A CROSS VALIDATION STUDY OF THE EFFECTIVENESS OF THREE BEHAVIOR RATING SCALES IN IDENTIFYING DIFFICULTIES IN PRIMARY SCHOOL CHILDREN

THERESA MARIE LEONARD

000321

NOTICE

AVIS

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

·LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

A CROSS VALIDATION STUDY OF THE EFFECTIVENESS

OF THREE BEHAVIOR RATING SCALES IN IDENTIFYING

DIFFICULTIES IN PRIMARY SCHOOL CHILDREN

by

Ⓒ   Theresa Marie Leonard

A thesis presented to
the Department of Educational Psychology
Memorial University of Newfoundland

In partial fulfillment
of the requirements for the degree
Master of Education
December 1979

St. John's   Newfoundland

ABSTRACT

This study was initiated to investigate the usefulness of teacher completed behavior rating scales as a means of detecting high and low risk students. The study also sought to find out which, if any, scale best demonstrated potential risk of school difficulties in primary school age children. The three behavior rating scales chosen for this exploration were: the AML Behavior Rating Scale by Van Vleet (1970), the Children's Behavior Questionnaire by Rutter (1967), and the Devereaux Elementary School Behavior Rating Scale by Spivack and Swift (1967). All behavior rating scales were compared with the Wide Range Achievement Test by Jastak and Jastak (1978). Two groups of subjects were studied firstly as a total group and then as subgroups of the total group.

Subjects were categorized into high and low risk firstly by teacher ranking of the likelihood of school difficulties and secondly by level of reading achievement as measured by the Wide Range Achievement Test.

The scores of the three behavior rating scales and the Wide Range Achievement Test were statistically analyzed using Pearson and Kendal's Tau correlation coefficients and a one way and two way analysis of variance.

The results indicated that all three bheavior rating scales correlate highly with one another, with teacher rank, as well as with scores of the Wide Range Achievement Test. These findings suggest that

ii

these three behavior rating scales can differentiate between groups of high and low risk students, split by teacher rank as well as by reading achievement.

Results of this research generally suggest that, firstly, these three behavior rating scales are useful devices for detecting children who are likely to have school difficulties, secondly, no one scale is better than another, and thirdly, the three behavior rating scales can be used by the teacher to detect high and low risk children.

ACKNOWLEDGEMENTS

This study is the result of the wise counsel and the generous contributions of a number of persons.

To Dr. Gary Jeffery, my supervisor, my sincere gratitude for his advice and assistance. His criticisms, suggestions and confidence have guided my efforts in this study from their formative stages to the concluding statements.

My thanks to Dr. Terry Boak for helping with the statistical aspects of this research.

To the principals, teachers and children who participated in this study, my sincere appreciation for their help and support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

PURPOSE AND RATIONALE

I. PURPOSE OF THE STUDY

This research seeks to explore the usefulness of behavior rating scales as a means of detecting children who have or are likely to have school difficulties. More specifically, the research assesses the relative effectiveness of three of the better behavior rating scales in identifying potential school difficulties in primary school age Newfoundland children.

II. RATIONALE FOR THE STUDY

Elementary school counselors are faced with the very real problem of helping teachers identify and cope with children who are vulnerable to school failure. There is a need to provide a means whereby the counselor can readily and indirectly gain valid information on a child. A lack of sufficient funding also often renders impossible the adequate evaluation of potentially high risk children by a trained counselor. The teacher can potentially be that indirect informational source by taking the time to observe and compare children during the day-to-day instruction of his or her class.

While teachers may have considerable information on a child it is not always readily accessible to the counselor. A possible way of obtaining such information from the teacher might be through the

1

use of a behavior rating scale. This approach is attractive in that a large number of potentially high risk children might be screened at an early age and hence be directed toward early help by the classroom teacher or, where necessary, by the counselor; thereby initiating a more preventative than corrective type program (Rutter, 1967).

Unfortunately, at present there are few adequately developed instruments which allow the counselor to readily and accurately collect descriptive information on high risk children (Pritchard, 1963). If a means existed whereby such information could be collected and conveyed to the school counselor, the counselor could then potentially have more time to devote to a more productive advisory role and also potentially be able to spend more time working with the more serious child difficulties.

Most instruments which have been developed are principally designed to be used by counselors or researchers and not classroom teachers. For example, Buros' (1961) in his Tests in Print refers to only a small number of scales and inventories that purport to assess classroom adjustment. The best known of these according to Buros' (1961) are: the S.R.A. Rating Scale for Pupil Adjustment (SRA, 1950-53), the N.Y. Rating Scale for School Habits (Cornell, Coxe, and Orleans, 1927), Personal and Social Development Program (SRA, 1956), and the Pupil Adjustment Inventory (Education Services Bureau, 1957). These instruments were designed to assess a pupil's academic success and temperament or school adjustment. On these instruments, school adjustment is mainly being defined by such traits as courtesy, initiative, sociability, physical health, and industry. Most scales do not

address themselves to the problems experienced by counselors.

In using behavior rating scales one must consider the fact that many who create and use scales or recommend them for use in the school systems unfortunately do not put the required emphasis on 1) the purpose of the measuring device, 2) the choice of items, 3) the reliability and validity, and 4) the nature and quality of the norms (Spivack and Swift, 1973). Because teachers are not trained to evaluate such measurement devices they may use these devices without awareness of their intended purpose and without realizing their limited usefulness in the classroom. Thus, the guidance counselor has the responsibility of assessing the strengths and weaknesses of such measurement devices before recommending particular instruments to teachers.

The researchers who have developed behavior rating scales typically have not been educators but mental health oriented professionals and hence have been more concerned with emotional adjustment than classroom behavior and achievement. Because of this, additional work aimed at validating measures of classroom behavior should be carried out. What is particularly needed is work related to a standard of academic accomplishment such as teacher grades and/or academic test scores, if these instruments are to provide useful knowledge about childhood adaptation to the classroom setting. This view is supported by Davidson and Greenberg (1967); Kim, Anderson and Bashaw (1968); and Miller (1968) who also believe that certain overt behaviors in the classroom are related to academic capacity and achievement and that these behaviors better reflect academic success than do general dimensions of adjustment or personality functioning.

The study was initiated in an attempt to deal with these difficulties by identifying a useful means whereby information can be gathered on a child who is likely to fail. More specifically, the research hopes to assess the effectiveness of three behavior rating scales in "identifying" actual or potential difficulties in primary school age Newfoundland children.

III. RESEARCH QUESTIONS

The research is aimed at exploring the usefulness of behavior rating scales completed by teachers as a means of detecting potential high risk students. The study also seeks to find out which if any scale best demonstrates potential risk. More specifically, the study seeks to discover if three behavior rating scales, namely, the AML Behavior Rating Scale (AML) by Van Vleet (1970), the Devereaux Elementary School Behavior Rating Scale (DES) by Spivack and Swift (1967), and the Children's Behavior Questionnaire (CBQ) by Rutter (1967) are each individually capable of differentiating between higher and lower risk school age children. High risk children are defined in this study as those being likely to have school difficulties and low risk children as those unlikely to have school difficulties. Risk was assessed by both teacher rank and scores on the Wide Range Achievement Test (WRAT).

In operational terms, the following research questions were explored:

1. Do behavior rating scales correlate significantly

   a) with each other,

   b) with teacher rank, and

c) with Wide Range Achievement Test scores?

2. Are high and low risk students as determined by ranking made by teachers rated significantly different on teacher completed behavior rating scales AML, CBQ and the DES?

3. Are high and low risk students as determined by scores obtained on the Wide Range Achievement reading sub-test rated significantly different on teacher completed behavior rating scales AML, CBQ and the DES?

CHAPTER II

RELATED LITERATURE REVIEW

I. INTRODUCTION

This chapter reviews the literature discussing the usefulness
of behavior rating scales as a means of detecting children who have or
are likely to have school difficulties. It is divided into three major
segments, namely:

(I) The Relationship of Classroom Behavior and Achievement;

(II) Review of Literature on Behavior Rating Scales; and

(III) The Hypotheses for this Research.

II. THE RELATIONSHIP OF CLASSROOM BEHAVIOR
AND ACHIEVEMENT

Observations of a child's classroom behavior are regarded as a
significant factor in the prediction of his/her school achievement.
The following literature review explores the relationship between
observed classroom behavior and achievement.

(1) Classroom Behavior and Achievement

The relationship of classroom behavior to achievement was
established in the last twenty years. Research was carried out on
related areas such as task set, work attitude, and emotional impairment
before this link was made. Research in each of these areas will be
reviewed.

a. Work attitude and task set

One of the first to associate work attitude (that is, the extent to which the child displays interest, curiosity, and assertiveness) and task set (the child's ability to organize himself) was Buehler (1935) who believed that work attitude and task set were developed at a very early age and in several stages from the second to the sixth year. Buehler considered intentional work play and striving towards a product a crucial development step and a prerequisite of school success.

