INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
HELWIG, John Joseph, 1936-
MEASUREMENT OF VISUAL-VERBAL FEEDBACK ON CHANGES
IN MANUSCRIPT LETTER FORMATION.

The Ohio State University, Ph.D., 1975
Education, special

Xerox University Microfilms, Ann Arbor, Michigan 48106

© Copyright by
John Joseph Helwig
1975
MEASUREMENT OF VISUAL-VERBAL FEEDBACK ON
CHANGES IN MANUSCRIPT LETTER FORMATION

DISSERTATION

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

By

John Joseph Helwig, B.S., M.A.

The Ohio State University

1975

Reading Committee:
John O. Cooper
Thomas M. Stephens
Daryl Siedentop

Approved by

John O. Cooper
Adviser
Department for Exceptional Children
To my GOD

"My Strength"
ACKNOWLEDGMENTS

I wish to thank my committee for their guidance, counsel, and encouragement. I want to especially thank Dr. Daryl Siedentop for taking time to give me feedback about the research design and my writing skills on past research, Dr. Thomas M. Stephens for his continued expression of faith in me over the last three years. His personal involvement and support of students will be a model for me to follow years after I leave Ohio State University. Dr. John O. Cooper, my adviser, has spent many unselfish hours in reading, analyzing, and recommending changes in this Dissertation. A special appreciation to Dr. Walter Barbe for the hours spent discussing handwriting and the special art work so generously donated, and the Zaner-Bloser Company, Columbus, Ohio, for the materials furnished, professional advice and the grant necessary for this study. A special thanks to two friends, James Norman and John Johns, for the many hours spent collecting data. Their willingness to sacrifice many hours contributed generously to this study.

To my wife, Marjorie, and my three children Paul, John, and Kim, my thanks for their continual support and encouragement and their constant willingness to help in whatever
endeavors necessary while I wrote this final draft. With this Dissertation completed, I can return to the role of husband and father.
VITA

February 22, 1936 .......... Born - Grand Rapids, Michigan

1961 ..................... B.S. Education., Western Illinois University, Macomb, Illinois

1960-1972 ................. Classroom Teacher, Keokuk Junior High, Keokuk, Iowa

1972-1973 ................ Graduate Research Assistant, Department for Exceptional Children, The Ohio State University, Columbus, Ohio.

1973 ....................... M.A. Education, The Ohio State University, Columbus, Ohio

1973-1975 ................ Graduate Research Assistant, Department for Exceptional Children, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Learning and Behavioral Disorders

Studies in Behavioral Analysis.  Professor John O. Cooper

TABLE OF CONTENTS

DEDICATION ......................................... ii
ACKNOWLEDGMENTS ....................................... iii
VITA .................................................... v
LIST OF TABLES .................................... viii
LIST OF FIGURES ..................................... ix

CHAPTER

I. INTRODUCTION ..................................... 1
   Background ........................................ 1
   Statement of the Problem ......................... 4
   Operational Definition of Terms .................. 7

II. REVIEW OF THE LITERATURE ..................... 10
   Introduction ...................................... 10
   Characteristics of Cursive and Manuscript
      Handwriting ..................................... 11
   Legibility of Handwriting ......................... 13
   Measurement of Legibility ......................... 15
   Reliability of Handwriting Scales ................. 23
   Individual Factors Affecting the
      Legibility of Handwriting ...................... 27
   Legibility of Letter Forms ....................... 35
   How Perception Effects Legibility ............... 40
   Measures of Selected Physiological and Motor
      Correlates of Handwriting Legibility .......... 45
   Summary .......................................... 50

III. METHODS AND PROCEDURES .................... 52
   Subjects .......................................... 53
   Setting .......................................... 53
   Materials ........................................ 54
   Behavior Definition and Recording ............... 58
## CHAPTER

