

AN INVESTIGATION OF CHILDREN'S AROUSAL LEVELS
AS THEY READ GRADED MATERIALS

by

HARRIET WILLIS BRYANT
B.Ed. (Elem.), University of British Columbia, 1971

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE DEPARTMENT OF READING EDUCATION

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August, 1976

© Harriet Willis Bryant, 1976

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study.

I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of BOG EDUCATION

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date OCT 1/76

ABSTRACT

The major purpose of this study was to attempt to determine word recognition and comprehension criteria for the frustrational level of reading by means of the galvanic skin response (GSR), in order to substantiate the conventional criteria used by authorities in the field.

The sample consisted of 60 children in grades two and three, for whom parental consent for inclusion in the study had been obtained. Scores from a standardized reading achievement test were used to stratify all second and third grade children on reading achievement level (below, average, and above average). Ten children for each grade and reading achievement level were then randomly selected (N=60).

Subjects were required to read the words on the Wide Range Achievement Test-Level 1 (1965) while being monitored on the GSR in order to obtain an index of each child's high arousal level. This level was then to have been used to indicate the frustrational level of reading as the child read the passages of the Diagnostic Reading Scales (1972), that was treated as an informal reading inventory. At this level, word recognition and comprehension accuracy scores were to have been obtained, and these compared to the existing criteria. However, during the reading of the Diagnostic Reading Scales passages no subject attained the arousal level obtained on the Wide Range Achievement Test, and therefore frustrational level of reading had to be determined from the criteria stated in Johnson and Kress (1965). As a result, alternate hypotheses were postulated.

GSR arousal levels were obtained for the instructional and pre-

frustrational levels of reading, for both the oral reading and comprehension sections of the test, and these statistically compared. At the frustrational level of reading, comparisons were made between the oral reading and comprehension sections of the test, the three reading achievement groups, and for the boys and the girls.

Data were analyzed using a two-way analysis of variance for a repeated measures design.

It was found that there was no significant difference between the oral reading of the passage and the comprehension section at both the instructional and frustrational levels of reading. The three reading achievement levels did not differ significantly in arousal at the frustrational level of reading.

Implications of the results were discussed.

TABLE OF CONTENTS

Chapter

| | |
|--|----|
| I. STATEMENT OF THE PROBLEM | 1 |
| Introduction | 1 |
| The Purpose | 3 |
| Limitations of the Study | 4 |
| Operational Definitions | 5 |
| Procedures of the Study | 9 |
| Selection of the Subjects | 9 |
| Testing | 9 |
| Treatment of the Data | 10 |
| II. REVIEW OF THE LITERATURE | 14 |
| Introduction | 14 |
| Description of the IRI | 15 |
| Historical Development of the IRI | 16 |
| Controversies Related to the Criteria for the IRI | 22 |
| Reliability and Validity of the IRI | 25 |
| Research Related to the Scoring of the IRI | 27 |
| Physiological Support for the Reading Levels of the IRI | 29 |
| THE GALVANIC SKIN RESPONSE | 32 |
| Introduction | 32 |
| Terminology | 33 |
| Early History | 33 |
| Mechanics | 33 |
| Research Issues | 35 |
| Electrodermal Research in Education | 36 |
| The IRI and Anxiety | 40 |
| III. METHOD | 43 |
| Selection of the Subjects | 43 |
| Materials | 44 |
| Apparatus | 44 |
| Measuring Equipment: GSR Recording | 45 |
| Procedure | 45 |
| Treatment of the Data | 46 |
| IV. RESULTS OF THE STUDY | 48 |
| Introduction | 48 |

CHAPTER

| | |
|---|----|
| V. SUMMARIES, CONCLUSIONS AND RECOMMENDATIONS | 54 |
| Summary of Methods | 54 |
| Summary of Findings and Conclusions | 57 |
| Implications | 58 |
| Recommendations for Further Study | 61 |
| BIBLIOGRAPHY | 62 |
| APPENDIX. | 68 |

LIST OF TABLES

TABLE

| | |
|--|----|
| I. Criteria for Determining the Levels of an IRI | 7 |
| II. Classification of Reading Achievement Groups According to Grade Level | 8 |
| III. Selection of the Subjects | 9 |
| IV. Comparison of Criteria for the Instructional Level of the IRI | 26 |
| V. Analysis of Variance: Comparison of Arousal Levels at the Instructional Level of Reading | 50 |
| VI. Analysis of Variance: Comparisons of Arousal Levels at the Frustrational Level of Reading | 51 |

ACKNOWLEDGEMENTS

An expression of sincere appreciation is extended to those who made possible the completion of this thesis.

In particular, Dr. K. Slade, without whose constant encouragement, advice and generous giving of time, this thesis could not have been completed.

To the other committee members, Dr. R. Bennet, Dr. C. Pennock, and Dr. T. Westermarck, whose time and advice were greatly appreciated.

To the many people who helped throughout the running of this study. Special thanks to graduate students Campbell Clarke, for sharing his time and knowledge, and Jack Tarasoff, who was always there whenever needed. To the boys and girls and staff members of the schools who took part in the investigation for their cooperation in the compilation of the data.

And to my family, for their understanding, encouragement, and unending support, the writer is most grateful.

CHAPTER I

STATEMENT OF THE PROBLEM

This study sought to investigate the galvanic skin response arousal levels of children in grades two and three reading at the mid-Grade 1 and above levels, as they read through the graded passages of the Diagnostic Reading Scales (Spache, 1972) that was treated as an informal reading inventory. The major purpose of this study was to determine word recognition and comprehension criteria for the frustrational level of reading by means of the galvanic skin response in order to substantiate the conventional criteria used by authorities in the field of reading.

Introduction

Literacy is of prime importance in North American culture, and is reflected in our schools emphasis on reading education. Educators agree that maximum learning is facilitated when a child receives instruction at his own level of competence. This concept is reflected in the theory of developmental reading which stresses that each child receives instruction in reading skills at his own level of performance. Teachers of reading are therefore faced with the problem of determining each child's best level for instruction.

To determine a child's best level for instruction in reading, Gray (1920) promoted the idea of informal reading achievement tests, as opposed to standardized reading achievement tests. Killgallon (1942) and Betts

were the first to apply performance criteria to testing with the informal reading inventory (IRI). Today IRI's are widely used, but are fraught with many problems in administrative procedures, scoring methods, and criteria for determining the three performance levels; independent, where the child can read effectively without assistance; instructional, where the child is able to progress with systematic instruction; and frustrational, where the child becomes unable to read the materials (Utsey, Wallen, and Beldin, 1965).

The major thrust of research with the IRI has been concerned with the criteria for determining a child's instructional level in reading. The initial performance criteria for accuracy in word recognition and comprehension determined by Killgallon (1942) has generally been adhered to by later authors, although these criteria lack a strong empirical base. Recent research by Cooper (1952), Powell (1968), and Dunkeld (1970) suggests the need for different criteria for establishing the reading levels of primary children as opposed to intermediate children. They also suggest less stringent performance criteria for these levels than was originally suggested by Killgallon. Powell (1968) has offered empirical data to support his criteria of 85 percent accuracy in word recognition and 70 percent accuracy in comprehension for the instructional level in primary grades, and 91 percent to 94 percent accuracy in word recognition and 70 percent accuracy in comprehension for the intermediate grades. Powell states that most reported criteria for an IRI are based on the intermediate grades, and that these are not appropriate for primary grade children.

The frustration level in reading, that level where the child is unable to handle the reading material, has largely been ignored in research. The criteria for this level, like those for the instructional

level, have seldom been empirically tested or validated, but rather, accepted as accurately defined by Betts and Killgallon.

Ekwall and English (1971) attempted to validate the frustrational reading level criteria of ninety percent accuracy in word recognition and fifty percent accuracy in comprehension using the galvanic skin response as a measure of frustration or arousal. Their major finding was that for children in grades three to five the ninety percent word recognition criterion was valid if each repetition of a word was counted as an error, and they recommended that, until further research could more specifically determine a comprehension criterion, "the 50 percent correct criteria is adequate for comprehension" (p. 42).

The youngest children sampled in Ekwall and English's study were in grade three. It is suggested that had they used younger children they might have found similar trends to the studies investigating the instructional level criteria, namely, that lower word recognition and comprehension scores were more appropriate for children at the grades one and two levels. The results of the galvanic skin response for children of this age level may have indicated higher arousal levels before reaching the 90 percent accuracy criterion for word recognition and 50 percent accuracy criterion for comprehension.

The Purpose

The major purpose of this study, which was descriptive in nature, was:

1. To determine the word recognition and comprehension criteria for the frustrational reading level as determined by the galvanic skin response (GSR).

Secondary purposes of this study were:

2. To compare the criteria for the frustrational reading level as established by Betts and Killgallon (1942) to the frustrational reading level as determined by the GSR.
3. To compare both word recognition and comprehension arousal levels, as measured by the GSR at the instructional, pre-frustrational, and frustrational reading levels of an informal reading inventory.
4. To compare the GSR measured frustrational reading levels and the percentages of word recognition and comprehension accuracy at three levels of reading achievement: below, average, and above average.
5. To compare the mean GSR arousal level of boys and girls at the frustrational level of reading.

Limitations of the Study

This study was limited by a number of factors. Firstly, the sample consisted of 60 children in Grades 2 and 3, from two elementary schools, for whom parental consent for inclusion in this study was obtained. Secondly, this study dealt with only the criteria for the frustrational reading level. Thirdly, the reading instrument used, the Diagnostic Reading Scales (Spache, 1963, 1972), is not an informal reading inventory, but a reliable and highly respected standardized test, which was scored as an IRI. Lastly, only oral reading of the passages was used in this study, since children at this age were felt to have not internalized the reading process to a point where silent reading is a practised and facile skill.

Operational Definitions

For the purpose of this study, the following terms were defined as follows:

1. Informal Reading Inventory: (IRI): An IRI is an individual test consisting of a series of graded passages usually prepared from each level of a basal reader which is being used as the main vehicle for reading instruction in the school, and to which the child has not had prior exposure in his reading program. The child reads orally and/or silently from these passages and a check of comprehension follows. Based on specific criteria for word recognition and comprehension accuracy, three critical levels are determined; independent, instructional, and frustrational.
2. Independent Level: At this level the child can read the material without help. His reading is free from observable signs of difficulty. It is at this level that the child reads for recreation (Johnson and Kress, 1965).
3. Instructional Level: At this level the reading materials are challenging for the child, but not so difficult that the child cannot read it without teacher supervision. Signs of difficulty such as head movements, finger pointing, and vocalizations are not present (Johnson and Kress, 1965).
4. Pre-frustrational Level: The term pre-frustrational level is used to designate the reading passage which is between the instructional level, and directly preceding the frustrational level.
5. Frustrational Level: At this level the child becomes completely unable to cope with the reading materials. The child shows signs of tension and discomfort such as finger pointing and vocalizations (Ekwall

and English, 1971).

6. Word Recognition Errors: The following types of errors were classified as word recognition errors:

- a. Substitutions: Any whole word stated which differs from the given word in the text.
- b. Insertions: Any word or groups of words inserted which did not appear at that point in the printed material from which the child read (McRae, 1974). Example: "The kitten played," is read, "The little kitten played."
- c. Omissions: Whole words or groups of words left out are counted as omissions errors (Spache, 1963). Example: "The little kitten played," is read, "The kitten played."
- d. Words Aided: Words which, when encountered by the child, he refuses to read, or is unable to successfully decode, and which are told to him after a five second interval (Spache, 1963).
- e. Repetitions: Reading authorities concerned with the scoring of an IRI disagree concerning the number of words which constitute a repetition error, and as to whether repetition errors should be included as scorable errors on an IRI. Johnson and Kress (1965) feel that the repetition of each word should be counted as an error. Spache (1963), on the other hand, considered only repetitions of two words or more as scorable repetition errors.