Danziger (1933) found similar results between work attitude and task involvement. She reported that 80% of first grade children failed because they had not developed a work attitude in their games before entering school. She reported that only 6% of the failures were in one subject, 50% in two subjects, and 44% in three subjects. Her conclusion was that failure in first grade was due to a general disability that showed itself whenever the child attempted to undertake anything.

This eventually led to a more contemporary approach expressed by Kohn (1973) who stated that during elementary school years the classroom behaviors affecting achievement include the child's ability to function within rules, limits and norms in the classroom as well as the child's ability to organize himself around specific tasks.

b. Emotional impairment

Studies on emotional impairment made an attempt to be more specific in relating school difficulties to the sex of the child. This research led to an eventual link between various school related difficulties and classroom observation of selected behaviors.

Wickman (1928) pioneered this research when he suggested more widespread problems were experienced among boys than girls in the elementary school years. He reported that 10% of boys as compared to 3% of girls were considered to have severe behavior problems. Rogers (1942) found percentages to be 18% for boys, 7% for girls; Ullman (1952) found a difference of 13% and 3% respectively. More recently, similar results have been found by Bower (1969).

Researchers in the 60's have looked at specific symptoms of emotional impairment. It has been typically observed that acting out behavior was more prevalent among boys than girls and that character- istics such as passivity, shyness and anxiety were found to occur either as often in boys as girls or somewhat more frequently in girls. For example, Cullen and Boundy (1966) and Rutter et al. (1970) reported antisocial behaviors as more prevalent among boys whereas characteristics such as thumbsucking and lethargy were more prevalent among girls.

The above stated studies led researchers such as Harris (1961), Stennett (1966), and Bower (1969) to relate emotional impairment with underachievement. Stennett (1966), for example, reported that emotionally handicapped children became progressively farther behind peers in academic achievement over the elementary grades.

c. Classroom behavior and achievement

Following in the 70's Rutter, Tizard and Whitmore (1970) made more associations between specific classroom behaviors and achievement. Rutter et al. (1970) found that underachievement was generally asso- ciated with antisocial behavior. Contrary results are presented by Emmerick (1977), Richards and McCandless (1972), and Kohn (1968) who

reported that apathy and withdrawal are more commonly related to under-
achievement in primary school age children and not antisocial behavior.
More recent research by Kohn (1973) also demonstrated that in primary
grade school age children, shy withdrawal behavior was a better indi-
cator of learning failure than angry defiant behavior.

As the previous discussions have suggested a relationship
between various school related difficulties such as poor work attitude,
emotional impairment, as well as poor achievement, and classroom obser-
vation of selected behaviors does exist. Such a relationship has
provided the bases on which behavior rating scales have been formulated.
Such scales are discussed below.

III. BEHAVIOR RATING SCALES

Potentially useful devices to aid in the screening of those
children to have school related difficulties are behavior rating scales.
These scales when completed by the teacher, particularly when he or she
has the opportunity to view large numbers of students, can be useful
screening devices. Unfortunately, as with other measures of child
behavior, there are few adequately developed instruments (Pritchard,
1963). The following sections briefly review the early behavior rating
scales and their inadequacies. The limitations of currently used scales,
along with a review of the currently available better behavior rating
scales are also included.

(1) Early Behavior Rating Scales

Wickman (1928) was first to attempt to measure a child's behavior
in the classroom. Haggerty (1929) and Olson (1930), after having revised

Wickman's scale, felt it had adequate reliability and validity. Buros'
(1961) refers to only a small number of scales and inventories that
purport to assess classroom adjustment, the better known of these being
the N.Y. Rating Scale for School Habits (Cornell, Coxe, and Orleans,
1927), the S.R.A. Rating Scale for Pupil Adjustment (SRA, 1950-53), the
Personal and Social Development Program (S.R.A., 1956), and the Pupil
Adjustment Inventory (Educational Service Bureau, 1957). These instruments
were mainly designed to focus on such categories as pupils' academic
levels and temperaments or school adjustment. School adjustment or
temperament was typically described as courtesy, initiative, sociability,
physical health and industry. These instruments did not address them-
selves to problems previously outlined in Chapter I, as validity and
reliability data were not presented nor was information on the construction
of these scales evident. Reviews of these scales in Buros' Mental
Measurement Yearbook, Volumes I and VI suggest caution in their use
because of the above shortcomings in their development.

Among the behavior rating scales for which there are some
published details on reliability and validity are: The Children's Guild
Symptom Check List (Eisenberg, Landowne, Wilner, and Inber, 1962);
The Mulligan School Performance Checklist (Mulligan, 1963); The Revised
School Performance Checklist (Mulligan, Douglas, Hammond, and Tizard,
1963); and The Bristol Social Adjustment Scale (Stott, 1963). These
scales have a number of disadvantages such as lack of diagnostic
distinction and a lack of recent and adequate validation. They are
also excessively time consuming.

Specific reviews by individual authors suggest that more recent
behavior rating scales are also lacking in basic requirements of good

scale construction (Dunn, 1967; Powers, 1977; Spivack and Swift, 1973).
The following are typical summaries of three such reviews.

Dunn (1967) in a review of the Child's Behavior Rating Scale
(CBRS) found weaknesses in item construction such that several items
were ambiguous and biased. Odd-even reliability was reported to be
weak at .59 to .87. Empirical validity was not established for the
judgment of psychologists who arranged the items of the CBRS into
categories. The apparent lack of attention to basic scale construction
suggests that the CBRS cannot be recommended for use where decisions
regarding children are to be made.

Powers (1974) in a review of the Vane Kindergarten Scale (VKS)
(Vane, 1968) recommended rejection of this scale on the basis of its
instability over time. Test retest reliability coefficients for this
scale range from .65 to .37. Powers (1969) found other weaknesses in
its ability to predict problem behavior.

Spivack and Swift (1973) reviewed the Teacher's Behavior Rating
Scale (TBRS) by Cowen, Izzo, Miles, Telschow, Trost and Zax (1963) and
found limitations also. Reliability data were found inadequate. There
were also no data reflecting the homogeneity of the scale itself and no
published norms. Because of the above shortcomings the TBRS was also
found inadequate.

The above information represents an overview of data available
and is by no means complete. The interested reader can seek further
information on behavior rating scales in sources such as:

1) Tests and Measurements in Child Development, Volume I (Johnson,

1971) and Volume II (Johnson, 1971);

2) <u>Tests and Measurements in Child Development</u>, Volume II (Johnson and Bommarito, 1976);

3) <u>Socio-emotional Measures</u> (Walker, 1973);

4) <u>Tests in Print</u> (Buros, 1961);

5) <u>Mental Measurement Yearbooks</u>, Volumes I-VII (Buros, 1938-1978).

(2) <u>General Limitations of Available Scales</u>

The previous review pointed out that most behavior rating scales cannot be recommended for classroom use due to poor reliability, validity and because of typically incomplete data. The following section summarizes limitations typically found in behavior rating scales as reported by leading authorities such as Walker (1973), Buros (1961), Johnson (1971), Johnson and Bommarito (1976), and Spivack (1966).

a. <u>General limitations of available scales.</u>

1) Most behavior rating scales lack operationally adequate definitions of the behaviors being measured (Walker, 1973).

2) The items found on behavior rating scales are typically ambiguous, biased, and do not deal with behaviors occurring in the classroom (Spivack, 1966).

3) The items found on behavior rating scales have inadequate levels of discriminative power (Rutter, 1967).

4) The items found on behavior rating scales have had criteria used for item development which are generally inadequate in identifying an individual's status relative to an established standard of performance (Johnson, 1971).

5) The reliability of the behavior rating scales is typically

lower than what is accepted on behavior rating instruments (Buros, 1961).

6) The number of items found on behavior rating scales is typically found to be inappropriate, that is, too much of the teacher's time is needed to complete a form for each child (Cowen, Dorr, and Orgel, 1971).

This brief review points out some of the general limitations of many behavior rating scales. In concluding it must be stressed that such weaknesses must be eliminated if behavior rating scales are to be useful.

(3) Selected Behavior Rating Scales Best Meeting Above Criteria

In an attempt to choose behavior rating scales for this research, it was necessary to review those behavior rating scales regarded as better by experts (Buros, 1961; Walker, 1973; Spivack, 1966) in the field. The authors of these better scales have made an attempt to meet basic scaling requirements and have succeeded to a degree in reducing weaknesses found in behavior rating scales.

