerating illness).

29. Has been late coming to class.

30. Has been observed in possession of stolen property (e.g., clothes, books, money, purse, lock).

31. Has treated me with rudeness and disrespect (e.g., back talk, verbal abuse, obscene gestures).

32. Has disrupted class by inappropriate talking (e.g., inappropriate questions, name calling, loudness, obscenity).

33. Has displayed lack of control (e.g., kicking door, hitting wall).

34. Has damaged or destroyed school property, making it unusable (e.g., breaking windows or equipment).

35. Has hit or thrown objects at me.

36. Has been unable to complete assignments satisfactorily because of the influence of non-prescription drugs or alcohol.

37. Has been observed in possession of illegal drugs or alcohol.

38. Has at school committed theft (e.g., from a locker) or robbery (from a person).

39. Has been observed in possession of a weapon or dangerous object.

40. Has been observed committing other major conduct violations (e.g., arson, battery, bomb threat, extortion, pulling fire alarm, shooting fireworks).

APPENDIX E
INSTRUCTIONS FOR CONTENT VALIDATION STUDY

Field Research: The Disruptive Student Behavior Scale

Dear Colleague:

Thank you for agreeing to participate in this research study. Please read all of the directions before beginning. _____ will answer any questions you may have.

1. Enclosed are two sets of slips. The slips labelled "Constructs" are categories of disruptive behavior. The slips labelled "Items" are statements of possible disruptive school behaviors.

2. Assign each item to one category of behavior if possible.

3. If an item obviously fits more than one category, write all category numbers on the slip but put it under the category it fits best.

4. If an item seems to fit no category, write "none" on the item slip. If the item can be edited to fit a category, please feel free to rewrite it and include it under the appropriate category.

5. Your comments on both the categories and items would be appreciated. Please write directly on the slips.

6. Clip the item slips to the appropriate category slips.

Please place all materials into the original envelope and return to _____'s mailbox as soon as possible. _____ will have for you an envelope containing your honorarium.

Again, thanks for assisting me in this project to develop an instrument for quantifying disruptive behavior. If you would like to receive a copy of the completed instrument, write your name below.

Sincerely,
/s/
Bill Moses
Date

APPENDIX F
THE DISRUPTIVE STUDENT BEHAVIOR SCALE (DSBS)

by

William Moses

Student's name_____ Grade___ Age___ Sex___

Rater's name_____ Subject_____

School_____Class Period_____

Date of rating_____

Directions: Please take a few minutes to recall your observations of this student during the current school year. Then respond to the statements on the following pages by placing an (X) through the appropriate number. Please return this form to_____ by_____

Example: In the current school year, this student:

1. Has been irresponsible in returning borrowed articles.

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

---

0 - None of the time = absolutely never

1 - Very infrequently = no more than once a month

2 - Sometimes = more than once a month

3 - Quite often = more than once a week

4 - Always = daily

(Blue Stock)

Key:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

In this school year, this student:

1. has refused to follow my disciplinary instructions (e.g., to sit down, be quiet).

   0 1 2 3 4

2. has threatened to harm other students.

   0 1 2 3 4

3. has disrupted class by inappropriate activities (e.g., throwing objects, passing notes, hitting, walking around).

   0 1 2 3 4

4. has contributed to the destruction of personal or school property through carelessness, inattention, or neglect (e.g., using equipment improperly, failing to report unsafe conditions).

   0 1 2 3 4

5. has falsely denied being involved in a disruptive activity (e.g., breaking rules, cheating, fighting).

   0 1 2 3 4

6. has been observed in possession of stolen property (e.g., clothes, books, money, purse, lock).

   0 1 2 3 4

7. has attempted to manipulate me (e.g., using friendship or flattery, exaggerating illness).

   0 1 2 3 4

8. has displayed lack of impulse control (e.g., kicking door, shouting out).

   0 1 2 3 4

Page Score

Key:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

---

In this school year, this student:

9. has been late coming to class.

0 1 2 3 4

10. has failed to complete classroom assignments satisfactorily from lack of interest (e.g., oral or written work, projects, dressing out).

0 1 2 3 4

11. has defaced school property (e.g., writing on walls or sidewalks, carving on desk).

0 1 2 3 4

12. has been deliberately rude or impolite to other students.

0 1 2 3 4

13. has broken rules resulting in a confrontation with me.

0 1 2 3 4

14. has been observed in possession of illegal drugs or alcohol.

0 1 2 3 4

15. has displayed impatience (e.g., demanded immediate attention, refused to wait turn).

0 1 2 3 4

16. has left class without permission.

0 1 2 3 4

17. has hit or thrown objects at other students.

0 1 2 3 4

Page Score

Key:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

---

In this school year, this student:

18. has avoided classroom assignments through passive activities (e.g., sleeping, not bringing materials, being under the influence of drugs).

    0 1 2 3 4

19. has been seen outside of class being disruptive (e.g., running, shouting, singing, playing in the halls; throwing food in cafeteria).

    0 1 2 3 4

20. has broken rules primarily for personal convenience (e.g., smoking, eating, dress code violations).

    0 1 2 3 4

21. has been present at homeroom but absent from class without permission.

    0 1 2 3 4

22. has acted impulsively, without regard for the consequences.

    0 1 2 3 4

23. has threatened to harm me.

    0 1 2 3 4

24. has at school committed theft (e.g., from a locker) or robbery (from a person).

    0 1 2 3 4

25. has cheated on assignments (e.g., using notes, copying).

    0 1 2 3 4

Key:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

In this school year, this student:

26. has refused to accept disciplinary interventions from me (e.g., sentences, detention, referral to dean).

0 1 2 3 4

27. has interfered when I was disciplining another student.

0 1 2 3 4

28. has alienated other students (e.g., using obscene language or gestures, name calling).

0 1 2 3 4

29. has blamed others for his or her actions.

0 1 2 3 4

30. has damaged or destroyed school property, making it unusable (e.g., breaking windows or equipment).

0 1 2 3 4

31. has treated me with rudeness and disrespect (e.g., back talk, verbal abuse, obscene gestures).

0 1 2 3 4

32. has hit or thrown objects at me.

0 1 2 3 4

33. has been observed in possession of a weapon or dangerous object.

0 1 2 3 4

Page Score

Key:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| None of the time | Very infrequently | Sometimes | Quite often | Always |

In this school year, this student:

34. has disrupted class by inappropriate talking (e.g., inappropriate questions, name calling, loudness, obscenity).

    0 1 2 3 4

35. has been unreliable (e.g., lying, failing to return report cards or borrowed articles).

    0 1 2 3 4

36. has fought with other students.

    0 1 2 3 4

37. has intentionally damaged or destroyed personal property (e.g., clothes, book, bicycle).

    0 1 2 3 4

38. has been observed committing other major conduct violations (e.g., arson, battery, bomb threat, extortion, pulling fire alarm, shooting fireworks).

    0 1 2 3 4

Page Score _____ Form Score _____

## APPENDIX G
## INSTRUCTIONS FOR SEVERITY FACTOR STUDY

Dear Colleague:

Thank you for your help with this study to develop an instrument for measuring disruptive school behavior.

Enclosed are 40 statements of disruptive behaviors which teachers have observed in middle and junior high schools. The purpose of this portion of the project is to determine a severity rating for each of the 40 statements. Severity is defined as "a prediction, stated quantitatively, of the potentially detrimental consequences a disruptive behavior would likely have for a student."

One system of identifying these consequences is by completing the following statement:

This behavior would probably be detrimental to the student's

1. relations with school personnel
2. relations with peers
3. future vocational opportunities
4. mental development
5. emotional development
6. physical development
7. learning of course material
8. course grade
9. remaining in school

To assign a severity factor to the 40 behavioral items, please write below each item the number(s), 1-9, of the consequences that are most likely to result from that behavior. Please work independently of the other judges.

After completing this task, please return all materials in the envelope provided. Thank you again for this help.

Sincerely,
/s/
Bill Moses

P.S. An honorarium is enclosed to partially compensate you for your time.