<table>
<thead>
<tr>
<th>Design Conditions</th>
<th>Design</th>
<th>Data Collection and Analysis</th>
<th>Interobserver Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### IV. PRESENTATION, ANALYSIS AND DISCUSSION OF DATA

<table>
<thead>
<tr>
<th>Interobserver Agreement</th>
<th>Presentation and Analysis of Data</th>
<th>Individual Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### V. DISCUSSION, LIMITATIONS, IMPLICATIONS, AND SUMMARY

<table>
<thead>
<tr>
<th>Discussion</th>
<th>Limitations</th>
<th>Implications</th>
<th>Recommendations for Further Research</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## APPENDIX

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>BIBLIOGRAPHY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### BIBLIOGRAPHY
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Average Number of Words Read Per Fixation, Average Duration of Fixation, and Average Number of Words Read Per Regression For Selections 1 to 11</td>
<td>30</td>
</tr>
<tr>
<td>2. Condition in Which Interobserver Observations Occurred Range and Average Percentage of Agreement</td>
<td>72</td>
</tr>
<tr>
<td>3. Percentage of Correct Manuscript Letter Strokes</td>
<td>75</td>
</tr>
<tr>
<td>4. Summary of Interobserver Agreement Measures of 40 Samples of Manuscript Letters</td>
<td>76</td>
</tr>
<tr>
<td>5. Mean Number of Correct Responses Out of a Possible Forty-Two</td>
<td>102</td>
</tr>
<tr>
<td>6. Percentage of Correct Responses for Untrained Letters</td>
<td>103</td>
</tr>
<tr>
<td>7. Percentage of Correct Responses for Trained Letters in Sentence Writing</td>
<td>104</td>
</tr>
<tr>
<td>8. Percentage of Correct Responses for Untrained Letters in Sentence Writing</td>
<td>104</td>
</tr>
<tr>
<td>9. Summary of Strokes Measured and Recorded Per Minute By Six Naive Evaluators</td>
<td>119</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Hypothetical Example of Multiple Baseline Design</td>
<td>67</td>
</tr>
<tr>
<td>2.</td>
<td>Percentage of Correct Letter Strokes of Six Subjects During Baseline Conditions</td>
<td>74</td>
</tr>
<tr>
<td>3.</td>
<td>Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Session (Subject One)</td>
<td>78</td>
</tr>
<tr>
<td>4.</td>
<td>Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject One)</td>
<td>79</td>
</tr>
<tr>
<td>5.</td>
<td>Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Sessions (Subject Two)</td>
<td>82</td>
</tr>
<tr>
<td>6.</td>
<td>Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject Two)</td>
<td>83</td>
</tr>
<tr>
<td>7.</td>
<td>Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Session (Subject Three)</td>
<td>86</td>
</tr>
<tr>
<td>8.</td>
<td>Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject Three)</td>
<td>87</td>
</tr>
<tr>
<td>9.</td>
<td>Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Sessions (Subject Four)</td>
<td>90</td>
</tr>
<tr>
<td>10.</td>
<td>Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject Four)</td>
<td>91</td>
</tr>
<tr>
<td>11.</td>
<td>Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Sessions (Subject Five)</td>
<td>95</td>
</tr>
<tr>
<td>12.</td>
<td>Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject Five)</td>
<td>96</td>
</tr>
</tbody>
</table>
13. Per cent of Trained and Untrained Letter Strokes Meeting Criterion by Sessions (Subject Six) ... 99

14. Total Number of Trained Letter Strokes Meeting Criterion During Each Session (Subject Six) ... 100
CHAPTER I

INTRODUCTION

Background

Starch (1913) considered "handwriting as a mode of conventional expression created by man and necessitating an apprenticeship of the hand in the same way as sewing and knitting with needles." Freeman (1954) stated "the ultimate purpose of handwriting is to express and communicate meaning." Webster's New World Dictionary (1971, p. 340) defines handwriting as "a style or way of forming letters and words by hand with pen, pencil, etc. when writing." The purpose of handwriting then is to communicate with others.