A brief review of such behavior rating scales follows. A description of each scale will not be provided at this point; however, an effort has been made to state if reliability, validity data and norms are available. The reader is cautioned that if the behavior rating scales cannot meet standards in the above named areas, they should not be used.

a. The better behavior rating scales (BRS)

The behavior rating scales reviewed are:

1) Emmerick Classroom Observation Rating Scale (ECO) by Emmerick (1971).

2) Peterson Problem Checklist (PPC) by Peterson (1961).

3) Classroom Behavior Inventory (CBI) by Schaefer, Aaronson and Small (1966).

4) Devereaux Elementary School Behavior Rating Scale (DES) by Spivack and Swift (1967).

5) AML Behavior Rating Scale (AML) by Van Vleet (1970).

6) Children's Behavior Questionnaire (CBQ) by Rutter (1967).

To aid in presenting the information on the above instruments the following summary table was developed by the author. Data presented in this table were collected from information provided by the authors of these scales.

Table II.1 indicates whether or not reliability and validity data are available for each scale and whether norms are provided. Reliability data is presented as a score or summary of scores completed on a particular scale. By the use of this table, scales can be pinpointed for further examinations with regard to their weaknesses and strengths.

Of the six scales reviewed all have reliability data available; however, the higher coefficients were on the DES, the AML, and the CBQ. Five scales provide validity studies, one, the PPC, does not. Only one scale is correlated with teacher grades, the DES; two are correlated with academic test scores, the DES and the CBQ; four are correlated with a mental health criteria, the DES, the AML, the CBQ and the ECO; two are correlated with other measures, the DES and the AML; and two have developed norms, the DES and the CBI.

TABLE II.1

SUMMARY TABLE OF THOSE BRS's FOR WHICH NORMS AND RELIABILITY AND VALIDITY DATA ARE AVAILABLE

	RELIABILITY		VALIDITY				NORMS
Behavior Rating Scale	Test Retest	Inter-Rater	Teacher Grade	Mental Health	Academic Test Scores	Other BRS's	
Emmerick (ECO)		.60's[1]		X^2			
Peterson (PPC)		.75					
Schaefer (CBI)	.70's	.70's			X		X
Spivack & Swift (DES)	.80's	.89	X	X	X	X	X
Van Vleet (AML)	.80's			X	X	X	
Rutter (CBQ)	.80's	.70's		X			

1. Number represents reliability scores obtained on these scales.

2. While specific score data were not available, these scores were reported as 'adequate' by test developers.

b. Summary

The review of better behavior rating scales demonstrates, that, in these scales more attention is focused on the principles of good scale construction (i.e., validity, reliability and norms). In spite of this, further studies are needed in the above areas. This research will attempt to meet that need.

Three of the better behavior rating scales were selected for further study.

(4) Instruments Used for this Research

There are virtually no behavior rating scales without some of the limitations presented in Section III; however, the three behavior rating scales with the fewest limitations were chosen for this research. The scales chosen were:

1) The Children's Behavior Questionnaire (CBQ) by Rutter (1967).

2) The Devereaux Elementary School Behavior Rating Scale (DES) by Spivack and Swift (1967).

3) The AML Behavior Rating Scale (AML) by Van Vleet (1970).

a. Description of behavior rating instruments

1. The Children's Behavior Questionnaire (CBQ) was developed to discriminate between children who show "disorders" and those who do not. The scale consists of 26 items each describing a different child behavior. The described behaviors range from motor and affective responses, through specific conduct problems, and immaturities to behaviors indicative of unhappiness, fearfulness and speech difficulties. A three-point scale is used to rate each item. The teacher must use a normal standard in his (her) judgment of each item. No definition of

normal is given to the teacher by Rutter. The total score derived from the instrument can range from 0 to 52 with the lower score indicating more acceptable behavior. Two subscale scores labelled "neurotic" and "antisocial" can also be derived. These purportedly allow the scale to differentiate not only between abnormal and normal children but also between two subclasses of emotionally disturbed children. Test-retest reliability has been established by Rutter (1967) at .89 on the total score and by Richman (1964), using a slightly modified version, at .85. Criterion related validity studies report the scales as reasonably efficient in differentiating a sample of normal children from a group of children attending a psychiatric clinic (Rutter, 1967). No norms are presented and the author does not describe the source of his items or his rationale for selecting them. This scale is easy to use and can be completed in a matter of minutes on each child.

2. The Devereaux Elementary School Behavior Rating Scale (DES) was developed to assess achievement relative to classroom behavior. Each item is rated on either a five-point or a seven-point scale. Eleven subscale scores can be derived. These include scores on categories labelled: classroom disturbance, impatience, disrespect, defiance, external blame, achievement anxiety, external reliance, comprehension, inattentive withdrawal, irrelevant responsiveness, creative initiative, and a need for closeness to the teacher. Test-retest reliability of the factors range from .85 to .91 (Spivack and Swift, 1967). Concurrent validity studies report all factors differentiate between normal and special class children (Spivack and Swift, 1966). Normative data report factor scores which correlate with age, sex and

educational level (Spivack and Swift, 1968). Of the three scales this form requires the most time (approximately ten minutes) to complete.

3. The third scale chosen for this study was the AML Behavior Rating Scale (AML). It was developed for identifying children with learning or behavior problems and contains eleven items. Scores derived range from 11 to 55 with the lower score indicating more acceptable behavior. Three subscale scores can also be derived. Categories for these are: A (aggressive, outgoing behavior), M (moodiness, withdrawal behavior), and L (degree of learning difficulty). Criterion related validity has been established at .84 with other similar measures by Cowen et al. (1971). Reliability studies based on the Kuder Richardson formula report the scale as effective in differentiating children of multiproblem families from children of normal families (Van Vleet et al., 1970). No norms are available (Van Vleet, 1970). This is a very simple device that can be completed in about two minutes.

IV. HYPOTHESES

In attempting to explore the research questions posed in this study (see page 4, Chapter I), the following specific hypotheses have been tested:

1. There is no significant correlation between scores obtained on:

    a) the AML and the DES

    b) the AML and the CBQ

    c) the CBQ and the DES

    d) the AML and the Wide Range Achievement Test, Reading Raw Score

e) the AML and the Wide Range Achievement Test, Spelling Raw Score

f) the AML and the Wide Range Achievement Test, Arithmetic Raw Score

g) the CBQ and the Wide Range Achievement Test, Reading Raw Score

h) the CBQ and the Wide Range Achievement Test, Spelling Raw Score

i) the CBQ and the Wide Range Achievement Test, Arithmetic Raw Score

j) the DES and the Wide Range Achievement Test, Reading Raw Score

k) the DES and the Wide Range Achievement Test, Spelling Raw Score

l) the DES and the Wide Range Achievement Test, Arithmetic Raw Score

m) the AML and the teacher rank (TR)

n) the CBQ and the teacher rank

o) the DES and the teacher rank

2. (a) It is hypothesized that there are no significant differences between scores obtained by those students rated by teachers as high risk (group I_{TR}) and those students rated by teachers as low risk (group II_{TR}) on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS).

(b) It is hypothesized that there are no significant differences between scores obtained by boys and girls on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS).

(c) It is hypothesized that there is no significant interaction between factor I, teacher rating, and factor II, sex of students on the six dependent variables (AML, CBQ, DES, RRS, SRS, and ARS).

3. It is hypothesized that there are no significant differences between scores obtained by high reading achievement students (group 1_{RA}) and low reading achievement students (group II_{RA}) on the three dependent variables (AML, CBQ and DES).

CHAPTER III

METHODOLOGY

I. POPULATION AND SAMPLE

A sample of children was chosen from a population consisting
of all first grade students, who had not failed grade one, from six
rural schools under the Roman Catholic School Board for Conception Bay
Centre. This district is approximately forty miles outside St. John's,
Newfoundland. Sixty-four subjects out of a total population of one
hundred fifty-five students made up the sample.

A table of random numbers (Runyon-Haber, 1971) was used in
selecting subjects. Because sex was included as a variable, an equal
number of boys and girls were used for the study. Thus sixty-four
subjects were chosen, thirty-two girls and thirty-two boys with a mean
age of six years ten months. The total age range was from sixty-six
months to ninety months with a standard deviation of 4.63 months.

II. SELECTION OF THE WIDE RANGE ACHIEVEMENT TEST (WRAT)

The level I Wide Range Achievement Test was selected as the
measurement of achievement. The WRAT was chosen to measure reading,
spelling and arithmetic ability for the following reasons:

1. It is a recent test. Although it was originally prepared
in 1936 by J.F. Jastak, it has undergone four revisions: 1946, 1965,
1976 and 1978. It has been researched on many thousands of persons

from pre-school to advanced old age.

2. The reliability of the test has been established through many testing samples. Typically, reliability scores range from .85 to .98 (Jastak & Jastak, 1978).