For the purpose of this study, the following definition of a repetition was adopted, that is, a repetition was scored when the child repeated one or more words. Each word(s) repeated a different number of times than other words was considered an additional error (Ekwall and English, 1971). For example, in sentence one below there are three

errors, but in sentence two there is only one error.

1. The little kitten played.

2. The little kitten played.

7. Comprehension: Reading comprehension is determined by asking the child a number of questions based on the subject matter in the passage he has just read. The score is expressed as a percentage of correct answers.

8. Criteria for the IRI: The criteria for determining the three reading levels of an IRI are expressed in percentages for both word recognition accuracy and comprehension accuracy for each level. Table I shows the various criteria for determining the reading levels of an IRI.

TABLE I

CRITERIA FOR DETERMINING THE LEVELS OF AN IRI

| Reading Level | Word Recognition | Comprehension |
|---------------|------------------|---------------|
| Independent | 99% | 90% |
| Instructional | 95% | 75% |
| Frustrational | 90% | 50% |

(Johnson and Kress, 1965)

9. Classification of Reading Achievement Levels: Table II specifies the reading achievement levels for above average readers, average readers, and below average readers used in this study. The criteria are after Bond and Tinker (1967).

TABLE II
CLASSIFICATION OF READING ACHIEVEMENT GROUPS ACCORDING
TO GRADE LEVEL

| | Below Average | Average | Above Average |
|---------|---------------|------------|---------------|
| Grade 2 | below 2.3 | 2.4 to 2.7 | above 2.8 |
| Grade 3 | below 3.2 | 3.3 to 3.8 | above 3.9 |

10. Galvanic Skin Response Terminology: The following definitions are based on Chaplin (1968) and Sternbach (1966).

a. Galvanic Skin Response: (GSR): Galvanic skin response refers to changes in the electrical resistance of the skin and is detected by a galvanometer. A response may be measured by passing a weak electrical current through the skin and the resistance to this current measured. It is best thought of as an index of activation and arousal (Sternbach, 1966).

b. Arousal: This term refers to the level of autonomic alertness following sensory stimulation, which may lead to a heightened state of tension, anxiety, and activity level. This is the level when the subject is responding to a specific stimulus, resulting in a skin resistance decrease (Chaplin, 1968).

c. Baseline Level: This is the level designated when the subject is at rest or in a relaxed state, and is considered a global function of mental, emotional, and environmental set (Sternbach, 1966).

d. Criterion Arousal Level: CAL: The criterion arousal level was determined by having each subject read through the Wide Range Achievement Test-Level 1 (1965) until the GSR had plateaued and habituation began to occur. It was measured in centimeters from the baseline level to the maximum level of deflection.

Procedures of the Study

Selection of the Subjects: The Stanford Achievement Test (1972) was administered to intact classrooms of second and third grade children in the schools allotted by the Vancouver School Board. Level II of the test was administered to the second grade population and Level III to the third grade population. From the results of this test the children were stratified on achievement levels according to Table II (p. 8). Parental consent forms (see Appendix) were sent home with each child. From the select population of children whose families consented to the study involving their children, thirty children were selected from each grade, ten children in each of the reading achievement levels using a table of random numbers (Glass and Stanley, 1970), as is shown in Table III.

TABLE III
SELECTION OF THE SUBJECTS

| | Reading Achievement Levels | | | |
|---------|----------------------------|---------|------------------|-------|
| | Below Average | Average | Above Average | Total |
| Grade 2 | 10 | 10 | 10 | 30 |
| Grade 3 | 10 | 10 | 10 | 30 |
| Total | 20 | 20 | 20 | 60 |

Procedures were tested on seven children in order to develop testing techniques and scoring procedures. A polygraph expert examined the galvanic skin response profiles for sensitivity and accuracy of recording. These records showed a good response pattern.

Testing: Each subject was escorted to the testing room where the

electrodes were attached to the index and ring fingers of his non-preferred hand. After an attempt to establish rapport was undertaken, the subject was given the Wide Range Achievement Test-Level 1 (1965) (WRAT) and asked to read the words. The child read until a plateau had been reached on the galvanic skin response (GSR) and habituation had begun. Each subject's reaction was measured on the GSR print-out as the maximum deflection in centimeters from the baseline level, and this used as the subject's criterion arousal level (CAL).

In addition, each subject was asked to blow up a balloon 'until it bursts'. This testing situation was terminated for each child before the balloon burst. The measure obtained served as an index of the maximum arousal level in a non-verbal situation. In all cases this measure provided a greater arousal level than did the Wide Range Achievement Test. This enabled the experimenter to compare how close the criterion arousal level was to the higher non-verbal arousal level.

After the subject had again relaxed, he was given the first passage of the Diagnostic Reading Scales (1972) to read orally. Upon completion, the booklet was removed and the comprehension questions asked. The next highest level was then administered and the same procedure followed until the GSR indicated that the subject's CAL had been reached. However, in the data gathering situation none of the subjects attained their criterion arousal level during the reading of the passages, regardless of passage difficulty, as had children during the testing of the procedures. Because of this, frustrational level in reading had to be determined by the criteria set forth in Johnson and Kress (1965).

Treatment of the Data: The passage where the GSR indicated the subject's CAL had been reached was to have been scored for word recognition

and comprehension accuracy. An average word recognition and comprehension score was to have been computed using the total sample and these scores statistically compared to the criteria as set forth in Johnson and Kress (1965). This treatment of the data could not be completed since the frustrational level of reading had to be determined using the criteria stated in Johnson and Kress rather than from the criterion arousal level.

In order to analyze the data, an index of arousal for the IRI passages at the frustrational, pre-frustrational, and instructional levels was determined for both the oral reading and the comprehension sections, and a mean arousal level computed for each subject at the frustrational level of reading. This was obtained by measuring the maximum point of deflection from the baseline level in centimeters. An average level of arousal was computed for the total sample and used in the comparisons of arousal levels between the instructional and pre-frustrational levels.

Statistical analyses were conducted using a two-way analysis of variance for a repeated measures design using ANOV:23 (1969). The following null hypotheses were originally postulated.

Hypothesis 1

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level and the level on the GSR where the 90 percent word recognition criterion on the IRI was applied.

Hypothesis 2

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level and the level on the GSR where the 50 percent comprehension criterion on the IRI was applied.

Hypothesis 3

At the instructional level of reading as determined by the administration of an IRI there will be no significant difference between the

GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

Hypothesis 4

At the pre-frustrational level of reading there will be no significant difference between the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

Hypothesis 5

At the frustrational level of reading as determined by the subject's criterion arousal level (CAL) there will be no significant difference between the galvanic skin (GSR) measured arousal level found during the oral reading and that found during the comprehension section of the test.

Hypothesis 6

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading at grade level.

Hypothesis 7

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading below average.

Hypothesis 8

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading at grade level compared to children reading below average.

Hypothesis 9

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the boys and the girls.

Because of the failure of the sample to obtain uniformly a criterion arousal level at the frustrational level of reading, it became necessary to determine the frustrational level of reading by the criteria set forth in Johnson and Kress (1965) and the following alternate hypotheses

were generated on a post hoc basis.

Hypothesis 1

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

Hypothesis 2

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading at grade level.

Hypothesis 3

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading below average.

Hypothesis 4

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading at grade level compared to children reading below average.

Hypothesis 5

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of the boys and the girls.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

In recent years, educators and psychologists have given increasing attention to the emotional aspects of reading. Gates (1947) stated that there was substantial evidence to support the conclusion that success in reading depended to a considerable extent upon the kind of mental and emotional adjustment that the pupil made to the learning situation. In order that the learning situation be the least stressful and most beneficial to the pupil, it is essential that instruction, particularly reading instruction, be based "upon the student's ability to comprehend it" (Fay, 1969, cited in McRae, 1974, p. 4). For this to happen instructional materials must be at the pupil's instructional level. Harris emphatically stated, "The most important question to answer about a child's reading is, 'How difficult a book can this child read?'" (1970, p. 139).

One method available to the classroom teacher for determining the level of printed materials to use for the instruction of a pupil is the informal reading inventory (IRI). The IRI is generally agreed to be a valuable device in determining the various functional reading levels of a pupil, as well as for diagnosis of reading difficulties.

It is the purpose of this review to define, trace the development of the IRI, and report research related to the IRI. The galvanic skin response is also described.

Description of the IRI

The informal reading inventory is perhaps best described as a structured reading interview in which a child reads, both orally and silently, selections of increasing difficulty. As the child reads, the examiner notes the child's pronunciation errors, comprehension errors, and other behavior related to his reading. After the informal reading inventory is completed, the teacher compares the child's reading behavior with a set of criteria which enables her to identify the child's functional reading levels--independent, instructional, and frustration (Utsey, Wallen, and Beldin, 1965, p. 1).

This succinctly stated description includes the basic concepts of an IRI. The informal inventory is individually administered to a child, and usually consists of graded passages that have been chosen from the textbooks used in the classroom. The examiner listens to the child's oral reading of a passage, and records his word recognition errors. Silent reading of the passage may also be included. Comprehension is appraised by asking the child a number of questions concerning the content of the passage he has just read. The child's reading performance is not "compared to what other students can do . . . given the same job", but is evaluated "in terms of absolute standards" (Johnson and Kress, 1972, pp. 185-186).

There are a number of purposes attributed to the IRI, as identified by Johnson and Kress. Firstly, they see the IRI as a means of appraising achievement levels in reading, in order that the teacher be more able to adjust reading material to the instructional level of the child. Secondly, the IRI can be used diagnostically. By knowing the child's areas of weaknesses, the teacher can teach the specific skills needed by the child in order for him to become a more proficient reader. Lastly, Johnson and Kress suggest that the IRI could be used repeatedly in the evaluation of the child's progress, so the continued growth can be

measured (1965).

Standardized reading tests are another way of determining reading level and diagnosing reading problems. These tests are often lengthy to administer, and often fail to diagnose reading inability accurately. A major criticism is their inability to accurately place pupils at their proper instructional level. While these tests measure reading grade level, and relate information about the pupil in terms of grade achievement, they tend to overestimate the pupil's actual instructional reading level by one-and-a-half to four years (Farr and Anastasiow, 1969). Many authorities have demonstrated that standardized tests are actually indicating reading frustration level rather than the instructional level (Betts, 1957; Millsap, 1962; Sipay, 1964; Botel, 1957; Dunkeld, 1970).

This is not to denigrate the use of standardized reading tests. Standardized tests are often used in a comparative way to discriminate the poorer readers from the more able readers. The IRI, on the other hand, is used to determine the level of material that the child can read with adequate word recognition and comprehension skills. Both of these instruments have their place in a reading program.

In order to meet the needs of pupils in reading, and avoid some of the problems associated with the use of standardized reading tests, it has been recommended that an IRI be used to supplement the use of a standardized reading test.

Historical Development of the IRI

In the early part of the century, the standardized reading test was the most popular means for determining a child's reading level. In 1920 Gray advocated the use of informal tests. In doing so, Gray was not

intending that informal tests be used as a substitute for standardized tests. He felt that ". . . informal tests must be used frequently along with standardized tests, if testing is to be continuous and most effective" (Powell and Dunkeld, 1971, p. 638).