## APPENDIX H
## SCORING TEMPLATE FOR THE DSBS

Directions:

1. Align the arrow on the template (green) with the arrow on the rating scale (blue).

2. Read the score above the rater's mark (X) and record on the rating scale (blue) as indicated.

3. Total the student's scores for each page and for the completed form.

4. Record the form score onto the summary of Teacher Ratings (yellow).

(Trimmed to Line)

SCORING TEMPLATE FOR THE DSBS

| | | | | | |
|---|---|---|---|---|---|
| 1. | 0 | 2 | 4 | 6 | 8 |
| 2. | 0 | 1 | 2 | 3 | 4 |
| 3. | 0 | 2 | 4 | 6 | 8 |
| 4. | 0 | 1 | 2 | 3 | 4 |
| 5. | 0 | 1 | 2 | 3 | 4 |
| 6. | 0 | 2 | 4 | 6 | 8 |
| 7. | 0 | 2 | 4 | 6 | 8 |
| 8. | 0 | 2 | 4 | 6 | 8 |

(Green Stock)

| | | | | |
|---|---|---|---|---|
| 9. 0 | 1 | 2 | 3 | 4 |
| 10. 0 | 2 | 4 | 6 | 8 |
| 11. 0 | 1 | 2 | 3 | 4 |
| 12. 0 | 2 | 4 | 6 | 8 |
| 13. 0 | 1 | 2 | 3 | 4 |
| 14. 0 | 3 | 6 | 9 | 12 |
| 15. 0 | 2 | 4 | 6 | 8 |
| 16. 0 | 2 | 4 | 6 | 8 |
| 17. 0 | 2 | 4 | 6 | 8 |

(Green Stock)

| | | | | |
|---|---|---|---|---|
| 18. 0 | 2 | 4 | 6 | 8 |
| 19. 0 | 1 | 2 | 3 | 4 |
| 20. 0 | 1 | 2 | 3 | 4 |
| 21. 0 | 2 | 4 | 6 | 8 |
| 22. 0 | 2 | 4 | 6 | 8 |
| 23. 0 | 2 | 4 | 6 | 8 |
| 24. 0 | 3 | 6 | 9 | 12 |
| 25. 0 | 2 | 4 | 6 | 8 |

(Green Stock)

| | | | | |
|---|---|---|---|---|
| 26. 0 | 1 | 2 | 3 | 4 |
| 27. 0 | 1 | 2 | 3 | 4 |
| 28. 0 | 1 | 2 | 3 | 4 |
| 29. 0 | 2 | 4 | 6 | 8 |
| 30. 0 | 2 | 4 | 6 | 8 |
| 31. 0 | 1 | 2 | 3 | 4 |
| 32. 0 | 2 | 4 | 6 | 8 |
| 33. 0 | 3 | 6 | 9 | 12 |

(Green Stock)

| | | | | |
|---|---|---|---|---|
| 34. 0 | 3 | 6 | 9 | 12 |
| 35. 0 | 1 | 2 | 3 | 4 |
| 36. 0 | 2 | 4 | 6 | 8 |
| 37. 0 | 2 | 4 | 6 | 8 |
| 38. 0 | 3 | 6 | 9 | 12 |

(Green Stock)

APPENDIX I
SUMMARY OF TEACHER RATINGS ON THE DSBS

Directions:

1. Enter below the information from each DSBS form (blue).
2. Add the form scores to obtain the student's total score.
3. Divide the total score by the number of raters to obtain the student's disruptive school behavior (DSB) rating.
4. Enter your local DSB norm and SD from the norming study.
5. Complete the calculations to obtain the deviation (in standard deviation units) from your local norm.

Student's name_____ Grade_____ Age_____ Sex_____

School_____ Evaluator_____ Title_____

| Form score | Rater's name | Subject | Period |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| _____ | ÷ | _____ | = | \|_____\| | - | _____ | = | _____ | ÷ | _____ | = | \|_____\| |
| Total score | | Nbr of raters | | Student's DSBS rating | | Local DSBS norm | | Deviation from local norm | | Local SD | | SD's from local norm |

(Yellow Stock)

APPENDIX J
PRESCRIPTIVE PROFILE WORKSHEET FOR THE DSBS

Student_____ Grade___ Age___ Sex___ School_____

Evaluator_____ Title_____

Directions:

1. Using the DSBS form (blue) and the scoring template (green), enter below each teacher's rating for each scale item. Do not use raw scores from the DSBS. Note the different sequence of the item numbers below.

2. Add the ratings across by item and down by rater.

3. Add the Item Totals down and the Rater Totals across. The two totals must agree.

4. Add the item totals for each profile category and enter in the appropriate box. Profile definitions are provided for guidance in preparing qualitative behavioral descriptions and suggesting prescriptive interventions. They are not intended for use in placement decisions.

| DSBS Item Number | | | | | | | Item Totals |
|---|---|---|---|---|---|---|---|
| 1. | _____ | _____ | _____ | _____ | _____ | _____ | _____ |
| 13. | _____ | _____ | _____ | _____ | _____ | _____ | _____ |
| 26. | _____ | _____ | _____ | _____ | _____ | _____ | _____ |
| 31. | _____ | _____ | _____ | _____ | _____ | _____ | _____ |
| Totals | (a)_____ | (b)_____ | (c)_____ | (d)_____ | (e)_____ | (f)_____ | \|_____\| Disobedience Rating |

(Beige Stock)

| DSBS Item Number | | | | | | | Item Totals | |
|---|---|---|---|---|---|---|---|---|
| 3. | | | | | | | | |
| 19. | | | | | | | | |
| 27. | | | | | | | | |
| 34. | | | | | | | | |
| Totals | (a) | (b) | (c) | (d) | (e) | (f) | | Disruptiveness Rating |
| 8. | | | | | | | | |
| 15. | | | | | | | | |
| 22. | | | | | | | | |
| Totals | (a) | (b) | (c) | (d) | (e) | (f) | | Impulsiveness Rating |
| 4. | | | | | | | | |
| 11. | | | | | | | | |
| 30. | | | | | | | | |
| 37. | | | | | | | | |
| Totals | (a) | (b) | (c) | (d) | (e) | (f) | | Destructiveness Rating |
| 2. | | | | | | | | |
| 17. | | | | | | | | |
| 23. | | | | | | | | |
| 32. | | | | | | | | |
| 38. | | | | | | | | |
| Totals | (a) | (b) | (c) | (d) | (e) | (f) | | Aggression Rating |

(Beige Stock)

| DSBS Item Number | | | | | | | Item Totals | |
|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | |
| 18. | | | | | | | | |
| 25. | | | | | | | | |
| Totals (a) | (b) | (c) | (d) | (e) | (f) | | | Academic Irresponsibility Rating |
| 5. | | | | | | | | |
| 20. | | | | | | | | |
| 29. | | | | | | | | |
| 35. | | | | | | | | |
| Totals (a) | (b) | (c) | (d) | (e) | (f) | | | Personal Irresponsibility Rating |
| 7. | | | | | | | | |
| 12. | | | | | | | | |
| 28. | | | | | | | | |
| Totals (a) | (b) | (c) | (d) | (e) | (f) | | | Ineffective Interpersonal Relationships Rating |
| 9. | | | | | | | | |
| 16. | | | | | | | | |
| 21. | | | | | | | | |
| Totals (a) | (b) | (c) | (d) | (e) | (f) | | | Attendance Violations Rating |

(Beige Stock)

| DSBS Item Number | | | | | | | Item Totals | |
|---|---|---|---|---|---|---|---|---|
| 6. | | | | | | | | |
| 14. | | | | | | | | |
| 24. | | | | | | | | |
| 33. | | | | | | | | |
| 36. | | | | | | | | |
| | | | | | | | | Law Violations |
| Totals | (a) | (b) | (c) | (d) | (e) | (f) | | Rating |
| Rater Totals | (a) | (b) | (c) | (d) | (e) | (f) | | DSBS Rating |

(Beige Stock)

Definitions of Profile Categories from the DSBS

These profile descriptions are provided for guidance in preparing qualitative behavioral descriptions and suggesting prescriptive interventions. These descriptions are not intended for use in placement decisions. DSBS item numbers are in ( ).

1. Disobedience
   a. defying or challenging legitimate authority (refusing legitimate request) (1, 26)
   b. disrespect to teachers and staff (31)
   c. intentionally breaking rules to show defiance (13)

2. Disruptiveness
   a. classroom behavior which interferes with learning by others and/or the teacher's attempts to teach (3, 27, 34)
   b. behavior outside the classroom which interferes with the orderly operation of the school (19)

3. Impulsiveness
   a. reacting immediately without concern for consequences (8, 22)
   b. failure to delay gratification (15)

4. Destructiveness
   a. intentional destroying of school or personal property (11, 30, 37)
   b. careless, inattentive behavior resulting in the destruction of property (4)

5. Aggression
   a. verbal or physical threats or attacks to a person (2, 17, 23, 32, 38)

6. Academic irresponsibility
   a. failure to complete assignments satisfactorily (10)
   b. task avoidance (18)
   c. cheating (25)

7. Personal irresponsibility
   a. denial of his/her involvement in an activity (5)
   b. blaming others for his/her actions (7)
   c. unreliability (35)
   d. intentionally breaking rules for personal convenience (20)

8. Ineffective interpersonal relationships
   a. alienation of peers resulting in avoidance by them (28)

b. lack of regard for others (12)
c. manipulation of others (7)

9. Attendance violations
   a. leaving class (16)
   b. excessive tardiness or absences (9, 21)

10. Law violations
   a. any behavior which violates criminal laws, e.g., drug possession, arson (6, 14, 24, 33, 36)

## APPENDIX K
## INSTRUCTIONS FOR THE PILOT STUDY

Thank you for agreeing to participate in this pilot study of an instrument to measure disruptive behavior. The latest Gallup education poll, the 16th, shows that the major public concern about education continues to be disruption in the schools. Education association (for example, the NEA) polls show that many teachers feel the same way. This instrument we are field testing here may help identify students who need special assistance in order to develop appropriate school behavior.

A number of students at this school have been randomly chosen as representative of all the students. The behaviors of these students are being measured to determine the usual behavior patterns here. Another group of students has been chosen on the basis of referrals to the deans for disciplinary reasons. Their behaviors are being measured to see how different they are from the other, or average, group.

Each of you has received a rating form, the blue form, for each period you teach one of the students selected for either group. Each form contains _(number)_ items descriptive of disruptive behavior. You are being asked to rate each student only on those behaviors included on the instrument and only those you have actually observed during this school year. Please follow along as I read the directions on the blue form. If at any time there is a question, please make a note on the form and I will answer all your questions after reading through the directions.