Handwriting is a common skill taught to all children receiving a formal education. It is a set of symbols used to communicate with others. These symbols used to convey our message must be easily read and interpreted by the person or persons for whom it was written. Starch (1913) considered that "handwriting is not an end in itself but is purely an agency of communication and therefore can be measured or evaluated in terms of utility." Educators have not completely agreed on the chief elements of the term "utility." When utility is considered synonymous with
useability, while using handwriting as a mode of communication, then its ease in reading a person's handwriting sample has been a way of measuring its legibility (Ayres, 1912; Quant, 1946; Craig, 1966). The characteristics of a person's handwriting that determine its readability or legibility are many. Thorndike (1910) considered legibility as a word with many possible meanings. One meaning, legibility or writing may be measured by the distance at which it can be read with a given accuracy and rate. In another, it can be measured by the rate at which a sample can be read with a given accuracy and rate. In another, it can be measured by the rate at which a sample can be read at a given distance and with a given accuracy. Still another, measure could be by the degree of fatigue or of discomfort resulting from reading a handwriting sample. Lehman (1973) stated:

As a quality of handwriting, legibility denotes the formal adequacy of the letters to communicate, and results from high formal correspondence between the handwriting and a model that the reader is prepared to recognize. Legibility is a natural consequence of handwriting produced with distinctive, simple letter models in proportioned form.

The measurement of manuscript and cursive handwriting legibility has mostly been by three procedures. The first major attempt to measure the legibility of handwriting was by Thorndike (1910). He developed a set of handwriting scales consisting of samples of handwriting from very bad to
very good. His work was quickly followed by two new sets of scales, Ayres (1912) and Freeman (1915). There were six additional scales developed during the seven years following Freeman's but they were duplications or extensions of existing scales. The only major scale development through 1956 was by West (1926) and the revisions by West (1957) and Freeman (1959) of earlier scales. Harris, Herrick and Rarick (1961) and Bezzi (1962) developed the first new major scales in thirty-five years.

Feldt (1962), Anderson (1965), and Askor, Otto and Askor (1971) found scales developed through 1959 to be unreliable when evaluating handwriting legibility. Research evaluating the reliability of Harris, Herrick and Rarick (1961) and Bezzi (1962) scales was not available. Quant (1946) and Craig (1966) approached the measurement of legibility by measuring the readability of handwriting. Quant measured eye fixations while Craig used oral reading as his criterion for reliability. Lehman (1973) evaluated the legibility of handwriting by measuring deviations of certain letter characteristics from a norm. Freeman (1954, p. 28) stated that "legibility depends in the first instance on the formation of the individual letters."

Studies by Quant (1946), Craig (1966), and Rondinella (1971) showed that incorrect letter formation contributed more to illegible handwriting than any other single factor. Lewis and Lewis (1969) and Newland (1969) measured the
numbers and types of incorrect letter forms. Lewis and Lewis studied manuscript letter forms while Newland examined cursive letters. Newland found that the letters a, e, r, and t contributed an average forty-six per cent of the illegibilities measured.

Many evaluative devices, mostly subjective, have been developed to measure both handwriting as a whole and the set of individual characteristics comprising handwriting. However, determining the degree of legibility samples of handwriting exhibited has eluded educators. Whenever an objective way of showing change in letter formation is discovered and combined with teaching correct letter form, a promising way to increase legibility may be available to teachers.

Statement of the Problem

The review of the research on measurement of handwriting legibility indicates that changes in letter formation have the greatest effect on changes in legibility. Most students learn to form letters by copying a model commercially prepared or constructed by the teacher. When the student and teacher evaluate how close the student's letters match the model's, there appears to be no objective measure available for this comparison.
It is proposed that incorporating the use of clear plastic evaluation overlays as a feedback source may produce an effective and efficient way to decrease the number errors in student's letter formation. The development of an objective procedure for measuring improvement in formation of individual letters should prelude letter formation training procedures. The present study's first priority was to develop a set of plastic overlays to measure the formation of Zaner-Bloser lower case manuscript letters. Ten lower case manuscript letters: q, j, e, b, v, x, h, m, u, f were chosen as the experimental model letters. Training in letter formation was in three phases. During any phase the subject was given a copy of the ten model letters, a sheet of lined paper, a pencil, and was asked to copy the ten letters.