Gates and Betts both published books in 1936 which made reference to the use of informal testing. In line with the contemporary view of the importance of oral reading performance, Betts wrote:

Oral reading tests, either formal or informal from standard readers, provide needed evidence. An experienced examiner cannot use of context clues and deficiencies in word analysis as well as tendencies to reverse forms, to repeat, to omit, and to substitute. Not infrequently, children can arrive at the meaning of a silent reading selection and still evidence faulty oral reading (1936, p. 98).

It was not until 1937 that an effort was made to differentiate reading with ease from reading with difficulty (Durrell, 1937). In this article, Durrell stated, "In the usual classroom practice, it appears that children find difficulty in mastering material containing more than one difficult word in twenty running words" (1937, p. 333, cited in Beldin, 1970).

Killgallon (1942), under the direction of Betts, conducted the first research study which helped to formulate the criteria for the identification of an instructional level in reading, as well as to set probable criteria for the independent, frustrational and capacity levels. Killgallon administered the Gates Reading Survey, Grades III to X, to 211 fourth grade children in order to study "certain aspects of the general problem of individual differences in reading". In order to check the validity of the data at the instructional reading level, he administered an informal reading inventory to every fifth pupil on the ranked distribution of the test results: a total of 41 pupils. The criteria for the

lower limits of the functional reading levels were arbitrarily established at a ratio of one error in every fourteen running words (92.86%) for word recognition, and 50 percent or better for comprehension. Despite these results, Killgallon shifted the criteria for evaluation to 95 percent word recognition accuracy and 75 percent comprehension accuracy. He justified this shift in criteria by stating:

The mean comprehension score at the instructional level was 71% (p. 162). . . . The mean percentage of word recognition error was 5.1; the limits of the range, 1.2 and 6.6. In corresponding terms, pupils made approximately five word-perception errors in every 100 running words, or one in every twenty, on the average (p. 165).

His final criteria were 95 percent accuracy in word recognition, and 75 percent accuracy for comprehension.

Beldin (1970) has suggested that the development of Killgallon's criteria may have been based on the early work of Bolenious and Durrell. Bolenious (1919), in his book, Teacher's Manual of Silent and Oral Reading, suggested that a child should grasp at least 50 percent of the ideas in a four hundred word passage. Durrell, as noted previously, suggested the criterion of "one difficult word in twenty running words" (1937, p. 333). Beldin hypothesizes that Killgallon arrived at the 95 percent word recognition accuracy criterion by rounding off the figure to facilitate its use by classroom teachers. Similarly Beldin feels that the initial comprehension figure of 50 percent was shifted to 75 percent, a rounding off of the actual mean comprehension score of 71 percent, for easier reference by teachers.

Despite the unorthodox manner by which Killgallon arrived at his criteria for the instructional level, his work has been widely quoted, and is recognized as the first empirical attempt, to establish criteria for an informal test, albeit that these criteria were not his major concern in

his thesis.

Betts, building upon the Killgallon study, went on to set criteria for the Independent Level, Instructional Level, and Frustration Level, as they are presently called (1936). It is generally conceded that Betts has been the main influence on the development of criteria for the IRI. Much of the research on the IRI have been attempts to modify these criteria.

Cooper (1952), in a study designed to, in part, establish criteria for the instructional level, concluded that the primary levels should have 98 percent accuracy in word recognition and at least 70 percent accuracy in comprehension, and at the intermediate levels, 96 percent accuracy in word recognition and 60 percent accuracy in comprehension. He arrived at these criteria by administering an IRI to approximately 1000 children in grades one through six. The children read the passages orally and reread them silently. Based on the word perception errors made, he classified them into five groups according to those who made the most and least errors. Two standardized reading tests were administered at the beginning and at the end of the year. By comparing these results with the IRI's for each group he found that, at the primary level, the groups making the greatest amount of progress in reading achievement, were characterized by 0-1.99 word perception errors per 100 running words at a comprehension of 70 percent. For the intermediate grades, the 0-1.99 and 2-3.99 groups made the greatest gains with 60 percent comprehension scores. He concluded that these groups had been placed in 'suitable' instructional materials, and that the criteria should be, for the primary grades, 98 percent accuracy in word recognition and 70 percent accuracy in comprehension, and for the intermediate grades, 96 percent accuracy in word recognition and 60 percent accuracy in comprehension.

This study, while being one of the first studies of its kind, does suffer from some methodological weaknesses. These include, in part, limited range of materials in the classes for the children to read, inter-rater reliability of IRI scoring, and the limited geographical location used in the study (McKinlay, 1973). Powell suggests that,

the only certain conclusion from his (Cooper's) data would seem to support the conception that children in frustration reading material do not grow in reading skill, as measured by a standardized test, at a rate commensurate with those children who are not in frustrational reading material (1968, p. 7).

Sipay (1964) also attempted to validate the criteria used for scoring an IRI. Sipay used Cooper's criteria for determining the instructional level, and Betts' criteria for determining the frustrational level. The Metropolitan Reading Test, the Gates Reading Survey, and the California Reading Test were administered to 202 fourth graders. Two forms of an IRI were constructed from the Scott Foresman reading series: one being used to determine the subjects' functional reading levels, the other to determine the reliability coefficient using 40 randomly chosen subjects. When Cooper's 'suitable' criteria of 96 percent accuracy in word recognition and 60 percent accuracy in comprehension were used, "all three standardized tests tended to over-estimate the instructional level by approximately one or more grade levels" (p. 267). When Cooper's 'marginal' criteria of 9-percent accuracy in word recognition and 60 percent accuracy in comprehension were used, the Metropolitan Reading Test, and the Gates Reading Survey tended to indicate the instructional level. Sipay stated:

These findings suggest that it is impossible to generalize as to whether standardized reading test scores tend to indicate the instructional or frustrational level. Rather, it appears that in making such judgements, one must consider the standardized reading tests used and the criteria employed to estimate the functional reading levels (p. 268).

He concluded that the use of standardized tests does not appear warranted when used in the selection of the most suitable instructional materials in reading.

The Standard Reading Inventory was constructed by McCracken (1964) as a "valid and reliable individual reading test for measuring the reading achievement of elementary school children". He used the criteria of Killgallon to determine the independent, instructional, and frustrational levels. The vocabulary of three basal readers was analyzed and test materials developed which were felt to have high validity. To check validity further, the Spache Readability Formula was used to determine the difficulty level of the passages at the primary grades, and the Dale-Chall formula used for passages at the intermediate grades. The subjective ratings of twenty-five national experts were used to further support the results. This test was then administered to 664 pupils in grades one through six. Two trained examiners administered the test and the reliability was established. Twelve Pearson product-moment correlations were computed, and ranged from .68 to .99, with a median of .91. McCracken's study resulted in the production of a structured diagnostic test; not an informal reading inventory. While his Standard Reading Inventory is similar to an informal reading inventory, it should be remembered that a tenet of an informal inventory is that it is made from materials that the teacher plans to use in classroom instruction. A further weakness with McCracken's Standard Reading Inventory is the efficacy of using Killgallon's evaluative criteria.

Controversies Related to the Criteria for the IRI

The original word recognition criteria stated by Killgallon, and later by Betts, are viewed by some reading authorities today as being much too stringent for determining reading levels. Powell asserts that only three (Killgallon, Cooper, and Powell) have attempted to offer data to support their criteria (Powell and Dunkeld, 1971). Powell and Dunkeld are not alone in their views, for Lowell (1970) agrees that there is no real evidence to support the criteria. Regardless of this disagreement among experts, the IRI is being widely used with the different word recognition criteria being accepted as norms.

The criteria for comprehension performance is another area of disagreement. Betts originally felt that, for the instructional level, a pupil should be able to recall three-fourths of what he read. There are those, however, who feel that this should be reduced to 70 percent, or 60 percent (Cooper, 1952; Powell, 1968; Dunkeld, 1970).

One reading authority who challenged the Betts-Killgallon criteria for word recognition and comprehension was Powell (1968). He conducted a study to test the hypothesis that "the word-recognition criterion was lower than the 95 percent level and to attempt to pinpoint the probable level" (p. 10). Three examiners collected 178 IRI protocols from average ability and middle-class children in grades one through six. The method Powell used was to hold comprehension constant at the lowest acceptable score for classification at the instructional level: 70 percent or higher. The point of entry into the word recognition column was determined by this comprehension score. Word recognition scores were scanned to determine the lowest percentage of word recognition accuracy within this comprehension range. The reason for choosing the 70 percent comprehension level,

rather than the 75 percent level, "was to mitigate (sic) the effects of questions asked of the subject. . . . Observation revealed that this precaution was not truly necessary, as only a very small number of cases would have been so affected" (p. 11). No clear explanation is given for why he would have considered a 75 percent comprehension criterion, but it can be surmised that this might have been based on the Killgallon criteria.

Powell's rationale for this procedure was, "that if the youngster's comprehension per cent remained continuously at an acceptable level, then the fluctuation in word pronunciation was tolerable to the reader" (pp. 11-12). The data from this study,

clearly indicate that pupils in grades one and two can tolerate on the average an 85 per cent word-recognition score and still maintain seventy per cent comprehension. The finding that pupils in grades three through six could tolerate on the average a 91 to 94 per cent word-recognition score while maintaining 70 per cent comprehension is commensurate with the data of Killgallon and Schummers. This finding was expected (pp. 13-14).

A key issue in this statement is Powell's use of the term 'tolerable'. Although he does not explicitly define the word, presumably he meant that any variation in the word recognition error would not be frustrating to the reader as long as the comprehension criterion of 7- percent was maintained.

Support for Powell's criteria was obtained by Dunkeld (1970) in a study designed to examine the relationship between children's gains in reading and the difficulty level of their instructional materials. A total of 212 children, 101 boys and 111 girls, in grades two through six, who had average intelligence and reading achievement, were administered the Gates-MacGinitie Reading Tests of vocabulary and comprehension as pre- and post-tests of reading achievement. An IRI was constructed and administered by six trained examiners. The IRI passages which best matched the

readability of the instructional materials were noted for each child. Dunkeld's instructional level criteria were obtained by comparing children's gains with the difficulty of the instructional materials. The boundaries for the instructional level that he reported were: second grade word recognition scores, 89.5%-95.0%; third grade word recognition scores, 89.0%-97.0%; and fourth, fifth, and sixth grade comprehension scores, 60%-90%. Dunkeld concurred with Powell that,

the instructional level criteria most widely encountered in professional literature, when applied to children's oral reading at sight, underestimates children's abilities, especially at the lower grade levels, and fails to recognize the need to apply different criteria at different stages of development (p. 6274-A).

Two interesting findings were reported in the Dunkeld study. Firstly, he noted that children's performance on the first passage read was often inferior to their later performance, suggesting that the first passage may not be a valid indicator of reading performance. Secondly, he noted the lessening correlation between word recognition scores and comprehension at the higher grade levels. This latter finding is in agreement with Becker, who found that fourth grade children with the lowest comprehension scores made the fewest oral reading errors, and vice-versa (1970, cited in Pennock, 1973).

Spache (1963) also disagreed with the Betts-Killgallon criteria. In developing his Diagnostic Reading Scales he found that the practice of applying the standard of one error in twenty running words not to be justified. Powell and Dunkeld (1970) analyzed five widely used instruments: the Diagnostic Reading Scales, the Durrell Analysis of Reading Difficulty, the Gilmore Oral Reading Test, the Gray Oral Reading Test, and the Gates-McKillop Reading Diagnostic Tests. For each of the tests a word count was made of each passage and this figure divided by the number of

permissible errors that could be made in order to stay within the limits of acceptable reading behavior according to the norms. This resulted in a word recognition error ratio. The data from this study indicated that "all the word recognition error ratios increase in error latitude as the difficulty of the material increases and the age-grade of the sample increases" (p. 641). They concluded that Powell's (1968) criteria for the IRI are appropriate, since the most suitable criteria for the IRI were those which attempted "to reflect the progression of the increase of language difficulty and the reader's response to this increase" (p. 641).