(Read the directions and the example for this scale.)

Are there questions? (Answer any.)

_____(Name)_____ is your school coordinator for this study. He/she will answer any questions that may arise later and also collect the completed forms in his/her mailbox. Please replace the forms in the original envelope and return to _____(Name)_____.

Today is __(day)__, the _(date)_. What is a reasonable date for returning the rating forms? (Determine reasonable deadline and get majority concurrence.)

Now, please look at the last page in your packet. I am very interested in your reactions to this instrument and would appreciate, after you complete your ratings, if you would answer the few questions on this survey form.

Remember to rate each student on the basis of only the behaviors included on the instrument.

Thank you all very much for your valuable contributions. Instruments have to be field-tested and you are the people best qualified to do this.

## APPENDIX L
## RATER'S EVALUATION OF THE DSBS

Please use the following guidelines for your answers:

0 - No
1 - Somewhat
2 - Yes

1. The directions could be followed easily.

   0 1 2

2. The length of time to complete the ratings was reasonable.

   0 1 2

3. The items seemed related to disruptive behavior.

   0 1 2

4. The rating task was interesting.

   0 1 2

5. The items were understandable.

   0 1 2

6. The information needed to respond to the items was known to you.

   0 1 2

APPENDIX M
ASSIGNMENT OF CONSTRUCTS BY ITEM NUMBER

| DSBS Item Numbers | Construct Number |
|---|---|
| 1 | 1 |
| 2 | 5 |
| 3 | 2 |
| 4 | 4 |
| 5 | 7 |
| 6 | 10 |
| 7 | 8 |
| 8 | 3 |
| 9 | 9 |
| 10 | 6 |
| 11 | 4 |
| 12 | 8 |
| 13 | 1 |
| 14 | 10 |
| 15 | 3 |
| 16 | 9 |
| 17 | 5 |
| 18 | 6 |
| 19 | 2 |
| 20 | 7 |
| 21 | 9 |
| 22 | 3 |
| 23 | 5 |
| 24 | 10 |
| 25 | 6 |
| 26 | 1 |
| 27 | 2 |
| 28 | 8 |
| 29 | 7 |
| 30 | 4 |
| 31 | 1 |
| 32 | 5 |
| 33 | 10 |
| 34 | 2 |
| 35 | 7 |
| 36 | 10 |
| 37 | 4 |
| 38 | 5 |

## REFERENCES


Abikoff, H., Gittelman, R., & Klein, D. (1980). Classroom observation code for hyperactive children: A replication of validity. Journal of Consulting and Clinical Psychology, 48(5), 555-565.

Achenbach, T. (1978). The child behavior profile: I boys aged 6-11. Journal of Consulting and Clinical Psychology, 46(3), 478-488.

Achenbach, T., & Edelbrock, C. (1978). The classification of child pychopathology. Psychological Bulletin, 85 1275-1301.

Albaum, G., Best, R., & Hawkins, D. (1981). Continuous vs. discrete semantic differential rating scales. Psychological Reports, 49, 83-86.

Algozzine, B. (1977). The emotionally disturbed child: Disturbed or disturbing? Journal of Abnormal Child Psychology, 5, 205-211.

Algozzine, B. (1979). The disturbing child: A validation report. Minneapolis, Minn.: University of Minnesota, Institute for Research on Learning Disabilities.

Algozzine, B. (1980). The disturbing child: A matter of opinion. Behavioral Disorders, 5(2), 112-115.

Algozzine, B., & Ysseldyke, J. (1981). Special education services for normal children: Better safe than sorry? Exceptional Children, 48(3), 238-243.

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

American Psychological Association. (1966). Standards for educational and psychological tests and manuals. Washington, D.C.: Author.

Amos, K. (1980). Competency testing: Will the LD student be included? Exceptional Children, 47(3), 194-197.

Arnove, R., & Strout, T. (1978). Alternative schools and cultural pluralism: Promise and reality. Educational Research Quarterly, 2(4), 74-95.

Bailey, D., Bender, W., & Montgomery, D. (1983). Comparison of teacher, peer, and self-ratings of classroom and social behavior of adolescents. Behavior Disorders, 8(3), 153-159.

Balow, H. (1979). Definitional and prevalence problems in behavior disorders of children. School Psychology Digest, 8, 348-354.

Bardo, J., & Yeager, S. (1982). Note on reliability of fixed-response formats. Perceptual and Motor Skills, 54, 1163-1166.

Bardo, J., Yeager, S., & Klingsporn, M. (1982). A preliminary assessment of format-specific central tendency and leniency error in summated rating scales. Perceptual and Motor Skills, 54, 227-234.

Bass, B., Cascio, W., & O'Conner, E. (1974). Magnitude estimation of expressions of frequency and amount. Journal of Applied Psychology, 34, 313-320.

Bayh, B. (1975). Our nation's schools, a report card: "A" in school violence and vandalism. Washington, D.C.: Senate Committee on the Judiciary, Subcommittee to Investigate Juvenile Delinquency.

Beatty, R., Schneier, C., & Beatty, J. (1977). An empirical investigation of perceptions of ratee behavior frequency and ratee behavior change using Behavioral Expectation Scales (BES). Personnel Psychology, 30, 647-658.

Behar, D., & Stewart, M. (1984). Aggressive conduct disorder: The influence of social class, sex, and age on the clinical picture. Journal of Child Psychology and Psychiatry, 25(1), 119-124.

Beitchman, J., & Raman, S. (1979). The assessment of childhood psychopathology: The construction of a new self-report psychiatric rating scale for children. Multivariate Experimental Clinical Research, 4(192), 23-31.

Bell, T. (May 26, 1984). Education chief worried by dropout rate. St. Petersburg Times, p. A-6.

Beltramini, R. (1982). Rating-scale variations and discriminability. Psychological Reports, 50, 299-302.

Bennett, C., & Harris, J. (1982). Suspensions and expulsions of male and Black students. Urban Education, 16(4), 399-423.

Benson, J., & Clark, F. (1982). A guide for instrument development and validation. The American Journal of Occupational Therapy, 36(12), 789-800.

Bernardin, H., LaShells, M., Smith, P., & Alvares, K. (1976). Behavioral expectation scales. Journal of Applied Psychology, 61(1), 75-79.

Bernardin, H., & Pence, E. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bernardin, H., & Smith, P. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS). Journal of Applied Psychology, 66(4), 158-463.

Blanz, F., & Ghiselli, E. (1972). The mixed standard scale: A new rating system. Personnel Psychology, 25, 185-199.

Bower, E. (1958). A process for early identification of emotionally disturbed children. Bulletin of the California State Department of Education, 27, 1-111.

Bower, E. (1982). Defining emotional disturbances: Public policy and research. Psychology in the Schools, 19, 55-60.

Bower, E., & Lambert, N. (1971). In-school screening of children with emotional handicaps. In A. Long, O. Morse, & J. Newman (Eds.), Conflict in the classroom, (pp.142-148). Belmont, Cal.: Wadsworth.

Bower, R. (1969). Early identification of emotionally handicapped children in schools (2nd. ed.). Springfield, Ill.: Charles C. Thomas.

Broadbent, D., Cooper, P., Fitzgerald, P., & Parkes, K. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology, 21, 1-16.

Brown, L., & Hammill, D. (1978). The behavior rating profile. Austin, Tex.: Pro-ed.

Bruvold, W. (1969). Category and successive intervals scales for rating statements and stimulus objects. Journal of Experimental Psychology, 81, 230-234.

Buros, O. (Ed.). (1978). The eighth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press.

California Department of Education. (1973). Conflict and violence in California's high schools. Sacramento: Author.

Caliste, E. (1979). Students' adjustment from open to structured schools. Contemporary Education, 50(3), 138-151.

Camp, W. (1980). Educator perceptions of student discipline. Paper presented at the annual meeting of the American Educational Research Association (Boston, Mass., April 7-11) (ERIC Document Reproduction Service No. Ed 187 006).

Camp, W. (1981). Identifying problem behaviors: Indiana teachers look at student discipline. National Association of Secondary School Principals Bulletin, 65(441), 45-48.

Cattell, R. (1978). The scientific use of factor analysis in behavioral and life sciences. New York: Plenum.

Chase, C. (1969). Often is where you find it. American Psychologist, 24, 1043.

Conners, C. (1969). A teacher rating scale for use in drug studies with children. American Journal of Psychiatry, 126, 884-888.

Cooper, W. (1981). Ubiquitous halo. Psychological Bulletin, 90(2), 218-244.

Cooper, W. (1983). Internal homogeneity, descriptiveness, and halo. Personnel Psychology, 36, 489-502.

Cowen, E., Dorr, D., Clarfield, S., Kreling, B., McWilliams, S., Pokracki, F., Pratt, D., Terrell, D., & Wilson, A. (1973). The AML: A quick-screening device for early identification of school maladaptation. American Journal of Community Psychology, 1(1), 12-35.

Cronbach, L. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.

Cross, H., & Kohl, M. (1978). Predelinquent behavior in males: Perspectives and suggestions. Journal of Research and Development in Education, 11(2), 34-41.

Cullinan, D. (1975). Behavior modification with disruptive, delinquent, and otherwise deviant adolescents. Wilkes-Barre, Pa.: Wilkes College, Educational Development Center.