Phase One required the subject to place the plastic overlays over his letters and verbalize how his letters differed from the model letters. Phase Two required the same procedure as Phase One with the addition the student immediately wrote the letter a second time if it did not meet criterion. Phase Three added to Phase Two a possible reinforcing consequence for each subject's letter meeting criterion.

Questions the study addressed are:

1. Will the measurement procedures, using the plastic evaluation overlays, be sensitive enough to show
changes in manuscript letter formation?

2. Will student recognition of error as indicated by the measurement procedure and verbalization of the error increase or decrease the occurrence of the error?

3. If student verbalization of manuscript errors generates no change in accuracy of the letter formation, will immediately copying the letter a second time increase or decrease accuracy?

4. If student verbalization of manuscript errors and copying the letter a second time generates no change in accuracy of letter formation, will potential reinforcing consequences increase or decrease manuscript accuracy?

5. If a change occurs in one stroke used to construct a letter, will this change generalize to other non-trained letters using the same stroke?

6. If a change occurs in one letter of a letter-family, will that change generalize to other non-trained letters of that family?

7. If a change occurs in one letter of a letter-family, will that change generalize to non-trained letters being used in writing a sentence?
Operational Definition of Terms

For the purpose of this study the following operational definitions were used:

Closed Curve: the union of a set of segments and/or curves so that the end points of one is connected to the beginning of the next and so on until the end point of the last is joined to the beginning point of the first.

Deviation Range: a certain maximum measurement distance that letter strokes can vary from their position in the model letter.

Handwriting: a style or way of forming letters and words by hand with pen, pencil, etc., when writing.

Inner City School: a school situated in a part of the city constructed over fifty years ago with a school population derived from parents of lower socio-economic background.

Model Letters: any or all of the set of twenty-six lower case grade one letters from the Zaner-Bloser Creation Growth Manuscript Alphabet (1974).

Parameter: a constant mathematical measured distance between two segments, arcs, or circles.
Probe Sheet: a sheet of selected letters given to the subject at selected times without any form of directions other than to copy these letters.

Rural School: a school situated at least twenty miles from any major city with the majority of the school population bussed to school from a farming area.

Significant: the number of correct responses has increased until they are approaching the number necessary to function at criterion level.

Stroke: as defined by the Zaner-Bloser Creative Growth Alphabet (1974), a continuous set of points as demonstrated by the mark formed when a writing instrument touches the paper, is moved across the paper, and withdrawn.

Trained Letters: the set of ten letters, f, h, m, u, j, q, v, x, b, e, in all or in part used as models during each condition of the experiment.

Transparent Overlays: a sheet of clear acetate with certain closed curves imprinted on it through the use of the Diagonal Process.

Untrained Letters: the set of sixteen lower case letters, a, c, d, g, i, k, l, n, o, p, r, s, t, w, y, z.
Urban School: a school situated in a residential area less than twenty years old within fifteen miles of the city's center.
CHAPTER II

REVIEW OF THE LITERATURE

Introduction

Starch (1913) considered "handwriting as a mode of conventional expression created by man and necessitating an apprenticeship of the hand in the same way as sewing and knitting with needles." Freeman (1954) stated "the ultimate purpose of handwriting is to express and communicate meaning." Webster's New World Dictionary (1971, p. 340) defines handwriting as "a style or way of forming letters and words by hand with pen, pencil, etc. when writing." The purpose of handwriting then is to communicate with others. Symbols we use to convey our message must be easily read and interpreted by the person or persons for whom it was written. This paper examines how both manuscript and cursive forms of handwriting have been measured for legibility. Many evaluative devices, mostly subjective, have been developed to measure both cursive and manuscript handwriting as a whole, and the set of individual characteristics comprising cursive and manuscript handwriting. However, determining the degree of legibility samples of cursive and manuscript handwriting exhibited has eluded educators. This review will
examine the measurement of cursive and manuscript handwriting legibility with major emphasis on certain selected characteristics and how they effect legibility.