The research concerning the criteria for the instructional level of the IRI is far from conclusive in support for one criteria over another. While users of the IRI have many criteria to choose from, those suggested by Killgallon (1942) and later by Johnson and Kress (1965) appear to be the most widely used today.

Table IV summarizes some of the various criteria for determining the instructional level of an IRI.

Reliability and Validity of the IRI

Word recognition and comprehension criteria are not the only areas of disagreement concerning the IRI. Validity and reliability are seldom mentioned with respect to the instrument, but are necessary aspects when dealing with any test. Farr (1969) and Botel (1969) believe that, because the IRI is used to measure reading performance over a number of occasions, the reliability will be high. McGinnis (1970) stated that the reactions of the child in the situation are more important than the testing procedures. If one is to have confidence in test results, however, one must rely on standard administrative procedures. Criticism leveled at the IRI

TABLE IV.
COMPARISON OF CRITERIA FOR THE INSTRUCTIONAL
LEVEL OF THE IRI_a

| Author | Date | Word Recognition Percentage | Comprehension Percentage | Comment |
|--------------------|------|-----------------------------------|-----------------------------|--------------|
| Betts & Killgallon | 1942 | 95 | 75 | |
| Cooper | 1952 | 98 | 70 | Primary |
| | | 96 | 60 | Intermediate |
| Wheeler & Smith | 1957 | 95 | 70 | |
| N. B. Smith | 1959 | 80 | 70 | Primary |
| Austin & Heubner | 1961 | 95 | 75 | |
| Harris | 1961 | 95 | 75 | |
| Spache (b) | 1963 | normed by grade | 60 | Primary |
| Johnson & Kress | 1965 | 95-99 | 75-90 | |
| Powell | 1968 | 85 | 70 | Primary |
| | | 91-94 | 70 | Intermediate |

a. after Powell, 1968

b. Spache, 1963

has often touched on the inability of the examiner's themselves to accurately score IRI's. Lowell (1970) feels that teachers, who are the main users of the inventory, need extensive training before using it. Farr (1969) concurs with Lowell, that one of the factors limiting the use of the IRI is " . . . the ability to record errors" (1970, p. 2). Kender stated that, anyone who administered an IRI must be able to make decisions about a child's "word analysis skills, his oral reading, his silent reading, his comprehension, and many other factors involved in the reading process. . . . The usefulness, then, of an informal reading test is in direct proportion to the knowledge of the examiner who uses it" (1968, p. 341). It is for these reasons that several authorities maintain that the IRI should be used together with a standardized reading test.

Empirical studies to obtain validity and reliability coefficients

for an IRI have been conducted. McCracken and Mullen (1970) sought to determine if the data from two informal reading inventories, as well as one standardized reading achievement test, would be valid indicators of reading levels. The Standard Reading Inventory, Form A, the Botel Reading Inventory, the Stanford Achievement Tests, and the California Tests of Mental Maturity were administered to 147 children in grades one through six. The manual for both the Standard Reading Inventory and the Botel Inventory report the findings of this study. Concurrent validity coefficients for these tests were reported as being .80 and .88 (Botel, Bradley, and Kashuba, 1970).

The reliability of an informal reading inventory was obtained by McRae (1974). The Pupil Placement Test, a type of informal reading inventory, was administered to thirty-six sixth grade girls. The raw scores of the sub-tests were divided into halves and a split-half correlation computed. McRae concluded that this test was statistically reliable for his restricted sample. Correlations of .72 and .73 were obtained for the IRI and the Metropolitan Achievement Test, as well as the common finding that the standardized test's instructional level was approximately three years above the IRI's.

Research Related to the Scoring of the IRI

Scoring the word recognition section of an IRI also presents many problems. Experts do not agree on what word recognition errors to include when scoring the word recognition section of an IRI, nor on how to categorize scorable word recognition errors. Most reading tests include the following as scorable word recognition errors: mispronunciations, substitutions, omissions, insertions, words aided, and repetitions.

Some tests include the following as scorable word recognition errors: hesitations (Durrell, 1955), reversals (Spache, 1963), and self-corrections (Powell, 1968).

The word recognition error of repetitions is one category in which there is particular disagreement as to whether or not to include it as a scorable error, and if so, exactly how many words constitute a repetition error. Johnson and Kress (1965) and Ekwall (1976) state that each word repeated is a scorable repetition error. Spache (1963), on the other hand, states that repetitions are a normal part of the reading act, and should only be counted when two or more words are repeated. There is a definite lack of validation for any of the scoring criteria, and this is a limiting factor in determining the various reading levels of an IRI.

The question as to whether children should read both orally and silently in an IRI also lacks a consensus view by experts. Betts and Killgallon (1942) preceded oral reading by silent reading of the passages, Cooper (1952) used silent reading after oral reading. Johnson and Kress (1965) and Powell (1968) used oral reading only. Busboom (1974) attempted to offer empirical data concerning the relationship between the testing technique, either oral or silent reading, and the instructional and frustrational levels. The Pupil Placement Test was individually administered by one of six examiners to 204 second through fifth grade children. The children were randomly assigned to four treatment groups: (1) reading orally at sight followed by a comprehension test and an oral rereading, (2) silent reading of a passage followed by a comprehension test, then oral rereading, (3) silent reading preceding the oral reading of a passage, then the comprehension test, and (4) two successive oral readings of the same passage followed by the comprehension test. Powell's

criteria were used for determining the instructional and frustrational levels. Results indicated no differences at the instructional level for both word recognition and comprehension between the treatment groups. At the frustration level significant differences were found, the oral reading then oral rereading and comprehension groups having the highest word recognition scores. Busboom also found a significant grade level difference for the mean number of omissions and aided word miscues at both the instructional and frustrational levels, which is in agreement with criteria suggested by such authorities as Spache (1963) and Powell (1968).

Physiological Support for the Reading Levels of the IRI

A final limitation to be noted concerning the IRI is the lack of supporting evidence for the levels: independent, instructional, and frustrational. In a unique approach to the problem, Ekwall and English (1971) used the galvanic skin response to validate the criteria for scoring informal inventories, and also to determine whether any one set of criteria is applicable to pupils of various intelligence levels, ages, sexes, ethnic backgrounds, and reading levels. They were particularly interested in whether or not to count repetitions as errors. Sixty-two children in grades three through five were chosen; one-third were Anglo-American, one-third were Mexican-American, and one-third were Negro-American. The oral reading passages of the Diagnostic Reading Scales were read and recorded on tape while the child was being monitored by the polygraph in order to obtain data on the frustration or anxiety level of the child. Each child read until the polygraph indicated frustration, and on this passage the percentage of oral errors and comprehension errors were computed. Each inventory was scored twice, once counting repetitions as errors, and once

not counting repetitions as errors. The Wechsler Intelligence Scale for Children, the House-Tree-Person Test, and the Rorschach Test were also individually administered to each child.

No significant differences were found in polygraph-measured frustration reading at the three grade levels, for either sex, or ethnic group. For the total sample significant differences were found between polygraph-measured frustration and the ten percent oral reading errors (or 90 percent accuracy) when repetitions were not counted as errors, however, when repetitions were counted, no significant differences were found. For the comprehension criterion, a significant difference was found between polygraph-measured frustration and the 50 percent criterion. Polygraph-measured frustration for the total groups was reached when 58.39 percent errors had been made, suggesting to the authors that the 50 percent criterion was 'adequate' until further research could be conducted to verify if this criterion is truly the instructional level or the frustrational level (Ekwall, Solis, and Solis, 1973).

Ekwall and English (1971) analyzed the results according to intelligence level, reading achievement level, and personality type. When the sample was analyzed according to intelligence level, the only significant difference found was for the groups one standard deviation below the mean (85 IQ and below), who took an average of 22.40 percent word recognition errors without repetitions, and 23.00 percent errors with repetitions to reach frustrational reading level. This group differed significantly from children with normal intelligence and those one standard deviation above the mean (115 IQ and above). Since analyses were done with small numbers of children, the results must be regarded tenuously. Based on these tentative findings, however, it can be suggested that children who

have a low IQ can make more word recognition errors when reading before becoming frustrated. No significant differences were found for the comprehension criterion.

Ekwall and English (1971) found that when the results were analyzed according to reading achievement level, for word recognition errors, the group reading one year above grade level differed significantly from both the average and below average (one year or more below grade level).

When comprehension was considered, the above average readers frustrated significantly sooner than the below average, but no significant difference was found between above grade readers and average readers.

Rugel (1971) used the galvanic skin response (GSR) as a measure of arousal to determine if physiological arousal increases as the reader goes from the independent to the instructional to the frustrational levels, and also to evaluate the usefulness of the GSR as an indicator of anxiety in children with reading problems. Twelve girls and eight boys in grades two and three, who were average readers, were individually monitored on GSR as they read a preliminary series of graded materials. After this initial evaluation, passages at the child's independent, instructional, and frustrational levels were determined according to the Betts-Killgallon criteria, and the child was asked to read these passages. The order of presentation was counterbalanced so that each level appeared an equal number of times in the first, middle, and last position. Arousal was defined in terms of the number of GSR fluctuations per minute decreasing 1250 ohms or more. Results indicated that the level of arousal increased significantly from the independent level, but nonsignificantly from the instructional level to the frustrational level.

The results of Rugel's study, while supporting the hypothesis that

a child's level of arousal increased as reading difficulty increased, also showed that children were obviously not reading at a true frustrational level. Perhaps for children of this age, the frustrational reading level is not accurately designated by the Betts-Killgallon criteria. Just as Powell (1968) originally found that, for the instructional level, more word recognition errors could be tolerated by this age group, so it may be for the frustrational reading level. Rugel concluded that the GSR was sensitive to changes in reading difficulty and could "probably be a useful diagnostic tool with problem readers" (1971, p. 460).

The purpose of this section has been to trace the development of the IRI and to report the research related to this type of instrumentation. Its intent has not been to denounce the use of other alternatives, for a full reading program requires the use of many instruments, the IRI being one such means of measurement in reading. Different researchers have made attempts at establishing the most efficable criteria for scoring the instructional level on this instrument. Nine different researchers make claim to seven different percentages of word recognition and comprehension (Table IV). Much of the research has dealt with the instructional level of the IRI; the frustrational level has been largely ignored by researchers. This research project attempts to contribute to this paucity of information.

THE GALVANIC SKIN RESPONSE

Introduction

Considerable research has been published concerning the galvanic skin response (GSR). It is not the purpose of this review to recapitulate this research, but rather to give a brief description of this physiological measure.

Terminology

There are in use a wide variety of terms to designate the electrical activity of the skin. Two common types of techniques for measuring electrodermal activity are exosomatic, where a current is passed through the skin from an external source and resistance is measured, and endosomatic, where no current is used and the electrical activity of the skin itself is measured (Venables and Christie, 1973). This review is concerned with the first technique.

The term GSR is the most widely employed term, and commonly refers to either skin resistance or skin conductance, the latter being the reciprocal of the former ($C=1/R$).

Early History

The field of electrodermal research has been extensively reviewed (Landis and DeWicke, 1929; Landis, 1932; Neumann and Blanton, 1970).