3. The test provides raw scores as a measure of reading, spelling and arithmetic abilities. The reading subtest (RRS) consists of recognizing and naming letters and pronouncing words out of context. The spelling subtest (SRS) consists of copying marks resembling letters, writing his/her name and writing single words to dictation. The arithmetic subtest (ARS) contains counting, reading, number symbols, solving oral problems and performing written computations.

4. Administrative time was such that testing could be reasonably completed by the guidance counselor. The time required for testing was between twenty and thirty minutes per child.

III. SPECIFIC PROCEDURE

The specific procedure followed in this study is outlined below:

1. Eight teachers from whose classes the children were chosen ranked, from one to eight, the subjects selected according to the level of risk they felt best described the child based on the definition of risk as stated in Chapter I. This ranking was used to group subjects. Subjects ranked high risk (i.e., numbers 1, 2, 3) were classified as group I_{TR}. Subjects ranked low risk (i.e., numbers 6, 7, 8) were classified as group II_{TR}. Subjects ranked fourth and fifth (i.e., group III_{TR}) were excluded from the study. This resulted in two groups of

twenty-four subjects.

2. Two days after ranking the children, the teachers were given a packet that consisted of copies of three different behavior rating instruments, namely, the AML Behavior Rating Scale, the Children's Behavior Questionnaire, and the Devereaux Elementary School Behavior Rating Scale. Teachers were asked to complete one of each scale on each child. The subject's name had been previously written on the behavior rating scales. These scales were then randomly ordered (Runyon-Haber, 1971) and placed in a packet.

3. Teachers were asked to complete the behavior rating scales in the order presented and according to the names on each form. The teachers were asked to read the instructions provided by each behavior rating scale and to follow the directions as closely as possible. They were given four days to complete these forms. No further instructions were given to the teachers.

4. The writer completed the WRAT on all subjects under study. Reading, spelling and arithmetic scores were obtained on eight children from each class. The spelling subtest was completed in groups of four subjects. Arithmetic and reading subtests were completed on an individual basis in the visitor's office, on the tenth day of May of the school year 1979. Because the WRAT was used to validate the teacher rank as well as the behavior rating scales, grouping of students as to their level of risk was again carried out. This time, however, level of risk was based on the reading raw score of the WRAT and designated RA. The size of groups was set at twenty-four to match the size of the previous groups established on the basis of teacher rank. Thus subjects

forming group I_{RA} (i.e., high risk) obtained reading raw scores of 28
to 42 and subjects forming group II_{RA} (i.e., low risk) obtained reading
raw scores of 48 to 62. To have an even number of subjects in groups
I_{RA} and II_{RA}, one subject who received a score of 42 was deleted randomly
from group I_{RA} and one subject who received a score of 48 was added
randomly to group II_{RA}. All other subjects formed group III_{RA} which
was excluded from further study.

IV. STATISTICAL ANALYSIS

All statistical analyses were based on the SPSS 300 computer
programs (1976). The computer services division at Memorial University
provided the computer hardware. The data were analyzed by using
correlation coefficients, and a one and two way analysis of variance.
Specifically, to assess validity for the total sample of sixty-four
subjects correlations using the Pearson r coefficient of correlation
and Kendall's Tau (τ) coefficient were calculated. Correlations were
computed between all variables (namely, TR, AML, CBQ, DES, RRS, SRS and ARS).
In testing hypothesis I, comparisons between variables (TR, AML, CBQ,
DES, RRS, SRS and ARS) using correlations coefficients were used.

To test hypotheses 2a, 2b, 2c comparisons between high risk
students (group I_{TR}) and low risk students (group II_{TR}) were made using
a two way analysis of variance.

To test hypothesis 3 specific comparisons between high reading
achievement students (group I_{RA}) and low reading achievement students
(group II_{RA}) were carried out using a one way analysis of variance.

Data analysis is explained in the following summary table.

TABLE III.1

SUMMARY OF ANALYSIS USED (INCLUDING SAMPLE SIZES)

Factor	Group	Sex Male (number)	Female (number)	Total	Type of Analysis
	Total Group	32	32	64	Correlations
TR	Group I$_{TR}$	11	13	24	Two way ANOVA
	Group II$_{TR}$	13	11	24	
RA	Group I$_{RA}$	11	13	24	One way ANOVA
	Group II$_{RA}$	12	12	24	

The level of significance used for rejection of the null hypothesis was $\alpha \leq .05$.

The following table (III.2) further summarizes the statistical tests used to answer specific research hypotheses.

TABLE III.2

HYPOTHESES, STATISTICAL MEASURES USED

Hypothesis	Statistical Measure Used
1. There are no significant correlations between scores obtained by a) the three behavior rating scales, b) teacher rank, and c) the Wide Range Achievement Test scores.	Correlations Pearson r correlation and Kendall's Tau correlation.[1]

(cont'd.)

Table III.2 (cont'd.)

Hypothesis	Statistical Measure Used
2a. There are no significant differences between scores obtained by high risk students (group I_{TR}) and low risk students (group II_{TR}) on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS).	A--main effects (TR)
2b. There are no significant differences between scores obtained by boys and girls on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS).	B--main effects (SEX) two way ANOVA
2c. There is no significant interaction between factor I, teacher ranking, and factor II, sex of students, on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS).	AxB--interaction effect two way ANOVA
3. There are no significant differences between scores obtained by high reading achievement students (group I_{RA}) and low reading achievement students (group II_{RA}) on the three dependent variables (AML, CBQ and DES).	A--main effects (RA) one way ANOVA

1. Kendall Tau's coefficient was used in all correlations involving teacher rank (Glass and Stanley, 1970).

CHAPTER IV

STATISTICAL FINDINGS

I. STATISTICAL FINDINGS

The data gathered for analysis of this study consisted of the scores on the three behavior rating scales and on the Wide Range Achievement Test. The means (M), ranges (R), and standard deviations (SD) of the subjects' scores for the total group, groups I and II (teacher rank) and groups I and II (reading achievement) are presented in Tables IV.1 to IV.6.

TABLE IV.1

DESCRIPTIVE STATISTICS FOR THE AML BEHAVIOR RATING SCALE

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	18.73	11-47	7.85
GROUP I$_{TR}$	24	21.46	11-47	9.35
Male	11	22.18	11-47	9.82
Female	13	20.85	13-41	9.29
GROUP II$_{TR}$	24	15.79	11-30	5.99
Male	11	16.69	11-26	5.05
Female	13	15.72	11-30	5.50
GROUP I$_{RA}$	24	22.13	11-47	10.62
Male	11	22.72	11-47	10.06
Female	13	21.62	11-42	11.45
GROUP II$_{RA}$	24	15.75	11-30	4.54
Male	11	14.64	11-24	3.80
Female	13	16.69	12-30	5.04

TABLE IV.2

DESCRIPTIVE STATISTICS FOR THE CHILDREN''S BEHAVIOR QUESTIONNAIRE

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	4.90	0-28	5.36
GROUP I$_{TR}$	24	7.46	0-28	6.90
Male	11	8.27	1-28	8.51
Female	13	6.07	0-17	5.58
GROUP II$_{TR}$	24	3.46	0-17	3.58
Male	13	3.62	0-8	2.63
Female	11	3.27	0-17	4.58
GROUP I$_{RA}$	24	6.88	0-28	7.21
Male	11	8.81	1-28	8.46
Female	13	5.23	0-17	6.00
GROUP II$_{RA}$	24	3.12	0-16	3.53
Male	11	2.27	0-7	2.24
Female	13	3.85	0-16	4.30

TABLE IV.3

DESCRIPTIVE STATISTICS FOR THE DEVEREAUX ELEMENTARY SCHOOL
BEHAVIOR RATING SCALE

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	109.42	77-216	28.42
GROUP I_{TR}	24	124.00	88-216	36.66
Male	11	125.91	94-216	36.28
Female	13	122.38	88-203	38.37
GROUP II_{TR}	24	101.33	77-143	16.97
Male	13	100.00	77-132	16.27
Female	11	102.9	83-143	18.43
GROUP I_{RA}	24	120.58	84-216	38.54
Male	11	127.09	85-216	37.73
Female	13	116.54	84-203	39.61
GROUP II_{RA}	24	99.75	80-154	17.99
Male	11	95.55	86-119	11.47
Female	13	103.30	80-154	21.93

TABLE IV.4

DESCRIPTIVE STATISTICS FOR THE READING RAW SCORE OF THE
WIDE RANGE ACHIEVEMENT TEST

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	45.47	28–51	7.24
GROUP I$_{TR}$	24	40.54	28–51	6.67
Male	11	38.64	28–46	7.17
Female	13	42.15	31–50	6.03
GROUP II$_{TR}$	24	49.08	37–62	6.61
Male	11	48.76	37–62	5.64
Female	13	49.45	39–61	7.86
GROUP I$_{RA}$	24	38.20	28–42	3.97
Male	11	37.72	28–42	5.04
Female	13	38.62	33–42	3.35
GROUP II$_{RA}$	24	52.54	48–62	4.15
Male	11	52.36	48–62	4.08
Female	13	52.69	49–61	4.29