Davis, R., Butler, N., & Goldstein, N. (1972). From birth to seven: Second report of the National Child Development Study. London: Longman and National Children's Bureau.

Dayton, C. (1967). Technical manual: Pupil classroom behavior scale. College Park, Md.: University of Maryland.

Dean, R. (1980). Teachers as raters of aberrant behavior. Journal of School Psychology, 18(4), 354-359.

Deibert, J., & Hoy, W. (1977). "Custodial" high schools and self-actualization of students. Educational Research Quarterly, 2(2), 24-31.

Deno, S. (1979). A direct observation approach to measuring classroom behavior: Procedures and application. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Department of Health, Education, and Welfare (August 23, 1977). Implementation of part b., P.L. 94-142. Federal register, Part II, Section 121a.5.

Dickinson, D. (1978). Direct assessment of behavioral and emotional problems. Psychology in the Schools, 15(4), 472-477.

Dickinson, T., & Zellinger, P. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. Journal of Applied Psychology, 65(2), 147-154.

DiPrete, T. (1981). Discipline, order and student behavior in American high schools. Washington, D.C.: National Center for Education Statistics.

DiStefano, M., Pryer, M., & Erffmeyer, R. (1983). Application of content validity methods to the

development of a job-related performance rating criterion. Personnel Psychology, 36, 621-631.

Dreger, R. (1980). The initial standardization of the Adolescent Behavioral Classification Project Instrument. Journal of Abnormal Child Psychology, 8(3), 297-322.

Duke, D. (1978). Looking at the school as a rule-governed organization. Journal of Research and Development in Education, 11(4), 116-126.

Dunlap, W., & Dillard, J. (1980). Problem behavior from the principal's perception: A new view. Behavioral Disorders, 5(2), 108-111.

Dunnett, S., Koun, S., & Barber, P. (1981). Social desirability in the Eysenck Personality Inventory. British Journal of Psychology, 72, 19-26.

Dunnette, M. (1963). A modified model for test validation and selection research. Journal of Applied Psychology, 47, 317-323.

Duval County Public Schools. (1980). Code of student conduct. Jacksonville, Fla.: Author.

Duval County School Board. (1979). Project ACT: Accountability in citizenship training manual. Jacksonville, Fla.: Author.

Eaves, R. (1975). Teacher race, student race, and the Behavior Problem Checklist. Journal of Abnormal Child Psychology, 3(1), 1-9.

Edelbrock, C. (1979). Empirical classification of children's behavior disorders: Progress based on parent and teacher ratings. School Psychology Digest, 8(4), 355-369.

Edelbrock, C., & Achenbach, T. (1978). Child behavior profile patterns of children referred for clinical services. Bethesda, Md.: National Institutes of Mental Health. (ERIC Document Reproduction Service No. ED 768 259)

Edelbrock, C., & Achenbach, T. (1980). A typology of Child Behavior Profile patterns: Distribution and correlates for disturbed children aged 6-16. Journal of Abnormal Child Psychology, 8(4), 441-470.

Edwards, A. (1957). Techniques of attitude scale construction. New York: Prentice-Hall.

Edwards, A., & Thurstone, L. (1952). An internal consistency check for scale values determined by the method of successive intervals. Psychometrika, 17, 169-180.

Edwards, D., Roundtree, G., Kent, S., & Parker, J. (1981). An investigation of truancy and subsequent antisocial behavior. Corrective and Social Psychiatry and Journal of Behavior Technology Methods and Therapy, 28(1), 5-9.

Epstein, M., Cullinan, D., & Rosemier, R. (1983). Behavior problems of behaviorally disordered and normal adolescents. Behavior Disorders, 8(3), 171-175.

Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. American Psychologist, 35(9), 790-806.

Erickson, K. (1963). The wayward Puritans. New York: Wiley.

Erickson, M. (1974). The effects of initial information and observational set upon social perception (Doctoral dissertation, University of Oregon, 1973). Dissertation Abstracts International, 34, 4659B.

Evertson, C., Anderson, L., & Brophy, J. (1979). Texas junior high school study (R&D Rep. No. 4061). Austin, Tex.: University of Texas, Research and Development Center for Teacher Education.

Evertson, C., & Veldman, D. (1981). Changes over time in process measures of classroom behavior. Journal of Educational Psychology, 73(2), 156-163.

Faretra, G. (1981). A profile of aggression from adolescence to adulthood: An 18-year follow-up of psychiatrically disturbed and violent adolescents. American Journal of Orthopsychiatry, 51(3), 439-453.

Fay, C., & Latham, G. (1982). Effects of training and rating scales on rating errors. Personnel Psychology, 35, 105-116.

Feldhusen, J. (1978). Behavior problems in secondary schools: Final report. Washington, D.C.: Department of Health, Education, and Welfare; National Institute of Education.

Filipczak, J. (1978). Social skills training in the community with behaviorally disruptive youth. Silver Spring, Md.: Institute for Behavioral Research. (ERIC Document Reproduction Service No. ED 170 632)

Finley, D., Osburn, H., Dubin, J., & Jeanneret, P. (1977). Behaviorally based rating scales: Effects of specific anchors and disguised scale continua. Personnel Psychology, 30, 659-669.

Florida Department of Education. (1975). District procedures for providing special education for exceptional students. Tallahassee, Fla.: Author.

Florida Department of Education. (1983). Report on public schools. Tallahassee, Fla.: Author.

Florida Department of Education. (1985). A resource manual for the development and evaluation of special programs for exceptional students (Vol. 1-B). Tallahassee, Fla.: Author.

Fogel, L., & Nelson, R. (1983). The effects of special education labels on teachers' behavioral observations, checklist scores, and grading of academic work. Journal of School Psychology, 21, 241-251.

Foley, M. (1982). Punishment and the disruptive student. Contemporary Education, 53(2), 92-96.

Forness, S. (1983). Diagnostic schooling for children or adolescents with behavioral disorders. Behavioral Disorders, 8(3), 176-187.

Forness, S., & Cantwell, D. (1982). DSM-III psychiatric diagnoses and special education categories. Journal of Special Education, 16(1), 49-63.

Forness, S., Sinclair, E., & Russell, A. (1984). Serving children with emotional or behavior disorders. American Journal of Orthopsychiatry, 54(1), 22-32.

Fraser, C. (1980). Measurement in psychology. British Journal of Psychology, 71, 23-34.

Fremont, T., & Wallbrown, F. (1979). Types of behavior problems that may be encountered in the classroom. Journal of Education, 161(2), 5-24.

Gajar, A., & Hale, R. (1982). Factor analysis of the Quay-Peterson Behavior Problem Checklist across racially different exceptional children. The Journal of Psychology, 112, 287-293.

Gallup, G. (1984). The 16th annual Gallup poll of the public's attitudes toward the public schools. Phi Delta Kappan, 66(1), 23-38.

Garbarino, J. (1980). Some thoughts on school size and its effectson adolescent development. Journal of Youth and Adolescence, 9(1), 19-29.

Garibaldi, A. (Ed.) (1979). In-school alternatives to suspension conference report. Washington, D.C.: Department of Health, Education, and Welfare; National Institute of Education.

Gaynor, J., & Gaynor, M. (1976). The Delaware function rater: A method of quantifying classroom behavior. Wilmington, Delaware: Author. (ERIC Document Reproduction Service No. ED 133-983)

Geiger, K., & Turiel, E. (1983). Disruptive school behavior and concepts of social convention in early adolescence. Journal of Educational Psychology, 75(5), 677-685.

Gerard, R. (1970). Institutional innovations in juvenile corrections. Federal Probation, 34, 37-44.

Ghodsian, M., Fogelman, K., Lambert, L., & Tibbenham, A. (1980). Changes in behaviour ratings of a national sample of children. British Journal of Social and Clinical Psychology, 19, 247-256.

Goldsmith, A. (1982). Codes of discipline: Developments, dimensions, directions. Education and Urban Society, 14(2), 185-195.

Goldstein, A., Apter, S., & Harootunian, B. (1984). School violence. Englewood Cliffs, N.J.: Prentice-Hall.

Goocher, B. (1965). Effects of attitude and experience on the selection of frequency adverbs. Journal of Verbal Learning and Verbal Behavior, 4, 193-195.

Gorsuch, R. (1974). Factor analysis. Philadelphia: Saunders.

Goslin, D. (1969). Guidelines for the collection, maintenance and dissemination of pupil records. Troy, N.Y.: Russell Sage Foundation.

Governor's Task Force on Disrupted Youth. (1973). Phase I report. Tallahassee, Fla.: State of Florida.

Governor's Task Force on Disrupted Youth. (1974). Phase II report. Tallahassee, Fla.: State of Florida.

Goyette, C., Conners, C., & Ulrich, R. (1978). Normative data on revised Conners' Parent and Teacher Rating Scales. Journal of Abnormal Child Psychology, 6, 221-236.

Graham, S. (1981). Predictive and concurrent validity of the Jesness Inventory Asocial Index: When does a delinquent become a delinquent? Journal of Consulting and Clinical Psychology, 49(5), 740-742.

Green, R., Bigelow, L., O'Brien, P., Stahl, D., & Wyatt, R. (1977). The inpatient behavioral rating scale. Psychological Reports, 40, 543-549

Green, R., & Brydon, J. (1975). Investing in youth: An approachto discipline in urban schools. In National Education Association (Ed.), Discipline and learning, (pp. 107-114). Washington, D.C.: NEA.