The study of arousal and physiological phenomenon is generally conceded to have begun in 1888 with the work of Féré. He proposed a theory that 'psychic energy' was affected by sensations, and by the intensity of those sensations (Neumann and Blanton, 1970). Féré experimented with sensory and emotional stimuli and their effect on skin resistance responses, noting they were accompanied by a decrease in skin resistance. Since this early beginning, a large body of literature has developed relating electrodermal responses and arousal theory.

Mechanics

Numerous theoretical approaches to arousal, based on the various components of the response have been hypothesized. All of these approaches deal with arousal as a unitary variable which should be manifested by

generalized increases in both the central and autonomic nervous systems. When an individual is aroused, sweating occurs in the 'emotional sweating areas': the palms of the hands. The amount of sweating of the palms reflects the degree of activation of the palmar sympathetic nervous fibres. Sweat, being a salty solution, conducts electricity well, and when an electrical current is passed through the skin, a certain amount of resistance is met. The higher the concentrate of sweat, or salty solution, the lower the resistance that is met (Sternbach, 1966). This decrease may be either short-term and referred to as a response, or it may be long-term, and referred to as baseline change. The former is stimulus specific; the latter is a global function of mental, emotional and environmental set, or called generalized arousal (Grossman, 1967).

Ohm's law is used to express the relationship among the electrical factors. This law can be stated in three ways: $V(\text{voltage}) = I(\text{current}) \times R(\text{resistance})$, or $R = V/I$, or $I = V/R$. The resistance of an individual is variable and is dependent on his arousal level, the individual can be considered as the variable resistor. When a small current is passed through the individual, changes in skin resistance can be measured in terms of voltage change, or conversely, the voltage may be held constant and changes in current can be measured, representing a response to a stimulus. The less sweating, the higher the resistance to the current; the more sweating, the lower the resistance to the current (Sternbach, 1966). Results can be stated in ohms, which are the units of measurement for resistance, or in mhos, which are the units of measurement for conductance.

Research Issues

The use of baseline change or tonic levels as an index of generalized arousal is strongly endorsed (Duffy, 1962; Raskin et al., 1969). Unfortunately, tonic levels vary, depending on a number of variables, including the individual, sex, environment, and electrode sites. Electrode placement can be easily controlled in research. Environment, another important variable, can be manipulated in order to reduce the error variance. Ideally the experimental situation should be between 20° to 30° Celsius, with random noises reduced to a minimum, and good lighting provided when tasks are required of the subjects (Scholander, 1963).

Some aspects of the individual which have significance in electrodermal research are age and sex of the subject. The work of Schmavian et al. (1965, 1968) suggests that, although male subjects show higher electrodermal responsivity, and lower skin response level, these sex differences are not observed in pre-puberty subjects. Physiological and metabolic differences within the individual are variables which are relatively uncontrollable. When comparing individuals using the basal level as a measure of generalized arousal, one must be particularly careful.

When considering the amplitude of response, it is necessary to refer to the 'law of initial value'. This law, simply stated, says that the size of a response is related to the level from which it started (Venables and Christie, 1973). While this law is still controversial in physiological literature, it has merit when using a measure such as GSR or amplitude of response. To illustrate this law, consider two individuals, one having a basal level of 100 kohms and the other 20 kohms. When presented with the same stimulus, one has the potential to decrease 100 kohms while the other 20 kohms. If one compares only absolute raw

differences, then one would conclude that the 100 kohm decrease is a more powerful response than the 20 kohm response. In fact, the responses are equal: they are maximal. The extent of a response is dependent upon the level of resistance before the response, and is a function of the pre-stimulus level. The degree of this dependence varies from person to person, and from one physiological variable to another. When measuring a response to stimulation this pre-stimulus level must be considered (Sternbach, 1966).

Lykken et al. (1966, 1968) have suggested a means of measuring relative response in terms of a person's range of values on that variable. They found that the skin conductance level shown by one subject after a 30 minute rest period was twice as high as the maximum of another subject's while blowing up a balloon until it bursts. In comparing these two subjects, or any subjects, each one's conductance levels must be expressed in terms of his own maximum-minimum scale. The procedure of having subjects blow up a balloon until it burst has been found to be quite effective as a means for producing a maximum value of arousal; the minimum value being simpler to obtain (Lykken et al., 1968; Lykken and Venables, 1971; Venables and Christie, 1973).

Electrodermal Research in Education

Research in education has tended to avoid the use of the electrodermal response as a variable in studies dealing with arousal. There is a definite paucity of research in this area. In psychology, interest has been directed towards individual differences in electrodermal activity and performance in a number of situations. This work often uses the orienting response (OR). The OR is a "generalized response to mild or moderately

intense or novel stimuli, and habituates upon repetition of the stimuli" (Raskin, 1973, p. 128). Research indicates that the amplitude of the OR seems to be related to individual differences in attentional, and learning capacities (Maltzman and Raskin, 1965). High OR's have been found to show higher levels of semantic learning (i.e., paired-associate learning) for males (Raskin, 1969).

In one of the few educational studies using electrodermal responses, the OR patterns of readers and nonreaders were compared (Hunter et al., 1972). Twenty boys with a mean age of 9.9 years, of normal intelligence, and no severe motor, sensory, neurological, or emotional deficits, but who were retarded in reading an average of 2.4 years, served as the 'nonreader' group. Twenty control subjects, with a mean age of 9.11 years were matched with the nonreaders on sex, age, race, intelligence, and socioeconomic status. The children listened to fifteen ten second auditory tones (habituation), and were required to push a button for five tones with a fifty second interval (reaction time). The results, with regard to electrodermal activity, indicated that the nonreaders had lower mean skin conductance levels across all trials, and a higher skin resistance response amplitude on trial one of both tasks. Hunter et al. found that reading ability was significantly negatively correlated with motor reaction time, and skin resistance response amplitude to novel stimuli. The nonreader's lower skin conductance level is indicative of a lower arousal, or attentional level.

There is support for Hunter et al.'s findings in a study dealing with learning disabled children, in which Boydstun et al. (1968) found that these children were physiologically less reactive or attentive to 'meaningful' stimuli. Dureman and Pålshammar (1970) also found that

children rated as having low persistence in school work had significantly lower levels of skin conductance which decreased over time, as compared to children rated as being high in persistence. It has been suggested that nonreaders have an attentional deficit, or are "lacking in those specific arousal or emotive supports necessary for sustained attention and learning" (Dykman et al., 1970). It has also been suggested that perhaps dyslexia is a symptom of hereditary immaturity (Money, 1966; Johnson and Myklebust, 1967): physiological immaturity being one indication that this may be so.

Two studies more directly related to reading have been done by Carter (1950) and Proctor (1953). Carter used college students as subjects in his study. Groups of superior and inferior readers were asked to read aloud and the changes in their skin resistance were recorded as they read. Significant differences between groups were not found in total changes recorded during the entire time of reading, but the differences were significant on the more difficult paragraphs, the inferior readers showing the greatest drop in resistance on those paragraphs. However, certain methodological factors cast doubt on the validity of the conclusions. Firstly, this study took no account of differences in GSR reactivity that has been found to exist between intelligence groups. Reading ability correlates closely with intelligence, and therefore any technique comparing GSR of groups of differing intelligence must allow for this relationship if results are to be valid. Secondly, the subjects total loss of resistance in ohms is a statistically unsound basis for analysis since it is not equivalent from one to another, and is not equal over the total range of a distribution (Haggard, 1949). Lastly, the Carter study cannot confidently be applied to elementary school children.

Elementary school children were used in the Proctor study (1953),

which sought to discover if the GSR of below class readers were more frequent or of greater amplitude than either average or above class readers. Using a standardized test of silent reading ability, Proctor divided the children on the basis of word meanings and paragraph comprehension into three groups. The average readers were within a half (.5) standard deviation from the class mean, and the above class readers were more than one standard deviation above the class mean, and the below class readers were more than one standard deviation below the class mean. Sixty-three subjects of normal intelligence, as defined by Proctor, (75 IQ and above), twenty-one per group, were randomly selected. The subjects were required to do four tasks while being monitored on a polygraph: (1) read digits, (2) respond to a word association test, (3) read the Grays Oral Reading Paragraphs, and (4) do a digit span test. The GSR of each subject which exceeded one micromho in amplitude was counted. The results suggested that two types of anxiety were operating; for the good readers an anticipatory anxiety, and for the poor readers a frustration anxiety. No differentiation between the groups was found on the non text reading tasks. The reading tasks, when increasing in difficulty, resulted in an increase in the numbers of GSR's per group, with amplitude increasing significantly for the average and below class readers, but non-significantly for the above class readers. The relationship between increase in difficulty of reading materials and increased frequency of GSR was significantly greater for the below class readers than for the other groups. Increases in amplitude were significantly greater for the above class readers. Ekwall and English's (1971) results concur with Proctor's (1953) in that above average readers seem to show higher arousal initially and this level does not change significantly with the difficulty of the reading materials,

whereas the average or below average readers show greater GSR increases, indicating higher arousal as the difficulty of the material increases.

The IRI and Anxiety

The child, as he reads an IRI, ideally proceeds from an independent level, through the instructional level to the frustrational level. At the frustrational level, the child is unable to handle the material, and it is at this point that arousal or anxiety should be apparent. Knowing the child's frustrational reading level is suggested as being important for two reasons. Firstly, materials at this level should be avoided for the child's instruction. Secondly, to give the teacher some indication of the rate at which the child might progress when taught at his instructional level (Johnson and Kress, 1972). If the child is frustrated just one passage above his instructional level, then there is the possibility that much work will be needed in order for that child to progress. Faster progress is more likely when these two levels are spread more evenly, although this has not been shown empirically.

Emotional states, such as frustration and anxiety, and the accompanying physiological level of arousal, have a detrimental effect on learning efficiency (Spence and Spence, 1966; Spielberger, 1966). When a child has a reading problem it is difficult to ascertain whether anxiety was the causal factor, or the resultant factor, but the presence of high levels of anxiety is known to be detrimental to learning.

The effect of anxiety on school learning has been researched by Saranson et al. (1960, 1964, 1966). They studied 670 children from grade one through their fourth grade. In grades one and three the Test of Anxiety for Children (TASC), the Lie Scale for Children, and the Defensiveness Scale for Children were administered. Each year the Lorge-Thorndike

Intelligence Test was administered, and at the end of the second and fourth grades, achievement tests were administered. On the basis of the first grade testing, the children were specified as having either a high anxiety level (HA), or a low anxiety level (LA). At the third grade level they were again retested and anxiety levels again specified. The results of this massive study indicated that reading scores were most affected by high anxiety levels. The LA subjects showed higher mean achievement scores than the HA subjects. Boys classified as HA in grade one were, on the average, five months behind the LA boys by grade two, and eight months behind by grade four. For the HA girls, they were behind the LA girls an average of six months in grade two, and nine months by grade four.

Saranson et al. (1964) concluded:

First, the depressing effect of anxiety upon achievement test performance is stronger for both boys and girls than previous research--cross-sectional in nature--had indicated. Secondly, the differences in achievement scores are too large to be accounted for solely by the differences in IQ scores (p. 28).

With regards to intelligence and anxiety, Feldhusen and Klausmeier (1962) state:

Superior mental ability may make it possible for a child to assess more adequately the real and present danger in any current threatening object, situation, or person. Thus, his fears may be specific and ascertainable and unrelated to variations in an already high mental ability. Children of low and average IQ may show greater anxiety or generalized fear and a close relation between IQ and anxiety because of their limited intellectual capacity to evaluate the threat of danger in objects, situations, or people that confront them. In this study, an upper limit of intellectual advantages seemed to be reached at an IQ of 120, for a nonsignificant correlation was found with IQ's of 120-146 (cited in Saranson et al., 1964, p. 41).