TABLE IV.5

DESCRIPTIVE STATISTICS FOR THE SPELLING RAW SCORE OF THE WIDE RANGE ACHIEVEMENT TEST

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	29.55	18–37	3.67
GROUP I$_{TR}$	24	27.88	18–35	4.16
Male	11	27.90	20–35	4.18
Female	13	27.85	18–33	4.14
GROUP II$_{TR}$	24	30.92	24–37	3.21
Male	13	30.92	24–37	3.20
Female	11	30.90	26–37	3.30
GROUP I$_{RA}$	24	26.54	18–31	3.40
Male	12	26.73	20–31	3.66
Female	12	26.38	18–31	3.31
GROUP II$_{RA}$	24	32.04	23–37	2.88
Male	11	31.91	23–37	3.67
Female	13	32.15	29–37	2.15

TABLE IV.6

DESCRIPTIVE STATISTICS FOR THE ARITHMETIC RAW SCORE OF THE
WIDE RANGE ACHIEVEMENT TEST

	Total Number	Mean	Range	Standard Deviation
TOTAL GROUP	64	22.52	16-26	2.28
GROUP I$_{TR}$	24	21.67	16-26	2.48
Male	11	21.63	16-26	3.14
Female	13	21.69	19-25	1.89
GROUP II$_{TR}$	24	23.25	17-26	1.96
Male	13	23.52	19-25	1.61
Female	11	22.90	17-26	2.34
GROUP I$_{RA}$	24	21.50	16-26	2.64
Male	11	21.36	16-26	3.14
Female	13	21.62	18-24	2.26
GROUP II$_{RA}$	24	23.75	21-26	1.36
Male	11	24.27	22-26	1.61
Female	13	23.30	21-26	1.44

II.   ANALYSIS OF HYPOTHESES / TABLE BY TABLE

Hypothesis 1. There are no significant correlations between

scores obtained on: a) the three behavior

rating scales (AML, CBQ, DES), b) the Wide

Range Achievement Test score (Reading,

Spelling, Arithmetic), and c) teacher rank.

Tables IV.7 and IV.8 present correlations used in testing this

hypothesis. The three behavior rating scales (AML, CBQ, DES) and the

Wide Range Achievement Test scores (Reading, Spelling, Arithmetic) are

compared using the Pearson coefficient (r) in Table IV.7.

TABLE IV.7

CORRELATION MATRIX FOR AML, CBQ, DES, READING RAW SCORE,
SPELLING RAW SCORE, ARITHMETIC RAW SCORE

$N = 64$	AML	CBQ	DES	RRS	SRS	ARS
AML	1.00	0.84*	0.81*	−0.45*	−0.52*	−0.44*
CBQ		1.00	0.79*	−0.43*	−0.54*	−0.37**
DES			1.00	−0.42*	−0.57*	−0.47*
RRS				1.00	0.80*	0.58*
SRS					1.00	0.57*
ARS						1.00

*$p \leq .001$
**$p \leq 0.01$

The AML, the CBQ, and the DES, correlate significantly as pre-sented in Table IV.7. The null hypothesis was rejected at the pre-determined level of significance p ≤ .05 for the three behavior rating scales, and the Wide Range Achievement Test scores. The highest correlation (r = .84) exists between the AML and the CBQ. The lowest correlation (r = .79) exists between the CBQ and the DES.

The AML, the CBQ and the DES correlate significantly with the RRS, the SRS, and the ARS. The null hypothesis was rejected at the predetermined level of significance (p ≤ .05). The highest correlations exist between the DES and the RRS (r = -.43); the DES and the SRS (r = -.57); and the DES and the ARS (r = -.47). The lowest correlations exist between the CBQ and the RRS (r = -.43); the CBQ and the SRS (r = -.54); and the CBQ and the ARS (r = -.37). The AML, the CBQ, and the DES correlate most highly with the SRS at -0.52, -0.54 and -0.57 respectively.

Table IV.8 presents data on the three behavior rating scales and teacher rank using Kendall's Tau (τ) coefficient of correlation.

TABLE IV.8

CORRELATIONS OF TEACHER RANK (TR) WITH BEHAVIOR RATING SCALES (AML, CBQ, DES) AND WITH WIDE RANGE ACHIEVEMENT TEST SCORES (RRS, SRS AND ARS) USING KENDALL'S CORRELATION COEFFICIENT

	AML	CBQ	DES	RRS	SRS	ARS
TR	-.23**	-.23**	-.24**	.40*	.29*	.30*

*p ≤ .001
**p ≤ .01

The AML, the CBQ and the DES correlate significantly with teacher rank.

The null hypothesis was rejected at the predetermined level of significance ($p \leq .05$) in comparisons between teacher rank and the three behavior rating scales.

The RRS, SRS and the ARS correlate significantly with the teacher rank.

The null hypothesis was rejected at the predetermined level of significance ($p \leq .05$) in comparisons between the Wide Range Achievement Test scores and teacher rank.

Table IV.9 shows analysis of variance data on the AML for groups I_{TR} and II_{TR}. Hypothesis 2a [There are no significant differences between scores obtained by high risk students (group I_{TR}) and low risk students (group II_{TR}) on the AML.] was rejected ($F \leq 5.647$, $p \leq .022$). Hypothesis 2b [There are no significant differences between scores obtained by boys and girls of groups I_{TR} and II_{TR} on the AML.] was accepted ($F \leq .266$, $p \leq .609$). Hypothesis 2c [There are no significant interactions between teacher ranking and sex of students on the AML.] was accepted.

Table IV.10 shows analysis of variance data on the CBQ for groups I_{TR} and II_{TR}. Hypothesis 2a [There are no significant differences between scores of groups I_{TR} and II_{TR} on the CBQ.] was rejected ($F \leq 6.731$, $p \leq .13$). Hypothesis 2b [There was no significant differences between scores obtained by girls and boys of groups I_{TR} and II_{TR} on the CBQ] was accepted ($F \leq 1.106$, $p \leq .299$). Hypothesis 2c [There is no significant interaction between teacher ranking and sex of students on

TABLE IV.9

ANALYSIS OF VARIANCE RESULTS OF THE AML FOR GROUPS I_A AND II_A

Source	S.S.	D.F.	M.S.	F	p
Main Effects	341.290	2	170.645	2.874	0.067
Teacher Rank	335.269	1	335.269	5.647	0.022[1]
Sex of Subject	15.769	1	15.769	0.266	0.609
Interactions	0.409	1	0.409	0.007	0.934
Explained	341.699	3	113.900	1.918	
Residual	2612.276	44	59.370		
Total	2953.975	47	62.851		

1. Underlined scores are significant at the predetermined level ($p \leq .05$).

TABLE IV.10

ANALYSIS OF VARIANCE RESULTS OF THE CBQ FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	225.566	2	112.783	3.717	0.032[1]
Teacher Rank	204.233	1	204.233	6.731	0.013
Sex of Subject	33.566	1	33.566	1.106	0.299
Interactions	21.259	1	21.259	0.701	0.407
Explained	246.826	3	82.275	2.712	
Residual	1135.087	44	30.343		
Total	1581.913	47	33.658		

1. Underlined scores are significant at the predetermined level ($p \leq .05$).

the CBQ] was accepted ($F \le .701$, $p \le .407$):

Table IV.11 shows analysis of variance data on the DES for groups I_{TR} and II_{TR}. Hypothesis 2a [There are no significant differences between groups I_{TR} and II_{TR} on the DES.] was rejected ($F \le 7.217$, $p \le .01$). Hypothesis 2b [There are no significant differences between scores obtained by girls and boys of groups I_{TR} and II_{TR} on the DES.] was accepted ($F \le .001$, $p \le .971$). Hypothesis 2c [There is no significant interaction between teacher ranking and sex of subjects on the DES.] was accepted ($F \le .145$, $p \le .705$).

TABLE IV.11

ANALYSIS OF VARIANCE RESULTS OF THE DES FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	6166.461	2	3083.230	3.626	0.035[1]
Teacher Rank	6136.375	1	6136.375	7.217	0.010
Sex of Subject	1.128	1	1.128	0.001	0.971
Interactions	123.309	1	123.309	0.145	0.705
Explained	6289.770	3	2096.590	2.466	
Residual	37412.824	44	850.291		
Total	43702.594	47	929.842		

1. Underlined scores are significant at the predetermined level ($p \le .05$).

Table IV.12 shows analysis of variance data on the reading raw score for groups I_{TR} and II_{TR}. Hypothesis 2a [There are no significant

differences between scores obtained by groups I_{TR} and II_{TR} on the reading raw score of the Wide Range Achievement Test.] was rejected ($F \leq 17.569$, $p \leq .000$). Hypothesis 2b [There are no significant differences between scores obtained by boys and girls on the reading raw score of the Wide Range Achievement Test.] was accepted ($F \leq .548$, $p \leq .463$). Hypothesis 2c [There is no significant interaction between teacher ranking and sex of students on the RRS.] was accepted ($F \leq .147$, $p \leq .704$).