Greenwood, C., Walker, H., & Hops, H. (1977). Issues in social interaction/withdrawal assessment. Exceptional Children, 43, 490-498.

Gresham, F. (1982). A model for the behavioral assessment of behavior disorders in children. Journal of School Psychology, 20(2), 131-144.

Grise, P. (1980). Florida's minimum competency testing program for handicapped students. Exceptional Children, 47(3), 186-191.

Grosek, R. (1979). Problem Behavior Rating Scale. Binghampton, N.Y.: Broome Developmental Center. (ERIC Document Reproduction Service No. EC 113 785)

Guion, R. (1977). Content validity: Three years of talk --what's the action? Public Personnel Management, 6, 407-414.

Guttman, L. (1944). A basis for scaling quantitative data. American Sociological Review, 9, 139-150.

Guttman, L. (1945). Questions and answers about scale analysis. Report D-2. Washington, D.C.: Research Branch, Information and Education Division, Army Service Forces.

Guttman, L. (1947a). On Festinger's evaluation of scale analysis. Psychological Bulletin, 44, 451-465.

Guttman, L. (1947b). The Cornell technique for scale and intensity analysis. Educational and Psychological Measurement, 1, 247-280.

Hale, R., & Zuckerman, C. (1981). Application of confirmatory factor analysis to verify the construct validity of the Behavior Problem Checklist and the Bristol Social Adjustment Guides. Educational and Psychological Measurement, 41, 843-850.

Harris, A., Kreil, D., & Orpet, R. (1977). The modification and validation of the Behavior Coding System for school settings. Educational and Psychological Measurement, 37, 1121-1126.

Harrop, L. (1979). Unreliability of classroom observation. Educational Research, 21(3), 207-211.

Harvey, R. (1982). The future of partial correlation as a means to reduce halo in performance ratings. Journal of Applied Psychology, 67, 171-176.

Haskell, S. (1979). A time-referenced Q-sort technique for evaluating behavioral change. American Journal of Orthopsychiatry, 49(1), 109-120.

Helson, H. (1969). Adaptation level theory. New York: Harper and Row.

Hewett, F., & Forness, S. (1982). Education of exceptional learners (3rd ed.). Boston: Allyn & Bacon.

Hinton, J., Webster, S., & O'Neill, M. (1978). Simple behaviour rating scales for maximum security patients. British Journal of Social and Clinical Psychology, 17(SEP), 255-259.

Hirshoren, A., & Heller, G. (1979). Programs for adolescents with behavior disorders: The state of the art. The Journal of Special Education, 13(3), 275-281.

Holzbach, R. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. Journal of Applied Psychology, 63(5), 579-588.

Hom, P., DeNisi, A., Kinicki, A., & Bannister, B. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. Journal of Applied Psychology, 67(5), 568-576.

Horne, M., & Larrivee, B. (1979). Behavior rating scales: Need for refining normative data. Perceptual and Motor Skills, 49, 383-388.

Horowitz, L., Inouye, D., & Siegelman, E. (1979). On averaging judges' ratings to increase their correlation with an external criterion. Journal of Consulting and Clinical Psychology, 47(3), 453-458.

Howell, K. (1978). Evaluation of behavior disorders. In R. Rutherford & A. Prieto (Eds.), Severe behavior disorders of children and youth. Bloomington, Ind.: Indiana University, Department of Special Education. (Monograph)

Hulin, C. (1982). Some reflections on general performance dimensions and halo rating error. Journal of Applied Psychology, 67, 165-170.

Hunter, J., Hunter, R., & Lopis, J. (1979). A causal analysis of attitudes toward leadership training in a classroom setting. Human Relations, 32(11), 889-907

Ingram, G., Gerard, R., Quay, H., & Levinson, R. (1970). An experimental program for the psychopathic delinquent. Journal of Research in Crime and Delinquency, Jan, 24-30.

Ivancevich, J. (1980). A longitudinal study of behavioral expectation scales: Attitudes and performance. Journal of Applied Psychology, 65(2), 139-146.

Jacob, T., Grounds, L., & Haley, R. (1982). Correspondence between parents' reports on the Behavior Problem Checklist. Journal of Abnormal Child Psychology, 10(4), 593-608.

Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. Personnel Psychology, 33, 595-604.

Jesness, C. (1970). The Jesness behavior checklist manual. Palo Alto, CA: Consulting Psychologists Press.

Jesness, C. (1972). The Jesness inventory manual. Palo Alto, CA: Consulting Psychologists Press.

Jessor, R. (1982). Problem behavior and developmental transition in adolescence. The Journal of School Health, 52(5), 295-300.

Jessor, R., & Jessor, S. (1977). Problem behavior and psychosocial development. New York: Academic Press.

Johnson, S., Smith, P., & Tucker, S. (1982). Response format of the Job Descriptive Index: Assessment of reliability and validity by the multitrait-multimethod matrix. Journal of Applied Psychology, 67(4), 500-505.

Jones, R., Reid, J., & Patterson, G. (1975). Naturalistic observations in clinical assessment. In P. McReynolds (Ed.), Advances in psychological assessment (Vol. 3, pp.42-95). San Francisco: Jossey-Bass.

Joshi, M. (1964). A factorial study of the problems of adjustment. Indian Psychological Review, 1, 42-46.

Kahn, P., & Ribner, S. (1982). A brief behavior rating scale for children in a school setting. Psychology in the Schools, 19, 113-116.

Kane, J., & Bernardin, H. (1982). Behavioral observation scales and the evaluation of performance appraisal effectiveness. Personnel Psychology, 35, 635-641.

Kassin, S., & Wrightsman, L. (1983). The construction and validation of a juror bias scale. Journal of Research in Personality, 17, 423-442.

Kaufman, A. (1975). Factor analysis of the WISC-R at 11 age levels between 6-1/2 and 16-1/2 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kaufman, A., Swan, W., & Wood, M. (1979). Dimensions of problem behaviors of emotionally disturbed children as seen by their parents and teachers. Psychology in the Schools, 16(2), 207-216.

Kazdin, A., Esveldt-Dawson, K., & Loar, L. (1983). Correspondence of teacher ratings and direct observations of classroom behavior of psychiatric inpatient children. Journal of Abnormal Child Psychology, 11(4), 549-564.

Keaveny, T., & McGann, A. (1975). A comparison of behavioral expectation scales and graphic rating scales. Journal of Applied Psychology, 60(6), 695-703.

Kelley C. (1981). Reliability of the Behavior Problem Checklist with institutionalized male delinquents. Journal of Abnormal Child Psychology, 9(2), 243-250.

Kenny, D., & Berman, J. (1980). Statistical approaches to the correction of correlational bias. Psychological Bulletin, 88(2), 288-295.

Kerlinger, F. (1972). Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

King, L., Hunter, J., & Schmidt, F. (1980). Halo in a multidimensional forced-choice performance evaluation scale. Journal of Applied Psychology, 65(5), 507-516.

Kingstrom, P., & Bass, A. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. Personnel Psychology, 34, 263-289.

Kohn, M., Koretzky, M., & Haft, M. (1979). An Adolescent Symptom Checklist for Juvenile Delinquents. Journal of Abnormal Child Psychology, 7(1), 15-29.

Konopasek, D. (1983). Behavior evaluation scale. Behavioral Disorders, 8(4), 280-281.

Kowalski, G., Adams, M., & Gundlach, J. (1983). Structural determinants of juvenile offenses in school. Urban Education, 18(2), 179-190.

Kreitler, S., & Kreitler, H. (1981). Test item content: Does it matter? Educational and Psychological Measurement, 41, 635-642.

Kulka, R., Klingel, D., & Mann, D. (1980). School crime and disruption as a function of student-school fit: An empirical assessment. Journal of Youth and Adolescence, 9(4), 353-370.

Lachar, D., & Gdowski, C. (1979). Problem-behavior factor correlates of Personality Inventory for Children profile scales. Journal of Consulting and Clinical Psychology, 47(1), 39-48.

Lahey, B., Green, K., & Forehand, R. (1980). On the independence of ratings of hyperactivity, conduct problems, and attention deficits in children: A multiple regression analysis. Journal of Consulting and Clinical Psychology, 48(5), 566-574.

Landy, F., Vance, R., & Barnes-Farrell, J. (1982). Statistical control of halo: A response. Journal of Applied Psychology, 67, 177-180.

Landy, F., Vance, R., Barnes-Farrell, J., & Steel, J. (1980). Statistical control of halo error in performance ratings. Journal of Applied Psychology, 65, 501-506.

Larrivee, B., & Bourque, M. (1980). A comparative study of two rating formats. Measurement and Evaluation in Guidance, 12(4), 223-228.

Latham, G., Fay, C., & Saari, L. (1979). The development of behavioral observation scales for appraising the performance of foremen. Personnel Psychology, 32, 299-311.

Latham, G., & Wexley, K. (1977). Behavioral observation scales. Personnel Psychology, 30, 255-268.

Latham, G., Wexley, K., & Pursell, E. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.

Law Enforcement Assistance Administration. (October 15, 1979). Grants to schools operating alternative education programs. Federal register.

Lawshe, C. (1975). A quantitative approach to content validity. Personnel Psychology, 28, 563-575.