This description of anxiety closely resembles the effect of reading ability and arousal that Proctor (1953) stated in his study, and to the results of Haggard's classic study of gifted children (1957).

Phillips (1967) also studied anxiety and its effect on school

achievement. Phillips's results support those of Saranson et al.'s, that a high level of anxiety is negatively correlated with reading achievement in the elementary grades; when anxiety decreases, reading achievement increases.

These studies highlight the importance of reducing high levels of anxiety, or physiological arousal, particularly in the reading situation. A child must be prevented from reading at his frustration level, where anxiety is high. In order to prevent this, his frustration level must be known and avoided. The IRI is one method for determining this level, but as the earlier part of this review suggested, the IRI is not without its problems. While most work has centered on the IRI and criteria for determining the instructional level, the frustration level has been practically ignored in research.

CHAPTER III

METHOD

Selection of the Subjects

The Stanford Achievement Test (1972) was administered to intact classrooms of second and third grade children in the schools allotted to the study by the Vancouver School Board. Level II of the test was administered to the second grade population and Level III to the third grade population. This involved 57 grade two children and 73 grade three children. From the results of this test the children were stratified into below average, average and above average reading groups (Table II).

TABLE II

CLASSIFICATION OF READING ACHIEVEMENT GROUPS ACCORDING
TO GRADE LEVEL

| | Below Average | Average | Above Average |
|---------|---------------|------------|---------------|
| Grade 2 | below 2.3 | 2.4 to 2.7 | above 2.8 |
| Grade 3 | below 3.2 | 3.3 to 3.8 | above 3.9 |

Parental consent forms (see Appendix) were sent home with each child. From the select population of children whose families consented to the study involving their children, thirty children were selected from each grade. Ten children were assigned to each of the reading achievement levels using a table of random numbers (Glass and Stanley, 1970, p. 510), as is shown in Table III.

TABLE III
SELECTION OF THE SUBJECTS

| | Reading Achievement Levels | | | |
|---------|----------------------------|---------|------------------|-------|
| | Below Average | Average | Above Average | Total |
| Grade 2 | 10 | 10 | 10 | 30 |
| Grade 3 | 10 | 10 | 10 | 30 |
| Total | 20 | 20 | 20 | 60 |

Materials

The Wide Range Achievement Test-Level I (Jastak, Bijou, and Jastak, 1965) was used to determine each subject's criterion arousal level (CAL).

The Diagnostic Reading Scales (Spache, 1963, 1972) were used as an informal reading inventory (IRI) in this study. The accurate readability ratings of the eleven passages, which range in difficulty from a 1.6 to 8.5 grade level, and secondly, the readability of these passages "appear to represent the closest approximation to equal-interval scaling among the tests analyzed" (Pennock, in press, 1976, p. 3) were the major reasons for choosing this test. This study was in accord with that of Ekwall and English (1971) who also used the Diagnostic Reading Scales as an IRI.

Apparatus

Each child's oral reading of the passages of the Diagnostic Reading Scales and his responses on the comprehension section for each passage were recorded to insure accuracy of scoring and test scoring verification.

Measuring Equipment: GSR Recording

The galvanic skin response (GSR) was continuously recorded on a Lafayette Model 7601TP D.C. amplifier with chart speed of 1mm/sec. The electrodes were placed on the dorsal surfaces of the index and ring fingers of the nonpreferred hand. Procedures were tested on seven children in order to develop testing techniques and scoring procedures. A polygraph expert examined the galvanic skin response profiles for sensitivity and accuracy of recording. These records showed a good response pattern.

Procedure

Each subject was escorted to the prepared room by the experimenter. Once in the room each was assured that the treatment would not be painful. Subjects were seated in a chair and the electrodes attached. During this time it was explained that there would be no pain or effects from the electrodes. Each subject was given an adaptation period during which the baseline level was established.

Two techniques were used to measure physiological arousal in each subject. Firstly, the subject was asked to read the words on the Wide Range Achievement Test-Level 1 (1965) in order to establish an arousal level based on performance in a reading task. Each child read until the GSR had plateaued and habituation began. Each subject's reaction was recorded on the GSR print-out and this measured in centimeters as the maximum deflection from the baseline level. This measure was used as the subject's criterion arousal level (CAL). Secondly, the subject was asked to blow up a balloon 'until it bursts'. This testing situation was terminated for each child before the balloon burst. The measure obtained served as an index of the child's maximum arousal level in a non-verbal

situation. In all children this measure provided a greater arousal level than did the Wide Range Achievement Test. This enabled the experimenter to compare how close the criterion arousal level was to the higher non-verbal arousal level.

After the subject had relaxed again, it was explained that he would read some passages orally, and be asked some questions about what he had read. The first passage was then given to the subject, and he read this orally. Upon completion, the booklet was taken away, and the comprehension questions asked. The subject was then administered the next highest level of the Diagnostic Reading Scales and the same procedure of measuring and recording followed. Each subject was to have continued reading through the graded passages of the Diagnostic Reading Scales until the GSR indicated that he had reached his criterion arousal level (CAL). In the data gathering situation the subjects did not exhibit the same characteristics as did the children used by the experimenter in testing the procedures. None of the subjects attained their criterion arousal level during the reading of the passages, regardless of passage difficulty. Because of this, frustrational level in reading had to be determined by the criteria as set forth in Johnson and Kress (1965).

For each passage read, a notation was made on the GSR print-out of the number of the passage and a dividing line made to separate the oral reading of each passage from the comprehension section.

Treatment of the Data

The passage at which the GSR indicated that the subject's CAL had been reached was to have been scored for word recognition and comprehension accuracy. An average word recognition and comprehension accuracy score

was to be computed using the total sample and these scores statistically compared to the criteria as set forth in Johnson and Kress (1965). This treatment of the data could not be completed since the frustrational level of reading had to be determined using the criteria stated in Johnson and Kress (1965) rather than from the criterion arousal level.

In order to analyze the data, an index of arousal for the IRI passages at the frustrational, pre-frustrational, and instructional levels was determined for both the oral reading and the comprehension sections, and a mean level of arousal computed for each subject at the frustrational level of reading. This was obtained by measuring the maximum point of deflection from the baseline level in centimeters. An average level of arousal was computed for the total sample and used in the comparisons of arousal levels between the instructional and pre-frustrational levels.

Statistical analyses were conducted using a two-way analysis of variance for a repeated measures design using ANOV:23 (1969).

Verification of scoring accuracy in the informal reading inventory was obtained. A graduate student in the Reading Education Department of U.B.C. scored the protocols of randomly selected subjects. He concurred with the experimenter's scoring. The polygraph advisor surveyed the profiles for accuracy in scoring procedures.

CHAPTER IV

RESULTS OF THE STUDY

Introduction

The purpose of this chapter is to report the data obtained from the administration of an informal reading inventory (IRI) while the subjects were monitored on the galvanic skin response (GSR). The IRI was administered to grade two and three children who had been stratified on reading achievement in order to answer certain questions regarding the frustrational level of reading and arousal levels of children reading at different achievement levels. Specifically, the purposes set for this study were:

1. To determine the word recognition and comprehension criteria for the frustrational reading level as determined by the galvanic skin response (GSR).
2. To compare the criteria for the frustrational reading level as established by Betts and Killgallon (1942) to the frustrational reading level as determined by the GSR.
3. To compare both word recognition and comprehension arousal levels, as measured by the GSR at the instructional, pre-frustrational, and frustrational reading levels of an informal reading inventory.
4. To compare the GSR measured frustrational reading levels and the percentages of word recognition and comprehension accuracy at three levels of reading achievement: below, average, and above average.
5. To compare the mean GSR arousal levels of boys and girls at the

frustrational level of reading.

In order to analyze the data the hypotheses were stated in the null form.

Hypothesis 1

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level and the level on the GSR where the 90 percent word recognition criterion on the IRI was applied.

During the reading of the Diagnostic Reading Scales passages the criterion arousal level as determined by the GSR was never attained by any subject in the sample. The frustrational level of reading was therefore determined using the criteria stated in Johnson and Kress (1965). This hypothesis became, therefore, untestable.

Hypothesis 2

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level and the level on the GSR where the 50 percent comprehension criterion on the IRI was applied.

For reasons specified above, this hypothesis was untestable.

Hypothesis 3

At the instructional level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

The observed F ratio (.026) shown in Table V failed to obtain significance at the .05 level of confidence. Hypothesis 3, therefore, was not rejected.

TABLE V
ANALYSIS OF VARIANCE: COMPARISON OF AROUSAL LEVELS
AT THE INSTRUCTIONAL LEVEL OF READING

| Source | SS | DF | MS | F |
|--------------------------------------|--------|----|--------|-------|
| Word Recognition to Comprehension | 21.375 | 1 | 21.375 | 0.026 |

This would mean that at the instructional level of reading there was no significant difference in the anxiety level between the oral reading of the passage and the comprehension section of the test.

Hypothesis 4

At the pre-frustrational level of reading there will be no significant difference between the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

For many subjects there were no reading passages between their instructional level and their frustrational level of reading. For this reason the pre-frustrational level of reading had to be abandoned in the analyses, thus making this hypothesis untestable.

Hypothesis 5

At the frustrational level of reading as determined by the subject's criterion arousal level (CAL) there will be no significant difference between the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

As explained previously, the frustration level of reading had to be determined by using the criteria set forth by Johnson and Kress (1965). As a result, this hypothesis was untestable. As a consequence the following hypothesis was generated:

Hypothesis 5(a)

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between

the GSR measured arousal level found during the oral reading and that found during the comprehension section of the test.

The observed F ratio (1.572) shown in Table VI failed to obtain significance at the .05 level of confidence. Hypothesis 5(a), therefore was not rejected.

This would mean that in the sample of children the level of anxiety did not appear to be different during the oral reading and the comprehension sections of the test at the frustrational level of reading as determined by the criteria set forth by Johnson and Kress (1965).

TABLE VI
ANALYSIS OF VARIANCE: COMPARISONS OF AROUSAL LEVELS
AT THE FRUSTRATIONAL LEVEL OF READING

| Source | SS | DF | MS | F |
|---------------------------------------|----------|----|----------|-------|
| Word Recognition to Comprehension | 675.726 | 1 | 675.726 | 1.572 |
| Between Reading Achievement Groups | 9717.582 | 2 | 4858.789 | 0.884 |
| Sex | 9332.359 | 1 | 9332.359 | 1.739 |

Hypothesis 6

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading at grade level.

For reasons specified above, this hypothesis was untestable. As a consequence the following hypothesis was generated.

Hypothesis 6(a)

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading at grade level.

The observed F ratio (.884) shown in Table VI failed to obtain significance at the .05 level of confidence. Hypothesis 6(a), therefore, was not rejected.

This would mean that in the sample of children anxiety level did not appear to be different for children reading above average compared to children reading at grade level.

Hypothesis 7

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading below average.

For reasons specified above, this hypothesis was untestable. As a consequence the following hypothesis was generated.

Hypothesis 7(a)

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading above average compared to children reading below average.

The observed F ratio (.884) as shown in Table VI failed to obtain significance at the .05 level of confidence. Hypothesis 7(a), therefore, was not rejected.

This would mean that in the sample of children anxiety level did not appear to be different for children reading above average compared to children reading below average.

Hypothesis 8

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the GSR measured arousal level of children reading at grade level compared to children reading below average.

For reasons specified above, this hypothesis was untestable. As a consequence the following hypothesis was generated.