TABLE IV.12

ANALYSIS OF VARIANCE RESULTS OF THE RRS FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	776.098	2	388.049	8.862	0.001[1]
Teacher Rank	769.348	1	769.348	17.569	0.000
Sex of Subject	24.015	1	24.015	0.548	0.463
Interactions	6.425	1	6.425	0.147	0.704
Explained	782.523	3	260.841	5.957	
Residual	1926.724	44	43.789		
Total	2709.246	47	57.644		

1. Underlined scores are significant at the predetermined level ($p \leq .05$).

Table IV.13 shows analysis of variance data on the spelling raw score for groups I_{TR} and II_{TR}. Hypothesis 2a [There was no significant differences between scores obtained by groups I_{TR} and II_{TR} on the spelling

raw score of the Wide Range Achievement Test.] was rejected (F $\leq$ 7.63, p $\leq$ .008). Hypothesis 2b [There are no significant differences between scores obtained by boys and girls on the spelling raw score of the Wide Range Achievement Test.] was accepted (F $\leq$ .001, p $\leq$ .972). Hypothesis 2c [There is no significant interaction between teacher rating and sex of students on the SRS.] was accepted (F $\leq$ .000, p $\leq$ .982).

TABLE IV.13

ANALYSIS OF VARIANCE RESULTS OF THE SRS FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	111.038	2	55.519	3.850	0.029[1]
Teacher Rank	111.018	1	110.018	7.630	0.008
Sex of Subject	0.018	1	0.018	0.001	0.972
Interactions	0.007	1	0.007	0.000	0.982
Explained	111.046	3	37.015	2.567	
Residual	634.429	44	14.419		
Total	745.475	47	15.861		

1. Underlined scores are significant at the predetermined level (p $\leq$ .05).

Table IV.14 shows analysis of variance data on the arithmetic raw score of groups I_{TR} and II_{TR}. Hypothesis 2a [There are no significant differences between scores obtained by groups I_{TR} and II_{TR} on the arithmetic raw score of the Wide Range Achievement Test.] was rejected.

(F $\leq$ 5.606, p $\leq$ .022). Hypothesis 2b [There are no significant differences between scores obtained by boys and girls on the arithmetic raw score of the Wide Range Achievement Test.] was accepted (F $\leq$ .190, p $\leq$ .665). Hypothesis 2c [There is no significant interaction between teacher ranking and sex of students on the ARS.] was accepted (F $\leq$ .271, p $\leq$ .605).

TABLE IV.14

ANALYSIS OF VARIANCE RESULTS OF THE ARS FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	31.063	2	15.980	3.004	0.060
Teacher Rank	28.980	1	28.980	5.606	0.022[1]
Sex of Subject	0.980	1	0.980	0.190	0.665
Interactions	1.399	1	1.399	0.271	0.605
Explained	32.462	3	10.821	2.093	
Residual	227.454	44	5.169		
Total	259.916	47	5.530		

1. Underlined scores are significant at the predetermined level (p $\leq$ .05).

Table IV.15 presents analysis of variance results on the AML variable for groups I_{RA} and II_{RA}. Hypothesis 3 [There are no significant differences between scores obtained by high reading achieving students (group I_{RA}) and low reading achieving students (group II_{RA}) on the AML.] was rejected (F $\leq$ 7.314, p $\leq$ .01).

TABLE IV.15

ANALYSIS OF VARIANCE RESULTS OF THE AML FOR GROUPS I_{TR} AND II_{TR}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	487.687	1	487.687	7.314	0.010[1]
Reading Achievement	487.687	1	487.687	7.314	0.010
Explained	487.687	1	487.687	7.314	
Residual	3067.121	46	66.677		
Total	3554.808	47	75.634		

1. Underlined scores are significant at the predetermined level ($p \leq 05$).

Table IV.16 shows analysis of variance data on the CBQ for groups I_{RA} and II_{RA}. Hypothesis 3 [There are no significant differences between scores obtained by high reading achieving students (group I_{RA}) and low reading achieving students (group II_{RA}) on the CBQ.] was rejected ($F \leq 5.241$, $p \leq .027$).

TABLE IV.16

ANALYSIS OF VARIANCE RESULTS OF THE CBQ FOR GROUPS I_{RA} AND II_{RA}

Source	S.S.	D.F.	M.S.	F	P
Main Effects	168.750	1	168.750	5.241	0.027[1]
Reading Achievement	168.750	1	168.750	5.241	0.027
Explained	168.750	1	168.750	5.241	
Residual	1481.246	46	32.201		
Total	1649.996	47	35.106		

1. Underlined scores are significant at the predetermined level ($p \leq .05$).

Table IV.17 shows analysis of variance data on the DES for groups I_{RA} and II_{RA}. Hypothesis 3 [There are no significant differences between scores obtained by high reading achieving students (group I_{RA}) and low reading achieving students (group II_{RA}) on the DES.] was rejected ($F \leq 6.234$, $p \leq .016$).

TABLE IV.17

ANALYSIS OF VARIANCE RESULTS OF THE DES FOR GROUPS I_{RA} AND II_{RA}

Source	S.S.	D.F.	M.S.	F	p
Main Effects	5590.082	1	5590.082	6.234	0.016[1]
Reading Achievement	5590.082	1	5590.082	6.234	0.016
Explained	5590.082	1	5590.082	6.234	
Residual	41245.758	46	898.647		
Total	46835.840	47	996.507		

1. Underlined scores are significant at the predetermined level ($p \leq .05$).

SUMMARY

This chapter has presented the statistical analysis of the data for the study. The correlations proved statistically significant in the comparisons of all variables and combinations of variables.

The analyses of variance produced several significant F scores. These were: 1) in the main effects between scores obtained by high risk students (group I_{TR}) and low risk students (group II_{TR}) on the three

behavior rating scales and on the arithmetic, spelling and reading raw scores of the Wide Range Achievement Test; 2) in the main effects between scores obtained by high reading achievement students (group I_{TR}) and low reading students (group II_{TR}).

CHAPTER V

DISCUSSION OF RESULTS AND CONCLUSIONS

I. INTRODUCTION

This study investigated the usefulness of teacher completed Behavior Rating Scales as a means of detecting high and low risk students. The study also sought to find out which, if any, scale best assessed potential risk. Three behavior rating scales were chosen for exploration; these were the AML Behavior Ration Scale, the Children's Behavior Questionnaire and the Devereaux Elementary School Behavior Rating Scale. The Wide Range Achievement Test Battery was also chosen as an already established standard of performance against which to compare the behavior rating scales. Subjects of this research were studied as a total group and then were categorized as high and low risk by teacher rank and also by reading achievement subtests of the Wide Range Achievement Test.

In this chapter the statistical findings are discussed relative to the three research questions asked in Chapter I.

II. RESEARCH QUESTION ONE

1. Do behavior rating scales correlate significantly:
   a) with each other
   b) with teacher rank
   c) with Wide Range Achievement Test scores?

44

Findings (Table IV.7) indicate that all three behavior rating scales correlated significantly with each other. The highest correlation (r = .84) was between the AML and the CBQ, and the lowest correlation (r = .79) was between the CBQ and the DES. This is not surprising since items on the AML and the CBQ are similar; they are based on a mental health criterion, whereas items on the CBQ and the DES appear to differ somewhat; items on the DES are based on school behavior as pointed out in Chapter II, Section 11.a.

All three behavior rating scales correlate significantly with teacher rank (Table IV.8). The CBQ and the AML have the same correlation (r = -.23) with teacher rank and the DES has a correlation of (r = -.24) with teacher rank. These correlations were negative because those subjects who receive low ranks of 1, 2, or 3 were considered high risk by teachers while those subjects who receive high scores on the behavior rating scales were considered high risk. The fact that the three behavior rating scales correlate with teacher rank might be expected as both scores on a given child were obtained from the same person.

Teacher rank correlated significantly with all three subtests of the Wide Range Achievement Test (Table IV.8), having the reading raw score with a much higher correlation coefficient (r = .40) than the spelling raw score (r = .29) or the arithmetic raw score (r = .30). The relationship between teacher rank and the reading raw score might be expected because of the emphasis placed on reading at the grade one level).