Ledingham, J. Younger, A., Schwartzman, A., & Bergeron, G. (1982). Agreement among teacher, peer, and self-ratings of children's aggression, withdrawal, and likability. Journal of Abnormal Child Psychology, 10(3), 363-372.

Lee, R., Malone, M., & Greco, S. (1981). Multitrait-multimethod-multirater analysis of performance ratings for law enforcement personnel. Journal of Applied Psychology, 66(5), 625-632.

Lessing, E., & Clarke, C. (1982). Reliability and validity of IJR Behavior Checklist scores: Number versus pathology level of symptoms. Journal of Abnormal Child Psychology, 10(3), 337-362.

Lessing, E., Williams, V., & Gil, E. (1982). A cluster-analytically derived typology: Feasible alternative to clinical diagnostic classification of children. Journal of Abnormal Child Psychology, 10(4), 451-482.

Lessing, E., Williams, V., & Revelle, W. (1981). Parallel forms of the IJR Behavior Checklist for parents, teachers, and clinicians. Journal of Consulting and Clinical Psychology, 49(1), 34-50.

Levin, H. (1972). The costs to the nation of inadequate education. Washington, D.C.: Senate Select Committee on Equal Educational Opportunity. (ERIC Document Reproduction Service No. ED 064 437)

Levine, M. (1977). Sex differences in behavior ratings: Male and female teachers rate male and female pupils. American Journal of Community Psychology, 5(3), 347-353.

Leyser, Y., & Abrams, P. (1982). Teacher attitudes toward normal and exceptional groups. The Journal of Psychology, 110, 227-238.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, No. 140.

Lines, P. (1972). The case against short suspensions. In Inequality in education. Boston: Center for Law and Education.

Linton, H., & Chavez, C. (1979). Behavior checklist for junior high students. National Association of Secondary School Principals Bulletin, 63(431), 119-121.

Lissitz, R., & Green, S. (1975). Effect of the number of scale points on reliability: A Monte Carlo study. Journal of Applied Psychology, 60, 10-13.

Lobitz, G., & Johnson, S. (1975). Normal vs. deviant children: A multimethod comparison. Journal of Abnormal Child Psychology, 3, 353-374.

Loeber, R. (1982). The stability of antisocial and delinquent child behavior: A review. Child Development, 53, 1431-1446.

Loranger, M., Lacroix, O., & Kaley, R. (1982). Validity of teachers' evaluations of students' social behavior. Psychological Reports, 51, 915-920.

Lovitt, T. (1967). Assessment of children with learning disabilities. Exceptional Children, 34, 233-239.

Lufler, H. (1982). Past court cases and future school discipline. Education and Urban Society, 14(2), 169-184.

Lyness, K., & Cornelius, E. (1982). A comparison of holistic and decomposed judgment strategies in a performance rating simulation. Organizational Behavior and Human Performance, 29, 21-38.

Madle, R., Neisworth, J., & Kurtz, P. (1980). Biasing of hyperkinetic behavior ratings by diagnostic reports: Effects of observer training and assessment method. Journal of Learning Disabilities, 13, 35-38.

Marwit, K., Marwit, S., & Walker, E. (1978). Effects of student race and physical attractiveness on teachers' judgments of transgression. Journal of Educational Psychology, 70, 911-915.

Marwit, S. (1982). Students' race, physical attractiveness and teachers' judgments of transgressions: Follow-up and clarification. Psychological Reports, 50, 242.

Masterson, S. (1968). The adjective checklist technique: A review and critique. In P. McReynolds (Ed.), Advances in psychological assessment (pp. 275-312). Palo Alto, Calif.: Science and Behavior Books.

Matusek, P., & Oakland, T. (1979). Factors influencing teachers' and psychologists' recommendations regarding special class placement. Journal of School Psychology, 17, 116-125.

Mayer, G., & Butterworth, T. (1979). A preventive approach to school violence and vandalism: An

experimental study. Personnel and Guidance Journal, 57(9), 436-441.

McCarney, S., Leigh, J., & Cornbleet, J. (1983). Behavior evaluation scale. Columbia, Mo.: Educational Services.

McCarthy, J., & Paraskevapoulas, J. (1969). Behavior patterns of learning disabled, emotionally disturbed, and average children. Exceptional Children, 36, 69-74.

McDermott, P. (1980). Principal components analysis of the revised Bristol Social Adjustment Guides. British Journal of Educational Psychology, 50, 223-228.

McDermott, P. (1981). A syndromic typology for analyzing schoolchildren's disturbed social behavior. Philadelphia: University of Pennsylvania.

McDermott, P., & Hale, R. (1982). Validation of a systems-actuarial computer process for multidimensional classification of child psychopathology. Journal of Clinical Psychology, 38(3), 477-485.

McKelvie, S. (1978). Graphic rating scales--How many categories? British Journal of Psychology, 69, 185-202.

McKinney, J., & Forman, S. (1982). Classroom behavior patterns of EMH, LD, and EH students. Journal of School Psychology, 20(4), 271-279.

McSweeney, A., & Trout, B. (1979). Predicting treatment outcome through profile analysis of the Jesness Behavior Checklist. Paper presented at the annual convention of The Association for Behavior Analysis, Dearborn, Michigan. (ERIC Document Reproduction Service No. ED 177 436)

Mendelsohn, M., & Erdwins, C. (1978). The disruptive behavior scale: An objective assessment of unmanageable social behaviors in adolescents. Journal of Clinical Psychology, 34(2), 126-128.

Meredith, G. (1975). Toward a systems approach to student-based ratings of instruction. Journal of Psychology, 91, 235-246.

Meredith, G. (1981). Preferred length of scales for students' evaluation of instruction. Perceptual and Motor Skills, 53, 490.

Mesinger, J. (1982). Alternative education for behavior disordered and delinquent adolescent youth: What works--Maybe? Behavior Disorders, 7(2), 91-100.

Messick, S. (1964). Personality measurement and college performance. Proceedings of the 1963 invitational conference on testing problems. Princeton, N.J.: Educational Testing Service.

Messick, S. (1965). Personality measurement and the ethics of assessment. American Psychologist, 20, 136-142.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Miller, L. (1967). Louisville behavior check list for males 6-12 years of age. Psychological Reports, 21, 885-896.

Miller, L. (1980). Dimensions of adolescent psychopathology. Journal of Abnormal Child Psychology, 8(2), 161-173.

Mischel, W. (1969). Continuity and change in personality. American Psychologist, 24, 1012-1018.

Mitchell, S., & Rosa, P. (1981). Boyhood behavior problems as precursors of criminality: A fifteen-year follow-up study. Journal of Child Psychology and Psychiatry, 22, 19-33.

Mooney, R. (1942). Surveying high school students' problems by means of a problem check list. Educational Research Bulletin, 21, 57-69.

Morris, J., & Arrant, D. (1978). Behavior ratings of emotionally disturbed children by teachers, parents, and school psychologists. Psychology in the Schools, 15(3), 450-455.

Moses, W. (1974). A behavior checklist for teachers' use in selecting students for a behavior counseling program. Unpublished manuscript, University of South Florida.

Moses, W. (1976). Final report on Meaningful Alternatives to Regular Schooling--MARS (Proj. No. 75-AS-08-A104). Dade City, Fla.: Pasco County School District.

Moses, W. (1981). Conduct code violations, 1980-81. Jacksonville, Fla.: Duval County School District.

Mossholder, K., & Giles, W. (1983). The use of partial correlation to control halo in performance ratings. Educational and Psychological Measurement, 43, 977-984.

Mosteller, F., & Tukey, J. (1977). Data analysis and regression. Reading, Mass.: Addison-Wesley.

Moyer, T., & Motta, R. (1982). Alienation and school adjustment among Black and white adolescents. The Journal of Psychology, 112, 21-28.

Murphy, K. (1982). Difficulties in the statistical control of halo. Journal of Applied Psychology, 67, 161-164.

Murphy, K., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observation? Journal of Applied Psychology, 67(5), 562-567.

National Education Association (Ed.) (1975). Discipline and learning. Washington, D.C.: NEA.

National Education Association. (1980). The 1980 teacher-opinion poll. Washington, D.C.: Author.

National Institute of Education. (1978). Violent schools--Safe schools: The safe school study report to Congress. Washington, D.C.: Department of Health, Education, and Welfare.

National School Public Relations Association. (1973). Discipline crisis in schools: The problem, causes, and search for solutions. Arlington, Va.: Author.

Neumann, L., & Neumann, Y. (1981). Comparison of six lengths of rating scales: Students' attitudes toward instruction. Psychological Reports, 48, 399-404.

New York State United Teachers. (1979). Survey of classroom stressors. New York: American Federation of Teachers.

Nielsen, A., & Gerber, D. (1979). Psychosocial aspects of truancy in early adolescence. Adolescence, 14(54), 313-326.

Nihira, K., Foster, R., Shellhaas, M., & Leland, H. (1969). AAMD adaptive behavior scale. Washington, D.C.: American Association on Mental Deficiency.

O'Leary, K., & Johnson, S. (1979). Psychological assessment. In H. Quay & J. Werry (Eds.), Psychopathological disorders of childhood (2nd ed., pp. 162-175). New York: Wiley.

O'Leary, K., & Kent, R. (1973). Behavior modification for social action. In L. Hamerlynck, L. Bundy, & E. Mash (Eds.), Behavior change: Methodology, concepts, and practice (pp. 69-96). Champaign, Ill.: Research Press.