Hypothesis 8(a)

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of children reading at grade level compared to children reading below average.

The observed F ratio (.884) as shown in Table VI failed to obtain significance at the .05 level of confidence. Hypothesis 8(a), therefore, was not rejected.

This would mean that in the sample of children anxiety level did not appear to be different for children reading at grade level compared to children reading below average.

Hypothesis 9

At the frustrational level of reading as determined by the subject's CAL there will be no significant difference between the boys and the girls.

For reasons specified above, this hypothesis was untestable. As a consequence the following hypothesis was generated.

Hypothesis 9(a)

At the frustrational level of reading as determined by the administration of an IRI there will be no significant difference between the GSR measured arousal level of the boys and the girls.

The observed F ratio (1.739) as shown in Table VI failed to obtain significance at the .05 level of confidence. Hypothesis 9(a), therefore, was not rejected.

This would mean that in the sample of children anxiety level did not appear to be different for the boys and the girls.

CHAPTER V

SUMMARIES, CONCLUSIONS AND RECOMMENDATIONS

Summary of Methods

A review of the literature indicated that the instructional level of an informal reading inventory has been empirically researched over the years, with recent studies suggesting that the original performance criteria established by Betts and Killgallon may be too stringent for children in grades one through three. The criteria for determining the frustrational level of an informal reading inventory (IRI) have been largely ignored in the research.

Behavioral characteristics for the three levels of the IRI (independent, instructional, and frustrational) are often cited with the performance criteria for determining a level, but lack substantiation by empirical data. At the independent and instructional levels, Johnson and Kress (1965) state that the reading should be free from such signs as lip movements, finger pointing, head movements, subvocalizations, and anxiety about performance; these behavior characteristics being more likely to occur at the frustrational level of reading, where the reading is difficult and therefore likely to produce anxiety in the child.

The use of the galvanic skin response (GSR) seemed an appropriate instrument by which to validate empirically the criteria for the frustrational level of reading in an IRI. The GSR is commonly used to determine high levels of arousal, particularly with regard to truthfulness and

deceitfulness.

Ekwall and English (1971) used the GSR, as well as other physiological responses, to validate the criteria used to determine the frustrational level of reading in an IRI. A polygraph expert was used to predetermine each subject's high arousal level. Each subject read graded passages while being monitored on the polygraph and "at the frustration reading level stipulated by the polygraph examiner, the percentages of oral errors and comprehension errors were computed" (p. 16). Using this method the authors concluded, among other things, that the 90 percent word recognition criterion was valid if repetitions were scored, and that the 50 percent comprehension criterion was adequate.

The present study sought to investigate the GSR arousal levels of children reading at the mid grade one and above level, as they read through the graded passages of an IRI. The purposes of this study were:

1. To determine the word recognition and comprehension criteria for the frustrational reading level as determined by the galvanic skin response (GSR).
2. To compare the criteria for the frustrational reading level as established by Betts and Killgallon (1942) to the frustrational reading level as determined by the GSR.
3. To compare both word recognition and comprehension arousal levels, as measured by the GSR at the instructional, pre-frustrational, and frustrational reading levels of an IRI.
4. To compare the GSR measured frustrational reading levels and the percentages of word recognition and comprehension accuracy at three levels of reading achievement: below, average, and above average.
5. To compare the mean GSR arousal levels of boys and girls at the

frustrational level of reading.

The sample under study consisted of 60 children selected from grades 2 and 3 (30 from each grade). Parental consent was required in order for a child to be considered for inclusion in this study.

The Stanford Achievement Test (1972) was administered to all second and third grade children in order to determine the three reading achievement levels, as shown in Table II (p. 8). Level II of the Stanford Achievement Test (1972) was administered to the second grade children, and Level III to the third grade children. The children for whom parental consent for inclusion in the study had been obtained were stratified by grade and reading achievement level and then subjects were randomly selected for the study by means of a random number table. The children for each of the three reading achievement levels for each grade were selected (see Table III, p. 9).

Each subject was escorted to the testing room where the electrodes were attached to the index and ring fingers of the nonpreferred hand. After an attempt to establish rapport was undertaken, the subject was given the Wide Range Achievement Test-Level 1(1965) and asked to read the words. The child read until the galvanic skin response (GSR) had plateaued and habituation had begun. Each subject's reaction was measured on the GSR print-out as the maximum deflection in centimeters from the baseline level, and this used as the subject's criterion arousal level (CAL).

After the subject had again relaxed, he was given the first passage of the Diagnostic Reading Scales (1972) to read orally. Upon completion, the booklet was removed and the comprehension questions asked. The next highest level was then administered and the same procedure followed until the GSR indicated that the subject's CAL had been reached. Each subject's

oral reading and responses were tape recorded to insure accuracy of scoring.

Word recognition and comprehension accuracy scores were obtained for the reading passage where the GSR indicated that the subject's CAL had been reached. An average word recognition and comprehension score was computed for the total sample and these scores statistically compared to the criteria set forth by Johnson and Kress (1965).

An index of arousal for the IRI passages at the instructional and pre-frustrational levels were determined for both the oral reading and the comprehension sections. This was obtained by measuring the maximum point of deflection from the baseline in centimeters. An average level of arousal was computed for the total sample and used in the comparisons between the instructional and pre-frustrational levels.

Statistical analyses were conducted using a two-way analysis of variance for a repeated measures design using ANOV-23 (1969).

Summary of Findings and Conclusions

The use of the Wide Range Achievement Test (1965) (WRAT) in determining each subject's criterion arousal level proved to be ineffectual. The WRAT did yield a high arousal level index, however, this level was never attained by any child during the reading of the passages in the Diagnostic Reading Scales (1972), regardless of passage difficulty. Because of this, the frustrational level of reading had to be determined by using the criteria set forth by Johnson and Kress (1965).

Statistical analyses of the arousal levels at the frustrational level of reading were nonsignificant for comparisons between word recognition and comprehension, the three reading achievement levels, and the boys

and the girls.

At the instructional level of reading, statistical analysis of the arousal level found during the oral reading compared to that found during the comprehension section also yielded nonsignificant results.

In conclusion, none of the nine null hypotheses was rejected.

Implications

Some important points can be inferred from the results of this study. As the review of the literature indicated, few empirical studies in reading education have been conducted using the galvanic skin response. Of those conducted, the purpose and method of a study by Ekwall and English (1971) bore the most resemblance to the present study. In their study they were primarily attempting to validate the criteria for the frustration level of reading by means of the polygraph. Each subject read graded materials while being monitored on a polygraph. Recording units used included breathing unit, the galvanic skin response unit, and the plethysmograph (pulse rate) unit. Ekwall and English (1971) explain that:

The polygraph expert, and a neurologist then interpreted each polygraph test to indicate the point of apparent frustration as the students progressed through the reading passages. Tracings were graded individually by comparing the pre-test norm against the magnitude of changes in amplitude, baseline, rate and rises. At the frustration reading level stipulated by the polygraph examiner, the percentages of oral errors and comprehension errors were computed (p. 16).

Ekwall and English (1971) fail to report two important factors: the model of polygraph used in their study, and how the pre-test norm was obtained.

The present study, as stated in Chapter III, used a portable polygraph, and was concerned with only the galvanic skin response unit on the machine. It is quite possible that Ekwall and English (1971) used a more

powerful polygraph which would be more sensitive to actual changes in the responses being measured. Their use of three physiological responses measured during each subject's reading would possibly give them more information as to when the subject was frustrated, and perhaps one physiological indication of frustration was adequate to deem the child reading at frustrational level.

Ekwall and English (1971) also obtained a pre-set norm to which they compared the subject's physiological arousal during the reading in order to determine the subject's frustrational reading level. No mention was made, however, of how this was established. The present study attempted to obtain a pre-set norm by having each subject read difficult words, a task felt to be related to difficult reading. This pre-set norm, however, was never attained during the reading of the passages, regardless of passage difficulty.

The lack of full methodological reporting in the Ekwall and English (1971) study made it difficult to critically assess, and impossible to accurately replicate. The present study failed to verify their findings. Four possible reasons can be forwarded. Firstly, their polygraph may have been more sensitive to actual physiological changes in their subjects. Secondly, the frustrational level in their study was not determined by the galvanic skin response alone, but rather, in conjunction with two other physiological responses, of which all or one may have been necessary in order to indicate the frustrational level. Thirdly, their pre-set norm may have been a more reliable or attainable indicator of frustration. Lastly, magnitude of change in amplitude, baseline, rate and rises were used to measure the physiological responses, whereas the present study used only the baseline level and magnitude of change in amplitude. These possible explanations

highlight the importance of accurate reporting of empirical research.

The performance criteria are the primary means for determining the frustrational level of reading. Behavioral characteristics, as mentioned previously, may also be used to aid in determining this level. While there is no empirical validation of the reliability or accuracy of the commonly reported behavioral characteristics, the present experimenter did notice a change in the reading behavior of many subjects as the materials became more difficult. These behaviors sometimes happened prior to the performance criteria being met. While these behaviors were not of the intensity described by Ekwall (1976), they were apparent enough to be noticed by an unfamiliar examiner.

The informal reading inventory is basically intended for the use of classroom teachers, who would be more familiar with their pupils reading and reading behaviors. Classroom teachers may find that the behavioral characteristics of the pupils are a more feasible, and perhaps, more accurate way of determining this level, as opposed to rigid adherence to the performance criteria. It may be that teachers should be trained to be aware of the behavioral characteristics for the frustrational level of reading and use them as another clue in assessing a child.

The present study was not able to offer empirically support for the performance criteria used to determine the frustrational level of reading. For reasons discussed earlier, the results of the study by Ekwall and English (1971) were not replicated. The lack of significant results in this study did, however, have an important point to make concerning empirical research. From the data that were obtained it would appear that the reading of isolated word lists generates considerably more arousal in a reader than does the reading of prose. It may be that when

a child reads words for which he has no meaning referent this becomes frustrating where upon reading prose where a degree of understanding is taking place does not produce the same degree of arousal. In its own way this study might give support to the emphasis that reading educators place upon meaning as a basis of all developmental programmes.

Recommendations for Further Study

Further research resulting from this study would be:

1. A replication study using a more powerful polygraph, and scoring of the galvanic skin response for both frequency and amplitude of response might yield significant results.
2. A study to investigate the accuracy of the behavioral characteristics in determining the frustrational level of reading might prove useful.

BIBLIOGRAPHY

- Analysis of Variance Programme 23. University of Alberta, Edmonton, 1969.
- Beldin, H. O. Informal reading testing: Historical review and review of the research. In W. K. Durr (Ed.), Reading Difficulties: Diagnosis, Correction and Remediation. Newark: International Reading Association, 1970.
- Betts, E. A. The Prevention and Correction of Reading Difficulties. Evanston: Row, Peterson, 1936.
- Betts, E. A. Foundations of Reading Instruction. New York: American Book Company, 1957.
- Bolenious, E. M. Teacher's Manual of Silent and Oral Reading. Boston: Houghton Mifflin, 1919.
- Bond, G. L., & Tinker, M. A. Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, Inc., 1967.
- Botel, M. Newer techniques to the supervision of reading instruction. In Frontiers of Elementary Education. Syracuse, New York: Syracuse University Press, 1957, pp. 26-72.
- Botel, M. Ascertaining instructional levels. In S. H. Figurel (Ed.) Forging Ahead in Reading, 1969, 12, 170-175.
- Botel, M., Bradley, J., & Kashuba, M. The validity of informal reading testing. In W. K. Durr (Ed.), Reading Difficulties: Diagnosis, Correction and Remediation. Newark: International Reading Association, 1970, pp. 85-103.
- Boydston, J. A., Ackerman, P., Stevens, D. A., Clements, S. D., Peters, J. E., & Dykman, R. A. Physiologic and motor conditioning and generalization in children with minimal brain dysfunction. Conditional Reflex, 1968, 3, 2, 81-104.
- Busboom, J. B. An Investigation of the Relationship Among Various Testing Techniques on an Informal Reading Inventory at Instructional and Frustrational Level. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1974.
- Carter, H. L. J. A combined oral reading and psychogalvanic response technique for investigating certain reading abilities of college students. J. Applied Psychol., 1950, 34, 267-269.
- Chaplin, J. P. Dictionary of Psychology. New York: Dell Publishing Company, Inc., 1968.