Characteristics of Cursive and Manuscript Handwriting

Handwriting is a common skill taught to all children receiving a formal education. To be more effective in teaching this skill, some educators have attempted to discover what combination of characteristics exemplify good handwriting. Freeman (1918) reported that laboratory studies showed the motor movement used in writing is a complex and delicate coordination. The thumb and the first two fingers exert continual varying degrees of pressure upon the penholder as the different strokes of the letter are produced. Movement of the fingers and hand are combined in a complex manner. Some component elements of the movement are more prominent in some individuals, but each element has its characteristic part to play. Formation of the letter is attended by a series of fine pressure changes of the pen upon the paper. As the handwriter acquires skill, he improves particular letter formation with the characteristic pressure changes.

The speed of movement of the pen is also continually changing. Changes of speed are intimately adapted to the letters which are being written. A straight stroke increases
in speed toward the middle and usually decreases toward the end. When the direction of the stroke is changed, speed is reduced. Long strokes are made with greater speed than short strokes, with the result that strokes widely differing in length may be made in the same or nearly the same length of time. The speed of a writing stroke is also affected by the complexity of the stroke itself and by the complexity of the strokes which follow it. The movement of a writing instrument forming connected cursive letters can be called a speed curve when these movements are measured over time. A pause constitutes a break in the movement which may be regarded as marking the division point between successive units of the movement. The speed curve furnishes a convenient means of studying the way in which the whole writing movement is divided into sections by pauses or retardation in speed. The sections into which the writing movement is divided by pauses may be regarded as the units of the writing movement.

There appear to be certain general characteristics of a well coordinated movement which may be lacking in a poorly coordinated movement. The elements of good writing coordination include the following items:

1. The organization of the movement into units.
2. The arrangement of the division points between units to correspond to the forms to be produced.
3. The execution of each unit in a well coordinated manner:
a. With the greatest speed in general in the middle of the stroke.
b. With a correspondence between the speed of the movement and the length and simplicity of the stroke.

4. Consistency or the production of the successive strokes with similar speed changes (Freeman, 1918, p. 95).

When comparing the coordinated movements of good and bad handwriters there is a sharp contrast in the type of speed organization. The poor handwriters do not organize the movement into speed units and the units of movement do not conform to the units necessary to form letters.

Freeman's (1918) study of students' arm, hand, and finger movements of both good and poor handwriters has uncovered some differing characteristics. In contrast to good handwriters, poor handwriters have difficulty in the sideward movement across the page. This difficulty seems to occur because they rest their hand on the side. Another consequence of this side position is that the hand frequently gets into a cramped position at the end of a word or at the end of a line. No clear contrast with the movement of the good handwriter appeared in respect to the relative amount of finger or arm movement in forming letters.

Legibility of Handwriting

Starch (1913) considered that "handwriting is not an end in itself but is purely an agency of communication and
therefore be measured or evaluated in terms of utility."
Most educators agreed with Starch and proceeded to determine what characteristics determine utility. An analysis of cursive handwriting shows that its three chief elements are legibility, producibility and form. The first two are utilitarian qualities, the third is aesthetic. Perhaps the most important element is a combination of characteristics referred to as legibility. Handwriting is for the purpose of reading. Handwriting which can be read most easily may be adjudged the best.