Ostrov, E., Marohn, R., Offer, D., Curtiss, G., & Feczko, M. (1980). The adolescent antisocial behavior check list. Journal of clinical psychology, 36(2), 594-601.

Parducci, A. (1968). Often is often. American Psychologist, 23, 828.

Patterson, G., Ray, R., Shaw, D., & Cobb, J. (1969). A manual for coding of family interactions (Document No. 01234). New York: Microfiche Publications.

Peed, S., & Pinsker, M. (1978). Behavior change procedures. Education and Urban Society, 10(4), 501-520.

Pepper, S., & Prytulak, L. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expression. Journal of Research and Personality, 8, 95-101.

Peterson, D. (1961). Behavior problems of middle childhood. Journal of Consulting Psychology, 25, 205-209.

Pettegrew, L., & Wolf, G. (1982). Validating measures of teacher stress. American Educational Research Journal, 19(3), 373-396.

Pinellas County School District. (Rev. 8/82). Alternative school referral form, PCS Form 1709. St. Petersburg, Fla: Author.

Pinellas County Schools. (1983). Student code of conduct. St. Petersburg, Fla.: Author.

Pisarra, J., & Giblette, J. (1981). Administrators' perceptions of aggressive behaviors. National Association of Secondary School Principals Bulletin, 65(441), 49-53.

Pohl, N. (1981). Scale considerations in using vague quantifiers. Journal of Experimental Education, 49(4), 235-240.

Poulton, E. (1976). Quantitative subjective assessments are almost always biased, sometimes completely misleading. Bulletin of the British Psychological Society, 29, 385-387.

Prinz, R., & Kent, R. (1978). Recording parent-adolescent interactions without the use of frequency or interval-by-interval coding. Behavior Therapy, 9, 602-604.

Pursell, E., Dossett, D., & Latham, G. (1980). Obtaining valid predictors by minimizing rating errors in the criteria. Personnel Psychology, 33, 91-96.

Quay, H. (1964). Personality dimensions in delinquent males as inferred from the factor analysis of behavior ratings. Journal of Research into Crime and Delinquency, 1, 33-37.

Quay, H. (1977). Measuring the dimensions of deviant behavior: The behavior problem checklist. Journal of Abnormal Child Psychology, 5(3), 277-287.

Quay, H. (1978). Behavior disorders in the classroom. Journal of Research and Development in Education, 11(2), 8-17.

Quay, H., & Peterson, D. (1967). Manual for the behavior problem checklist. Champaign, Ill.: Children's Research Center, University of Illinois.

Quay, H., & Peterson, D. (1975). Manual for the behavior problem checklist. Miami, Fla.: Author.

Quay, H., & Peterson, D. (1979). Manual for the behavior problem checklist. New Brunswick, N.J.: Rutgers State University.

Reed, M., & Edelbrock, C. (1983). Reliability and validity of the Direct Observation Form of the Child

Behavior Checklist. Journal of Abnormal Child Psychology, 11(4), 521-530.

Reeves, J., Perkins, M., & Hollon, T. (1978). Issues in the effective use of the Walker Problem Behavior Identification Checklist. Spartanburg, S.C.: Spartanburg School District Six. (ERIC Document Reproduction Service No. ED 169 115)

Reisberg, L., Fudell, I., & Hudson, F. (1982). Comparison of responses to the Behavior Rating Profile for mild to moderate behaviorally disordered subjects. Psychological Reports, 50, 136-138.

Roberts, A., & Jenkins, P. (1982). Teachers' perceptions of assertive and aggressive behavior at school: A discriminant analysis. Psychological Reports, 50, 827-832.

Roberts, M., Milich, R., Loney, J., & Caputo, J. (1981). A multitrait-multimethod analysis of variance of teachers' ratings of aggression, hyperactivity, and inattention. Journal of Abnormal Child Psychology, 9(3), 371-380.

Robins, L. (1966). Deviant children grown up. Baltimore: Williams and Wilkins.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. New York: Holt, Rinehart, and Winston.

Ross, A., Lacey, H., & Parton, D. (1965). The development of a behavior checklist for boys. Child Development, 36, 1013-1027.

Ross, M., & Salvia, J. (1975). Attractiveness as a biasing factor in teacher judgments. American Journal of Mental Deficiency, 80, 96-98.

Rubel, R. (1977). The unruly school: Disorders, disruptions, and crimes. Lexington, Mass.: Lexington Books.

Rubel, R. (1980). Crime and violence in public schools: Emerging perspectives of the 1980's. Contemporary Education, 52(1), 5.

Russell, P., Lankford, M., & Grinnell, R. (1981). A management evaluation scale for social workers. The Journal of Psychology, 107, 127-130.

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. Journal of Child Psychology and Psychiatry, 8, 1-11.

Ryan, F. (1958). Trait ratings of high school students by teachers. Journal of Educational Psychology, 49(3), 124-128.

Safer, D., Heaton, R., & Parker, F. (1981). A behavioral program for disruptive junior high school students: Results and follow-up. Journal of Abnormal Child Psychology, 9(4), 483-494.

Salvia, J., & Ysseldyke, J. (1981). Assessment procedures in special and remedial education (2nd ed.). Boston: Houghton-Mifflin.

Sanson-Fisher, R., & Mulligan, B. (1977). The validity of a behavioral rating scale. The Journal of Multivariate Behavioral Research, 12, 357-372.

Schenck, S. (1980). The diagnostic/instructional link in Individualized Education Programs. The Journal of Special Education, 14, 337-345.

Schriesheim, C. (1981a). The effect of grouping or randomizing items on leniency response bias. Educational and Psychological Measurement, 41, 401-411.

Schriesheim, C. (1981b). Leniency effects on convergent and discriminant validity for grouped questionnaire items: A further investigation. Educational and Psychological Measurement, 41, 1093-1099.

Schriesheim, C., & DeNisi, A. (1980). Item presentation as an influence on questionnaire validity. Educational and Psychological Measurement, 40, 175-182.

Schriesheim, C., & Hill, K. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. Educational and Psychological Measurement, 41, 1101-1114.

Schriesheim, C., & Schriesheim, J. (1974). Development and empirical verification of new response categories to increase the validity of multiple response alternative questionnaires. Education and Psychological Measurement, 34, 877-884.

Schriesheim, C., & Schriesheim, J. (1978). The invariance of anchor points obtained by magnitude estimation and pair-comparison treatment of complete ranks scaling procedures. Educational and Psychological Measurement, 38, 977-983.

Schwab, D., Heneman, H., & DeCotiis, T. (1975). Behaviorally anchored rating scales: A review of the literature. Personnel Psychology, 28, 549-562.

Searls, E., Isett, R., & Bowders, T. (1981). Examination of item weighting on the Adaptive Behavior Scale Part II. Perceptual and Motor Skills, 53, 154.

Seidman, E., Rappaport, J., Kramer, J., Linney, J., Herzberger, S., & Alden, L. (1979). Assessment of classroom behavior: A multiattribute, multisource approach to instrument development and validation. Journal of Educational Psychology, 71(4), 451-464.

Sherif, M. (1954). Integrating field work and laboratory in small group research. American Sociological Review, 11, 759-771.

Sherry, S. (1979). Behavioral characteristics of educable mentally retarded, emotionally handicapped, and learning disabled students. Washington, D.C.: Department of Health, Education and Welfare, Bureau of Education for the Handicapped.

Shuller, D., & McNamara, J. (1976). Expectancy factors in behavioral observation. Behavior Therapy, 7, 519-527.

Siegel, L., Dragovich, S., & Marholin, D. (1976). The effects of biasing information on behavioral observations and rating scales. Journal of Abnormal Child Psychology, 4(3), 221-233.

Sikes, M. (1975). Law and order and race in the classroom. In National Education Association (Ed.). Discipline and learning (pp.115-119). Washington, D.C.: NEA.

Silvern, L. (1978). Masculinity-feminity in children's self-concepts: The relationship to teacher's judgments of social adjustment and academic ability, classroom behavior, and popularity. Sex Roles, 4(6), 929-949.

Sinclair, E. (1980). Relationship of psychoeducational diagnosis to educational placement. Journal of School Psychology, 18, 349-353.

Sinclair, E., & Kheifets, L. (1982). Use of clustering techniques in deriving psychoeducational profiles. Contemporary Educational Psychology, 7, 81-89.

Smith, C. (1976). Identification of youngsters with emotional disabilities. Des Moines, Iowa: Iowa State Department of Public Instruction. (ERIC Document Reproduction Service No. ED 177 750)

Smith, P., & Kendall, L. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47(2), 149-155.

Solomon, D., & Kendall, A. (1975). Teachers' perceptions of and reactions to misbehavior in traditional and open classrooms. Journal of Educational Psychology. 67(4), 528-530.

Solomon, D., & Kendall, A. (1977). Dimensions of children's classroom behavior, as perceived by teachers. American Educational Research Journal, 14(4), 411-421.

Spector, P. (1976). Choosing response categories for summated rating scales. Journal of Applied Psychology, 61, 374-375.

Spivack, G., Haimes, P., & Spotts, J. (1967). Devereux adolescent behavior rating scale manual. Pennsylvania: The Devereux Foundation.

Spivack, G., & Swift, M. (1966). The Devereux elementary school behavior rating scales. Journal of Special Education, 1, 71-90.