The fact that teacher rank correlates significantly with the behavior rating scales and with the subscales of the Wide Range Achieve-

ment Test leads one to suggest that teacher ratings and rankings are accurate means of judging children who are likely to have school difficulties at least as defined by the Wide Range Achievement Test. Based on these judgments the three behavior rating scales might be a plausible way of receiving information on children who are at risk of failure. These instruments may not be more acceptable than teacher rating in categorizing children; however, they can provide useful clinical information to the counselor. The results suggest that the three behavior rating scales when used on whole classes by teachers can be the first step in identifying at risk children and thus could be the first step in a preventative program rather than a corrective type program.

The three behavior rating scales correlate significantly with all three subscores on the Wide Range Achievement Test (Table IV.7). Negative correlations result because those who score high on the reading raw score and the spelling raw score, and the arithmetic raw score, score low on the AML, the DES and the CBQ. Specifically the lowest correlation ($r = -.37$) is between the CBQ and the arithmetic raw score; however, this is a significant correlation at the .01 level of confidence. The highest correlations exist between the SRS and the three behavior rating scales. These correlations are -.52, -.54, and -.57 respectively. The DES correlates highest with the achievement scores at -.57, -.47, and -.37. This might be expected since the DES was developed to identify achievement related to classroom behavior, whereas the CBQ and the AML were developed to assess disorders or learning problems and were based on a mental health criterion (see Chapter II, section IV).

Because all three behavior rating scales correlate significantly with one another, with the Wide Range Achievement Test scores and teacher rank, the next question to consider is 'which scale is best?' The following table, Table V.1, presents a summary of correlations between the behavior rating scales, the Wide Range Achievement Test, and teacher rank in an attempt to answer this question.

Results show that no one behavior rating scale stands out over the other two. The DES correlates highly with the Wide Range Achievement Reading subtest.

TABLE V.1

BEHAVIOR RATING SCALE'S CORRELATIONS WITH THE WIDE RANGE ACHIEVEMENT TEST[1], TEACHER RANK[2] AND OTHER SCALES[1]

Scale	RRS	SRS	ARS	TR	CBQ	AML	DES
DES	-.43	-.57	-.47	-.24	.79	.81	-
AML	-.45	-.52	-.44	-.23	.84	-	-
CBQ	-.43	-.54	-.37	-.23	-	-	-

1. Pearson correlation was used to compare BRS's with WRAT and other scales.
2. Kendall's Tau was used to compare TR with the behavior rating scales.

The AML correlates higher with the CBQ and the DES, and the DES correlates highly with teacher rank. It appears that because all three scales correlate significantly and highly with each other, with the WRAT and the teacher rank, no one scale can be recommended over another. Thus,

in choosing one of these behavior rating scales for use it is recom-
mended that one make his/her choice based on 1) the suitability of the
behavior rating scale to the specific purpose intended in its use, and
2) the amount of time available to those completing the behavior rating
scale.

III. RESEARCH QUESTION TWO

Are high and low risk students as determined by rankings made
by teachers rated significantly different on teacher completed behavior
rating scales (AML, DES and CBQ)? In order to answer this question a
null hypothesis design format was used. Hypotheses 2a, 2b and 3 were
aimed at exploring this question.

(1) Hypothesis 2a [There are no significant differences between
scores obtained by high risk students (group I_{TR}) and low risk students
(group II_{TR}) on the six dependent variables (AML, CBQ, DES, RRS, SRS,
ARS)]. Tables IV.9 through IV.14 presented these results statistically.
The null hypothesis was rejected for all three behavior rating scales
at or below the predetermined level of significance (p $\leq$ .05). Such a
finding reflects the view proposed by Yellott et al. (1969) who stated
that behavior rating by teachers do not merely reflect teacher biases.
The null hypothesis was rejected for all three subscores of the WRAT at
the predetermined level of significance (p $\leq$ .05). (Tables IV.12, IV.13
and IV.14). High risk students (group I_{TR}) and low risk students
(group II_{TR}) show significant differences in scores on reading, spelling
and arithmetic with the reading scores showing the greatest significant
difference. The most probable explanation of these results is that

subjects classified high risk (group I_{TR}) by teachers do more poorly in reading, spelling and arithmetic than do subjects classified as low risk (group II_{TR}) showing clearly that teacher judgment as to level of risk on these behavior rating scales is consistent with findings on other behavior rating scales as reported in Table II.1.

(2) Hypothesis 2b [There are no significant differences between scores obtained by boys and girls on the six dependent variables (AML, CBQ, DES, RRS, SRS and ARS]. Tables IV.9 through Table IV.14 reported these findings. The null hypothesis was accepted. The failure of this study to find differences related to the sex of the subject does seem at variance with the findings in the literature reported by authors such as Wickman (1928), Rogers (1942), Ullman (1952), and Bower (1969) who found male subjects having more problems than female subjects in the elementary grades. A closer look at scores reveal a tendency throughout for male subjects to score higher than female subjects on the three behavior rating scales. No clear cut explanation of this outcome is readily available. One might be tempted to conclude that sex differences were controlled out of these instruments; however, no evidence of this was found.

(3) Hypothesis 2c [There is no significant interaction between teacher ranking and sex of students on the six dependent variables (AML, CBQ, DES, RRS, SRS, ARS]. (Tables IV.9 through IV.14) was accepted. An interaction effect was not found nor was there any indication of such an interaction effect reported in the literature searched.

IV. RESEARCH QUESTION THREE

Are high and low risk students as determined by scores obtained
on the Wide Range Achievement reading sub-test significantly different
on teacher completed behavior rating scales (AML, CBQ and DES)?

In order to answer this question a null hypothesis design format
was used. Hypothesis five was aimed at exploring this question.

(1) Hypothesis 3 [There is no significant difference between
scores obtained by high reading achievement students (group I_{RA}) and
low reading achievement students (group II_{RA}) on the three dependent
variables (DES, CBQ and AML)] was rejected at or below the predetermined
level of significance ($p \leq .05$) for the DES, CBQ and AML.   Thus, the
DES, CBQ and AML can determine differences between high and low reading
achievement subjects.

This finding that the AML, CBQ and DES reflect differences in
high and low reading achievement students adds support to similar
findings when subjects were split by teacher rank. Thus it appears
behavior rating scales can differentiate between high and low risk
students.

V. CONCLUSIONS

A number of conclusions were drawn from the preceding discus-
sion of the results. They are the following:

1. This study appears to add support to the view that selected
behavior rating scales are useful tools for detecting children who have
or are likely to have school difficulties.

2. There are available to the counselor selected behavior

rating scales which when completed by the teacher can be a useful means of identifying potentially high risk children.

3. These instruments correlated significantly with the Wide Range Achievement Test, thus it appears these three behavior rating scales relate to academic achievement.

4. No one scale appears better than another; however, in choosing one of these scales for use, your choice should be based on 1) suitability of the behavior rating scale to the specific purpose intended in its use, 2) the amount of time available to those completing the behavior rating scale, and 3) the qualities of the behavior rating scale as outlined in Chapter II.

VI. RECOMMENDATIONS FOR FUTURE RESEARCH

As a result of this study, other possibilities for research present themselves. Principally, the following areas could be explored:

1. This is but a single study demonstrating the usefulness of three behavior rating scales. Additional replications on other populations need be carried out.

2. Similar studies assessing other behavior rating scales need be carried out.

3. Additional studies need be carried out using different criteria than used in this study to assess the power of comparison with teacher grades, peer ratings, parent ratings, etc.

4. Studies need be carried out on variables such as sex and intelligence as influencing factors in determining an individual's score on behavior rating scales in relation to his achievement.

5. Studies need be carried out on the three behavior rating scales using special education students as a criterion group against "normal" children.

6. Studies of a long term nature need be carried out to further assess the validity of these behavior rating scales.

7. Studies need be carried out to assess if the three behavior rating scales are more useful than teacher ranking alone.

VII. SUMMARY

Several things have stood out as a result of this study. Firstly, this study lends support to the fact that behavior rating scales are useful devices for detecting children who are likely to have school difficulties. Secondly, the results of this study indicate no one of the behavior rating scales used in this study stands out more than the other; the three behavior rating scales appear of equal worth. Thirdly, it appears that the three behavior rating scales can detect high and low risk children when used by classroom teachers. Fourthly, these three behavior rating scales do not just reflect teacher biases.

In conclusion the behavior rating scales used by the guidance counselor to collect descriptive information on high risk children from the teacher leads one to believe that a large number of potentially high risk children might be screened at an early age. If so, such children could be directed toward early help by the classroom teacher, or where necessary, by the guidance counselor, thus initiating a more preventative than corrective type program.

BIBLIOGRAPHY

Bower, Eli M. Early Identification of Emotionally Handicapped Children in School (2nd ed.). Springfield, Ill.: Thomas, 1969.