Freeman (1954, p. 28) stated that "legibility depends in the first instance on the formation of the individual letters" and secondly after letter formation "the most important factor is probably spacing." Thorndike (1910) considered legibility as a word with many possible meanings. In one meaning, legibility of handwriting may be measured by the distance at which it can be read with a given accuracy and rate. In another, legibility is measured by the rate at which a handwriting sample can be read with a given accuracy. In still another, it can be measured by the rate at which a handwriting sample can be read at a given distance and with a given accuracy. Yet still another measure of legibility could be by the degree of fatigue or of discomfort resulting from reading a handwriting sample. Lehman (1973) stated:
As a quality of handwriting, legibility denotes the formal adequacy of the letters to communicate, and results from high formal correspondence between the handwriting and a model that the reader is prepared to recognize. Legibility is a natural consequence of handwriting produced with distinctive, simple letter models in proportioned form.

Measurement of Legibility

When someone states that a student's handwriting is legible it depends upon the person's ability to perceive what he considers legible. For someone to compare two students' handwriting samples, stating which is most legible and to get another person to agree with this evaluation is often difficult.

The first attempts at measuring cursive and manuscript handwriting were through the use of scales.

1. Use of Scales

In March, 1910, E. L. Thorndike first reported the development of a cursive handwriting scale for children and adult women. This contribution marked the beginning of the development of handwriting scales to measure and compare legibility of handwriting samples. Thorndike's scale measured beauty and character as well as legibility. His scale was constructed from one thousand samples of cursive handwriting, furnished by fifth to eighth grade pupils.
These samples were arranged in the order of merit by forty or more judges. This resulted in a scale of graded specimens ranging in quality from zero to eighteen, zero being absolutely illegible and eighteen being a perfect copper plate sample.¹ The differences between successive steps of the scale are equal in amount. Thus quality five was better than quality four as quality twelve was better than quality eleven. A given specimen of cursive handwriting was measured by putting it along side the scale and determining to what quality it was nearest.

The scale itself had a high degree of accuracy and was constructed on the basis of general merit or quality. The scale in reality measured primarily the form or aesthetic appearance of cursive handwriting.

Two years later Ayres (1912) announced a new scale for measuring cursive handwriting samples. Ayres' scale was designed to only measure legibility. It was constructed by measuring the reading speed of 1578 samples of children's cursive handwriting. The occurrence of words in these samples were randomized from their natural context. These specimens were then read by ten different persons and an average reading time was computed for each sample. From the entire group, typical specimens were selected which

¹A sample printed from a flat piece of copper etched or engraved to form perfect letters.
represent eight equally distant degrees of legibility. The scale consisted of three samples with letters having zero, medium, and extreme slant.

Three years later Freeman (1915) introduced the third scale. He constructed his scales to measure excellence in cursive handwriting. He considered general excellence as a sum of the specific evaluation of five factors: (1) uniformity of slant; (2) uniformity of alignment; (3) quality of line; (4) formation of letters; and (5) spacing between letters and words. The formation of letters was given a double weight when rating the samples for his scale. His scales consisted of five separate charts with three grades of excellence for each.

During the six years following Freeman's scale, several new scales were introduced: Gray (1915), Manuel (1915), Kelly (1915), Johnson and Stone (1916), Starch (1919), Gilchrist (1920), and Courtis (1921). These scales were not totally new but duplications or extensions of the Thorndike, Ayres and Freeman cursive handwriting scales. West (1926) developed a new scale which measured quality and speed of cursive writing with one set of seven levels of quality for each grade two through eight. These seven levels ranged from samples of poorest cursive handwriting with slowest speed to best cursive handwriting with fastest speed.

No new scales for measuring legibility were developed until West (1957) and Freeman (1959) revised their older
scales. West's new revision included seven scales, one for each Grade one through seven, with seven levels of quality for each scale. The quality varied from poorest sample with slowest speed to best sample with fastest speed. He considered quality and speed when constructing his scales. Freeman in his revised scale used a judgment of general excellence and did not consider the evaluation of specific factors as he had done in 1915. The new scale had five levels of quality for each grade.