Spivack, G., & Swift, M. (1971). Hahnemann High School Behavior rating scale (HHSB). Philadelphia: Hahnemann Medical College and Hospital, Department of Mental Health Services.

Spivack, G., & Swift, M. (1973). The classroom behavior of children: A critical review of teacher-administered rating scales. The Journal of Special Education, 7(1), 55-89. (Monograph)

Spivack, G., & Swift, M. (1977). The Hahnemann High School Behavior (HHSB) rating scale. Journal of Abnormal Child Psychology, 5(3), 299-307.

Stavig, G. (1978). The median controversy. Perceptual and Motor Skills, 47, 1175-1178.

Stavig, G. (1982). The normalized mean. Perceptual and Motor Skills, 54, 51-54.

Stenner, A., & Smith, M. (1982). Testing construct theories. Perceptual and Motor Skills, 55, 415-426.

Stewart, D., & Deiker, T. (1976). An item factor analysis of the Mooney Problem Check List. Educational and Psychological Measurement, 36, 509-513.

Stott, D. (1971). Classification of behavior disturbance among school-age students: Principles, epidemiology, and syndromes. Psychology in the Schools, 8, 232-239.

Stott, D. (1972). Manual for the Bristol social adjustment guides. San Diego: Educational and Industrial Testing Service.

Stott, D. (1978). Epidemiological indicators of the origins of behavior disturbance as measured by the Bristol Social Adjustment Guides. Genetic Psychology Monographs, 97, 125-159.

Stott, D., Marston, N., & Neill, S. (1975). Taxonomy of behavior disturbance. London: University of London Press Ltd.

Stott, D., & Sykes, E. (1956). The Bristol social adjustment guides. London: University of London Press.

Stott, D., & Wilson, D. (1977). The adult criminal as juvenile. British Journal of Criminology, 7, 45-57.

Strahan, R. (1980). More on averaging judges' ratings: Determining the most reliable composite. Journal of Consulting and Clinical Psychology, 48(5), 587-589.

Strain, P., Cooke, T., & Apolloni, T. (1976). Teaching exceptional children: Assessing and modifying social behavior. New York: Academic Press.

Sulzbacher, S. (1973). Psychotropic medication with children: An evaluation of procedural biases in results of reported studies. Pediatrics, 51(3), 513-517.

Sundberg, N. (1961) The practice of psychological testing in clinical services in the United States. American Psychologist, 16, 79-83.

Tarte, R., Vernon, C., Luke, D., & Clark, H. (1982). Comparison of responses by normal and deviant populations to Louisville Behavior Checklist. Psychological Reports, 50, 99-106.

Taylor, R., Warren, S., & Slocumb, P. (1979). Categorizing behavior in terms of severity. American Journal of Mental Deficiency, 83(4), 411-414.

Thorndike, E. (1920). A constant error in psychological ratings. Journal of Applied Psychology, 4, 25-29.

Thorne, F. (1978). Methodological advances in the validation of inventory items, scales, profiles, and interpretations. Journal of Clinical Psychology, 34(2), 283-301.

Thurstone, L., & Chave, E. (1929). The measurement of attitude. Chicago: University of Chicago Press.

Tinto, V., Paclilio, E., & Cullen, F. (1978). The social patterning of deviant behaviors in school. Washington, D.C.: Department of Health, Education, and Welfare.

Torgerson, W. (1958). Theory and methods of scaling. New York: Wiley.

Touliatos, J., & Lindholm, B. (1981). Congruence of patients' and teachers' ratings of children's behavior problems. Journal of Abnormal Child Psychology, 9(3), 347-354.

Tygart, C. (1980). Student social structures and/or subcultures as factors in school crime: Toward a paradigm. Adolescence, 15(57), 13-22.

Tzeng, O. (1983). A comparative evaluation of four response formats in personality ratings. Educational and Psychological Measurement, 43, 935-950.

United States Department of Justice. (1982). Assessing the relationship of adult criminal careers to juvenile careers: A summary. Washington, D.C.: Author.

Vance, R., Kuhnert, K., & Farr, J. (1978). Interview judgments: Using external criteria to compare behavioral and graphic scale ratings. Organizational Behavior and Human Performance, 22, 279-294.

Vergon, C., & Williams, J. (Eds.). (1978). School suspension practices in Michigan. Student discipline: Policies, programs, and procedures. Ann Arbor, Mich.: University of Michigan, The Governor's Task Force on Disruptive Youth.

Vinter, R., Sarri, R., Vorwaller, D., & Schafer, W. (1966). Pupil behavior inventory: A manual for administration and scoring. Ann Arbor, Mich.: Campus Publishers.

Waksman, S., & Loveland, R. (1980). The Portland problem behavior checklist. Psychology in the Schools, 17(1), 25-29.

Walker, H. (1970). The Walker problem behavior identification checklist. Los Angeles: Western Psychological Services.

Walker, H. (1979). The acting-out child: Coping with classroom disruption. Boston: Allyn and Bacon.

Walker, H., & Holland, F. (1979). Issues, strategies, and perspectives in the management of disruptive child behavior in the classroom. Journal of Education, 161(2), 25-36.

Walker, H., & Hops, H. (1976). Use of normative peer data as a standard for evaluating classroom treatment effects. Journal of Applied Behavior Analysis, 9, 159-168.

Wallbrown, J., Wallbrown, F., & Blaha, J. (1976). The stability of teacher ratings on the Devereux Elementary School Behavior Rating Scale. Journal of Experimental Education, 44(4), 20-22.

Waters, L., Reardon, M., & Edwards, J. (1982). Multitrait-multimethod analysis of three rating formats. Perceptual and Motor Skills, 55, 927-933.

Weinrott, M. (1979). The utility of tracking child behavior. Journal of Abnormal Child Psychology, 7(3), 275-286.

Werry, J., Methuen, R., Fitzpatrick, J., & Dixon, H. (1983). The interrater reliability of DSM III in children. Journal of Abnormal Child Psychology, 11(3), 341-354.

White, W. (1977). The use of the Behavioral Maturity Scale in formative and summative evaluation of young children. Education, 98(1), 17-22.

Willems, E. (July 13-19, 1975). Relations of models to methods in behavior ecology. Paper presented at the Biennial Conference, International Society for the Study of Behavior Development, Guilford, Surrey, England.

Willingham, W., & Jones, M. (1958). On the identification of halo through analysis of variance. Educational and Psychological Measurement, 18, 403-407.

Wixson, S. (1980). Two resource room models for serving learning and behavior disordered pupils. Behavioral Disorders, 5(2), 116-125.

Wodarksi, J., & Pedi, S. (1978). The empirical evaluation of the effects of different group treatment strategies. Journal of Clinical Psychology, 34(2), 471-481.

Yard, G. (1977). Behavioral disorders. Des Moines, Iowa: Drake University, Midwest Regional Resource Center. (ERIC Document Reproduction Service No. ED 163 707)

Ysseldyke, J., & Algozzine, B. (1982). Bias among professionals who erroneously declare students eligible for special services. Journal of Experimental Education, 50(4), 223-228.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. (1979). The influence of test scores and naturally occurring pupil characteristics on psychoeducational decision making with children (Research Report No. 17). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. (1981). The influence of test scores and naturally-occurring pupil characteristics on psychoeducational decision making with children. Journal of School Psychology, 19(2), 167-177.

Ysseldyke, J., Algozzine, B., & Richey, L. (1982). Judgments under uncertainty: How many children are handicapped? Exceptional Children, 48(6), 531-534.

Ysseldyke, J., & Marston, D. (1982). Gathering decision making information through the use of non-test-based methods. Measurement and Evaluation in Guidance, 15(1), 58-69.

Zammuto, R., London, M., & Rowland, K. (1982). Organization and rater differences in performance appraisals. Personnel Psychology, 35, 643-658.

## BIOGRAPHICAL SKETCH

William L. Moses was born on March 20, 1936, in Macon, Georgia. He attended public schools in Georgia. In 1961 he graduated, summa cum laude, from Mercer University, with a Bachelor of Arts degree in economics and psychology. From 1961 to 1972 he was employed in various business ventures. From 1972 to 1974 Mr. Moses was a graduate student in the Departments of Psychology and Education at the University of South Florida, receiving the Master of Arts degree in school psychology. During this period, he developed and coordinated the Adolescent Resocialization Residential Project at Memorial Hospital in Tampa, Florida.

From 1974 to 1976 Mr. Moses was employed in Pasco County, Florida, as a school psychologist and coordinator of a federally-funded alternative schooling project, which was developed from his master's thesis research. In 1976, Mr. Moses joined Pasco-Hernando Community College as an instructor in psychology and business. In 1979 he received a one-year sabbatical leave to pursue a doctoral degree in counselor education at the University of Florida. After an additional year of leave conducting dissertation research while employed as a school psychologist in Duval

County, Florida, Mr. Moses returned to Pasco-Hernando Community College where he is currently employed.

Mr. Moses maintains an active family counseling practice and is licensed in Florida as a mental health counselor. He is also a National Certified Counselor and a Certified Clinical Mental Health Counselor.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Roderick McDavis, Chairman
Professor of Counselor Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Larry C. Loesch, Cochairman
Professor of Counselor Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert C. Tiller
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1986

David C. Smith (N)
Dean, College of Education

Madelyn Cockhart
Dean, Graduate School