Buehler, C. From Birth to Maturity. London: Routledge and Kegan Paul, 1935.

Buros, O. K. Mental Measurement Yearbook. Vols. I-VIII, 1938-1978.

Buros, O. K. Tests in Print. Highland Park, New Jersey: Gryhon Press, 1961.

Cornell, E., Coxe, W., and Orleans, J. New York Rating Scale for School Habits. New York: Harcourt, Brace and World, 1927.

Cossell, R. The Child's Behavior Rating Scale. Beverly Hills, California: Western Psychological Services, 1962.

Cowen, E. L., Dorr, D. A., Orgel, A. R. Interrelation among Screening Measures for Early Detection of School Dysfunction, Psychology in the Schools, 1971, Vol. 1, No. 1, pp. 135-139.

Cowen, E. L., Izzo, L. D., Miles, H., Telschow, E. F., Trost, M. A., and Fox, M. A Preventative Mental Health Program in the School Setting: Description and Evaluation. Journal of Psychology, Vol. 56, pp. 307-356, 1963.

Cullen, K. and J., and Bounty, A. P. The Prevention of Behavior Disorders in the Children of 1000 Western Australian Families. Medical Journal of Australia, 1966, Vol. 2, pp. 805-808.

Danziger, L. Cited in Kohn, M. "Competence and symptom factors in the preschool child." Unpublished manuscript, 1968. Available from Wm. Hanson White Institute, New York 10023.

Davidson, H. H., and Greenberg, J. W. School Achievers from a Deprived Background. Report, Project No. 2805, Department of Health, Education and Welfare, 1967.

Dunn, J. The Child's Behavior Rating Scale by Cossell, R. Report cited in Buros, 7th Mental Measurement Yearbook, 1962.

Education Services Bureau, Pupil Adjustment Inventory. New York: Houghton, Mifflin, 1957.

Eisenberg, L., Landowne, E. J., Wilmer, D. M., and Inber, S. O. The Use of Teacher Ratings in a Mental Health Study: A Method for

Measuring the Effectiveness of a Therapeutic Nursery Program. American Journal of Public Health, Vol. 52, pp. 18-28, 1962.

Emmerick, W. Structure and Development of Personal-Social Behavior in Economically Disadvantaged Preschool Children. Genetic Psychological Monographs. (1971).

Haggerty, M. E. The Incidence of Undesirable Behavior in Public School Children. Journal of Educational Research, 1952, Vol. 12, pp. 102-122.

Harris, F. D. Emotional Blocks to Learning. Glencoe, Ill.: Free Press, 1961.

Jastak, J. F., and Jastak, S. Wide Range Achievement Test Manual. Washington: Jastak Associates, 1978.

Johnston, O. Tests and Measurements in Child Development: Handbook I and II. Washington: Jossey-Boss Publishers, 1971.

Johnston, O., and Bommarito, J. Tests and Measurements in Child Development: A Handbook. Washington: Jossey-Boss Publishers, 1971.

Kim, Y., Anderson, H. E., Jr., and Bashaw, W. L. Social Maturity, Achievement, and Basic Ability. Educational and Psychological Measurement, 1968, 28, 535-543.

Kohn, M. "Competence and Symptom Factors in the Pre School Child." Unpublished manuscript, 1968, available from Alanson, W., White Institute, New York 10023.

Kohn, M., and Rosman, B. L. Cognitive Functioning in Five Year Old Boys as Related to Social-Emotional and Background Variables, Developmental Psychology, 1973, 8, 277-294.

Linton, M., Gallo, P. J. The Practical Statistician Simplified Handbook. California: Brooks/Cole Publishing Corp., 1975.

Mulligan, G. Personal Communication, cited in Ryle, A., Pond, D. A., Hamilton, M. (1965), The Prevalence and Pattern of Psychological Disturbance in Children of Primary Age. Journal of Child Psychology and Psychiatry, 1963.

Mulligan, G., Douglas, J. Hammond, W., and Tizard, J. Delinquency and Symptoms of Maladjustment: The Finding of a Longitudinal Study, Proc. R. Social Medicine, 1963, 56, 1083-1086.

Miller, L. School Behavior Checklist: An Inventory of Deviant Behavior for Elementary School Children. Journal of Consulting and Clinical Psychology, 1972, 38, pp. 134-144.

Olson, W. C. Problem Tendencies in Children. Minneapolis: University of Minnesota Press, 1930.

Peterson, D.R. Behavior Problems of Middle Childhood. Journal of Consulting Psychology, 1961, Vol. 125, pp. 205-209.

Powers, S.M. The Validity of the Vane Kindergarten Test in Predicting Achievement in Kindergarten and First Grade, Educational and Psychological Measurement, 1974, Vol. 34, pp. 1003-1007.

Powers, S.M. The Vane Kindergarten Test: Temporal Stability and Ability to Predict Behavior Criteria. Psychology in the Schools, 1977, January, Vol. 14(1), pp. 34-36.

Pritchard, D.G. Education and the Handicapped. London, Routledge and Paul, 1963.

Richman, N. "The Prevalence of Psychiatric Disturbance in a Hospital School for Epileptics." Unpublished D.P.M. dissertation, University of London, 1964.

Richards, H.C., and McCandless, B.R. Socialization Dimensions among Five Year Old Slum Children. Cited in M. Roff and D.F. Richards (Eds.), Life History Research in Psychopathology, Vol. 1. Minneapolis: University of Minnesota Press, 1970.

Rogers, C.A. Mental Health Findings in the Elementary School. Education Review Bulletin, Ohio State University, Vol. 21:3 1942.

Runyon, R., and Haber, A. Fundamentals of Behavioral Statistics, 2nd ed. Don Mills, Ontario: Addison-Wesley Publishing Corp., 1971.

Rutter, M. A Children's Behavior Questionnaire for Completion by Teachers: Preliminary Findings, Journal of Child Psychology and Psychiatry, Vol. 8, pp. 1-11, 1967.

Rutter, M. Classification and Categorization in Child Psychiatry. International Journal of Psychiatry, Vol. 3, pp. 161-167, 1967.

Rutter, M., Tizard, J. and Whitmore, K. Education, Health and Behavior. New York: Wiley, 1970.

Schaefer, E.S., Aaronson, M.R. Classroom Behavior Inventory. Unpublished manuscript, 1966 (available School of Public Health, University of North Carolina, Chapel Hill, N.C.).

Slaniagan, J. Personal and Social Development Program. New York: Scientific Research Associates, 1956.

Spivack, G., and Swift, M. The Devereaux Elementary School Behavior Rating Scale: A Study of the Nature and Organization of Achievement Related to Disturbed Classroom Behavior. Journal of Special Education, 1966, Vol. 1, pp. 71-91.

Spivack, G., and Swift, M.   Devereaux Elementary School Behavior Rating Scale Manual. Devon, Pa." Devereaux Foundation Press, 1967a.

Spivack, G., and Swift, M.   Patterns of Disturbed Classroom Behavior-- The Nature and Measurement of Academically Related Problem Behaviors, Devereaux Foundation Report of the Office of Education, Hew., 1967 (ERIC documents, Report Service No. ED12545(b).

Spivack, G., and Swift, M.   The Assessment of Achievement--Related Classroom Behavior. Journal of Special Education, 1968, Vol. 2, pp. 137-154.

Spivack, G., and Swift, M.   The Classroom Behavior of Children: A Critical Review of Teacher-Administered Rating Scales. Journal of Special Education, 1973, Vol. 7, No. 1.

Stennett, R.J. Emotional Handicaps in the Elementary School Years: Phase or Disease? American Journal of Orthopsychiatry, Vol. 36, pp. 444-449, 1966.

Stott, D.H.   The Social Adjustment of Children, Manual to the Bristol Social Adjustment Guide (2nd ed.). London: University of London Press, 1963.

Ullman, C.A.   Identification of Maladjusted School Children.   Public Health Monograph #7. Washington, D.C.: U.S. Government Printing Office, 1952.

Vane, J.R.   The Vane Kindergarten Test. Journal of Clinical Psychology, Monograph #24, 1968.

Van Vleet, P. and others.   Young Children at Risk: A Three Year Research and Demonstration Project Concerned with Early Inter- vention and Prevention of Learning and Behavior Problems. American Journal of Orthopsychiatry, Vol. 40, pp. 320-322, 1970.

Yellott, A.W., Liern, G.R., and Cowan, E.L. Relationships among Measures of Adjustment, Sociometric Status and Achievement in 3rd Graders. Psychology in the Schools, Vols. pp. 315-321, 1969.

Walker, D.K.   Socioemotional Measures for Preschool and Kindergarten Children: A Handbook. San Francisco, Jossey-Boss, 1973.

Wickman, F.K. Children Behavior and Teachers' Attitudes. Common- wealth Fund: New York: 1928.