Each of the scales for measuring cursive and manuscript handwriting legibility through 1959 compared a student's sample to one that most closely resembled it from all the samples in that particular scale. The problem was that legibility in handwriting was relative judgment, not an absolute one. Most judges could decide whether Sample A was better, the same, or worse than Sample B but not the extent of the difference. Harris, Herrick and Rarick (1961) examined all major cursive and manuscript handwriting scales designed to measure legibility from Thorndike's first (1910) to Freeman's last (1959) and concluded that:

(a) There is no real zero or perfect sample in a scale derived from these judgments.

(b) The number of quality levels possible in a scale is determined by the number of differences in quality reliably maintained by an adequate number of judges.

(c) Scales produced on this basis are ordinal and not cardinal in character. Level 1 is better or worse than Level 3; it is not however, three times worse or better.
(d) Scales produced on this basis are meaningful only within the population of samples which are used to produce the scale levels. A sixth grade scale produced on one sixth grade population of writing samples need not be similar either in range or level of quality with another sixth grade scale even though both distinguish six levels of handwriting quality (Harris et al., 1961, p. 24).

The Committee for Research in Basic Skills, Harris, Herrick, and Rarick, University of Wisconsin, decided in 1960 to develop a new handwriting scale to measure the legibility of right-handed students when writing in cursive. It was decided the new scale would have the following characteristics:

1. Define a comprehensive continuum of legibility so that:
   (a) The writing of any child in our experimental group would fall within its limits.
   (b) Any child would be able to find a writing sample in the scale "just like his" in general quality, size and slant.
   (c) The distance between any two points on this continuum could be determined.

2. Define slant and size characteristics of handwriting so that these characteristics for every sample would be known and comparable.

3. Apply a scaling procedure which would allow the development of a cardinal scale of legibility.

4. Develop a procedure of classification and use which would make the application of the scale to our research problems efficient (Harris et al., 1961; p. 31).

This new scale constructed by the Committee for Research in Basic Skills (1961) was the first to use these three procedures to give a numerical value to the legibility of a particular cursive handwriting sample. There were 677
samples collected of sixth grade students writing in cursive: "The quick brown fox jumps over the lazy dog." These handwriting samples were given to trained judges in packets of 100 samples with the directions to mark on a separate horizontal scale of 24 squares the rating they would give to each sample. They were to consider the left end of the scale as their evaluation of a sample of cursive handwriting "so poor that they would rarely run into such a sample." They were in turn told to consider the right end of the scale as representing a sample of cursive handwriting "as fine as any you have or ever will see." Each sample was rated by 44 to 64 judges and a mean scale placement was established for each.

Each sample was then measured by a set of squares from one millimeter on each side through ten millimeters on each side. Each letter was evaluated by recording the smallest square in millimeters in which it would fit. The product of the number of letters fitting within one millimeter squares and one was added to the product of the number of letters within the two millimeter square multiplied by two and so on. The sum of these products was divided by the total letters in the sample and a mean size in millimeters was recorded for this sample.

Finally, all loop letters (t h q k b f j p t h l z y d g) of each sample were measured for slant. The measuring instrument was placed over the loop letters individually.
and recorded in this procedure. A rating of 1 was a slant up to 20° left of perpendicular; 2 was 0° slant or perpendicular; 3 was a slant up to 20° to the right of the perpendicular; 4 a slant from 20° to 40° right of the perpendicular. A numerical value was given to each sample by finding the sum of the products of each slant value times the number of letters with that slant. All data were transferred to IBM cards and scaling was completed on an IBM 650 programmed for Attneane's Method of Graded Dichotomies for the Scaling of Judgments. The final product of this group from Wisconsin was a master set of 600 scaled items for Grades four, five, and six which they stated provides:

(a) A definition of a comprehensive continuum of writing legibility for this age school children.