- Cooper, J. L. The Effect of Adjustment of Basal Reading Materials on Reading Achievement. Unpublished doctoral dissertation, Boston University, 1952.
- Daniels, J. The effectiveness of various procedures in reading level placement. Elementary English, 1962, 39, 590-600.
- Davis, E. E. Reading frustration level as indicated by the polygraph. J. of Educ. Research, 1975, 286-288.
- Duffy, E. Activation and behavior. New York: Wiley, 1962.
- Dunkeld, C. G. The validity of the informal reading inventory for the designation of instructional reading levels: A study of the relationships between children's gains in reading achievement and the difficulty of instructional materials. Dissertation Abstracts, 1970, 31-A, 11, 6274-A.
- Dureman, I., & Pålshammar, Å. Differences in tracking skills and physiological activation dynamics in children high or low in persistence in schoolwork. Psychophysiology, 1970, 7, 1, 95-102.
- Durrell, D. Diagnosis and their implication with respect to instruction in reading. In National Society for the Study of Education, 36th Yearbook, Part I. Chicago: University of Chicago Press, 1937, pp. 325-356.
- Durrell, D. Improvement of Basic Reading Abilities. New York: World Book, 1940.
- Durrell, D. Durrell Analysis of Reading Difficulty: Manual of Directions. New York: Harcourt, Brace & World, Inc., 1955.
- Dykman, R. A., Walls, R. C., Ackerman, P., & Peters, J. E. Children with learning disabilities: conditioning differentiation, and the effect of distraction. Amer. J. of Orthopsychiatry, 1970, 40, 5, 766-782.
- Ekwall, E. E. Informal reading inventories: the instructional level. The Reading Teacher, 1976, 29, 7, 662-665.
- Ekwall, E. E., & English, J. K. Use of the Polygraph to Determine Elementary School Students' Frustration Reading Level: Final Report. U.S. Department of Health, Education, and Welfare, 1971.
- Ekwall, E. E., Solis, J. K., & Solis, E. Investigating informal reading inventory scoring criteria. Elementary English, 1973, 50, 2, 271-274, 323.
- Farr, R., & Anastasiow, N. Tests of Reading Readiness and Teaching: A Review and Evaluation. Newark: International Reading Association, 1969.

- Farr, R. Reading: What Can Be Measured? Newark: International Reading Association, 1969.
- Gates, A. The Improvement of Reading. New York: The MacMillan Company, 1936.
- Gates, A. The Improvement of Reading. New York: The MacMillan Company, 1947.
- Glass, G. V., & Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, N. J.: Prentice-Hall Inc., 1970, Appendix A, Table A, p. 510.
- Gray, W. S. Value of informal test of reading accomplishment. J. of Educational Research, 1920, 1, 2, 103-111.
- Grossman, S. P. A Textbook of Physiological Psychology. New York: John Wiley & Sons, Inc., 1967.
- Haggard, E. A. On the application of analysis of variance to GSR data: I. The Selection of an appropriate measure, II. Some Effects of the use of inappropriate measures. J. of Experimental Psychol., 1949, 39, 378-392, 861-867.
- Haggard, E. A. Socialization, personality, and achievement in gifted children. School Review, 1957, Winter Issue, 318-441.
- Hunter, E., Johnson, L., & Keefe, F. Electrodermal and cardiovascular responses in nonreaders. J. of Learning Disabilities, 1972, 5, 4, 187-197.
- Jastak, J. F., Bijou, S. W., & Jastak, S. R. Wide Range Achievement Test. Delaware: Guidance Association of Delaware, Inc., 1965.
- Johnson, D. J., & Myklebust, H. R. Learning Disabilities: Educational Principles and Practices. New York: Grune and Stratton, 1967.
- Johnson, M. S., & Kress, R. A. Informal Reading Inventories. Newark: International Reading Association, 1965.
- Johnson, M. S., & Kress, R. A. Individual reading inventories. In L. Schell, & P. Burns (Eds.) Remedial Reading: Classroom and Clinic. Boston: Allyn & Bacon, Inc., 1972, pp. 185-206.
- Kender, J. P. How useful are informal reading tests? J. of Reading, 1968, 11, 5, 337-342.
- Killgallon, P. A. A Study of Relationships Among Certain Pupil Adjustments in Language Situations. Unpublished doctoral dissertation, Pennsylvania State University, 1942.

- Landis, C. Electrical phenomena of the skin. Psychol. Bull., 1932, 26, 693-752.
- Landis, C., & DeWicke, H. M. The electrical phenomena of the skin. Psychol. Bull., 1929, 26, 64-119.
- Lowell, R. E. Problems in identifying reading levels with informal reading inventories. In W. K. Durr (ed.), Reading Difficulties: Diagnosis, Correction and Remediation. Newark: International Reading Association, 1970, pp. 120-126.
- Lykken, F. T., Rose, R., & Luther, R. Correcting physiological measures for individual differences in range. Psychol. Bull., 1966, 66, 6, 481-484.
- Lykken, D. T., Miller, R. D., & Strahan, R. F. Some properties of skin conductance and potential. Psychophysiology, 1968, 5, 3, 253-268.
- Lykken, D. T., & Venables, P. H. Direct measurement of skin conductance: A proposal for standardization. Psychophysiology, 1971, 8, 656-672.
- Maltzman, I., & Raskin, D. C. Effects of individual differences in the orienting reflex on conditioning and complex processes. J. of Exp. Rsch. in Pers., 1965, 1, 1-6.
- McCracken, R. A. The development and validation of the Standard Reading Inventory for the appraisal of reading performance from grades one through six. Improvement of Reading Through Classroom Practice. Proceedings of the I.R.A. Annual Convention, 1964, 9, pp. 310-313.
- McGinnis, I. D. Making the most of informal inventories. In D. DeBoer (Ed.), Reading Diagnosis and Evaluation. Chicago, 1970, pp. 93-99.
- McKinlay, D. C. The Establishment of a Revised Word Recognition Accuracy and Oral Comprehension Criteria for the Instructional Level of the I.R.I. Unpublished masters thesis, The University of British Columbia, 1973.
- McRae, L. G. A Measure of the Reliability of the Informal Reading Inventory. Unpublished doctoral dissertation, University of Northern Colorado, 1974.
- Millsap, L. N. A Study of Teacher's Awareness of Frustration Reading Level Among Their Pupils in Basal Readers. Unpublished doctoral dissertation, University of Oregon, 1962.
- Money, J. (ed.) The Disabled Reader. Baltimore: John Hopkins Press, 1966.
- Neumann, E., & Blanton, R. The early history of electrodermal research. Psychophysiology, 1970, 6, 4, 453-475.

- Pennock, C. D. Selecting instructional materials to meet students' reading performance level. The Alberta J. of Educ. Rsch, 1973, 19, 30-36.
- Pennock, C. D. Diagnostic Reading Scales (DSR)-1972 Revision. in press, 1976.
- Phillips, B. N. Anxiety as a function of early school experience. Psychol. in the Schools, 1967, 4, 4, 335-340.
- Powell, W. R. Reappraising the Criteria for Interpreting Informal Inventories. Unpublished paper presented at the Thirteenth Annual International Reading Convention, Boston, 1968.
- Powell, W. R., & Dunkeld, C. G. Validity of the I.R.I. reading levels. Elementary English, 1971, 48, 637-642.
- Proctor, P. W. The Galvanic Skin Response as a Function of Differences between Individual Reading Achievement: A Comparative Study of the Amplitude and Frequency of the Galvanic Skin Response of Children in a Reading Situation Who Have Reading Ability Below, Equal to, and Above the Average for their Class. Unpublished doctoral dissertation, New York University, 1953.
- Raskin, D. C., Kotses, H., & Bever, J. Autonomic indicators of orienting and defensive reflexes. J. of Experimental Psychol., 1969, 80, 423-433.
- Raskin, D. C. Semantic and generalizations of autonomic responses. J. of Experimental Psychol., 1969, 79, 69-76.
- Raskin, D. C. Attention and arousal. In W. F. Prokasy, & D. C. Raskin (Eds.), Electrodermal Activity in Psychological Research. New York: Academic Press, 1973, pp. 125-156.
- Rugel, R. P. Arousal and levels of reading difficulty. Reading Teacher, 1971, 24, 5, 458-460.
- Sarason, S., Davidson, K., Lighthall, F., Waite, R., & Reubush, B. K. Anxiety in Elementary School Children. New York: John Wiley & Sons, Inc., 1960.
- Sarason, S., Hill, K., & Zimbardo, P. A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. Monographs of the Society for Research in Child Dev., 1964, 29, 7, No. 98.
- Sarason, S., & Hill, K. The relation of test anxiety and defensiveness to test and school performance over the elementary school years: A further longitudinal study. Monographs of the Society for Research in Child Dev., 1966, 31, 2, No. 104.

- Scholander, T. Some measures of electrodermal activity and their relationships as affected by varied temperatures. J. of Psychosomatic Rsch., 1963, 7, 151-158.
- Shmavonian, B. M., Yarmat, A. J., & Cohen, S. I. Relationships between the autonomic nervous system and central nervous system in age differences in behavior. In A. T. Welford, and J. E. Birren (Eds.), Behavior Aging and the Nervous System. Springfield, Illinois: Thomas, 1965.
- Shmavonian, B. M., Miller, L. H., & Cohen, S. Differences among age and six groups in electrodermal conditioning. Psychophysiology, 1968, 5, 119-131.
- Sipay, E. R. A comparison of standardized reading scores and functional reading levels. Reading Teacher, 1964, 17, 265-268.
- Spache, G. D. Diagnostic Reading Scales. Monterey: California Test Bureau, 1963.
- Spache, G. D. Diagnostic Reading Scales. Monterey: California Test Brueau, 1972.
- Spence, J. T., & Spence, K. W. The motivational components of manifest anxiety: drive and drive stimuli. In C. D. Spielverger (Ed.), Anxiety and behavior. New York: Academic Press, 1966.
- Spielberger, C. D. The effects of anxiety on complex learning and academic achievement. In C. D. Spielberger (Ed.), Anxiety and behavior. New York: Academic Press, 1966.
- Sternbach, R. A. Principles of Psycholphysiology. New York: Academic Press, 1966.
- Truman, L. K., Nadden, R., Gardner, E., Terman, L. M., & Ruch, G. M. The Stanford Achievement Test. New York: Harcourt Brace Jovanovich, Inc., 1972.
- Utsey, J., Wallen, C., & Beldin, H. O. Diagnostic Techniques in Teaching Reading. Unpublished training manual, University of Oregon, 1965.
- Venables, P. H., & Christie, M. J. Mechanics, instrumentation, recording techniques, and quantification of responses. In W. F. Prokasy, & D. C. Raskin (Eds.), Electrodermal Activity in Psychological Research. New York: Academic Press, 1973, pp. 2-124.