(b) A cardinal scale of legibility which will permit computation of known differences in legibility between two samples of writing.

(c) A means for identifying scale samples of similar style through a procedure for measuring slant and size characteristics of writing.

(d) A research tool which has greater power for dealing with legibility in handwriting than any scale that now exists (Harris et al., 1961, p. 50).

Almost all work in the area of scale development was designed to measure cursive writing. Manuscript handwriting was introduced during the second decade of this century and few scales had been developed to measure it. With a large percentage of schools teaching manuscript handwriting in
the first grade, a measuring instrument was needed.

Bezzi (1962) stated: "The shortcomings of the scales which propose to measure manuscript make it evident that a new scale is needed." He proposed a new scale be developed with the following requirements:

1. The new scale should consist of a series of scales, one for each grade, that measures at least five degrees of quality at each grade level.
2. The new scale should be a measure of rate as well as quality.
3. The general criteria for the selection of specimens for the new scale should be legibility: color, size of letters, slant, letter spacing, beginning and ending strokes, word spacing, alignment, and letter forms.
4. Specimens for the new scale should be selected from a large sampling and the samples should be representative of children's handwriting throughout the United States.
5. Specimens for the scale should be selected through a recognized statistical technique that allows itself to render validity and reliability to the method of standardization.
6. The results of the measurement should lend itself to the present day methods of reporting pupil progress.
7. The new scale should show overlap in quality among the grades.
8. The new scale should be cross-validated with a cursive scale (Bezzi, 1962)

Seven thousand two hundred twelve samples of manuscript handwriting were collected from 130 schools throughout the United States. From these samples, fifty samples from each grade were selected by a systematic plan. A set of five scales for each grade one, two, and three were developed from a mean rank assigned by teachers of each grade to the fifty samples.
Three tentative quality rate-norm tables were developed, one for each grade. They were based upon teachers' opinions as to what elements constitute the five quality levels of manuscript writing. The quality norms were expressed in raw scores, percentiles and T-scores.

The overlap in quality of writing by the grades was very striking, but natural. There was a four-step overlap in the quality of writing between Grades 1 and 2, and 2 and 3. The proposed manuscript scales were cross-validated with a cursive scale. Eighty-four percent of the teachers in Grades 1, 2, and 3 noticed that the manuscript writing was one quality level better than the cursive writing (Bezzi, 1962).

Reliability of Handwriting Scales

Prentner (1914) was the first to measure the reliability of two handwriting scales when both were used to measure the cursive handwriting samples. He stated a scale "must give fairly accurate results with any intelligent individual, and that handwriting scale will be the most reliable that shows the least deviation in the judgments of many different individuals."

He conducted a study of reliability of the Ayres and Thorndike scales when both were used to measure cursive handwriting of 24 selected samples from an ordinary grade school. Thirty-three observers graded the 24 cursive handwriting samples by both the Ayres and Thorndike scales. Observers included students in an educational psychology course, classroom teachers, and Prentner. The samples were
representatives of good, medium, and bad cursive handwriting with every step on the Ayres scale from 20 to 80 and the Thorndike scale from 5 to 14 being represented with the exception of 12 on the Thorndike scale. Each sample was ranked from lowest to highest using the Ayres and then the Thorndike scale. The ranking of the samples using the two different scales was almost identical. The correlation coefficient of the ranks was .98. If the observers had a partiality for one scale, it was Ayres due to the convenience of the scale and its use of percentage in step differentials. The Thorndike scale showed a slightly higher degree of reliability and uniformity in results. This study demonstrated that observers could use both Ayres' and Thorndike's scale for measuring cursive handwriting and get similar results. This study did not answer the question: Can the Ayres' and/or Thorndike scales reliably measure cursive handwriting? It did show that the use of both scales give similar results.

Feldt (1962) used Freeman's 1959 handwriting evaluation scales grade one and two which accompany "Guiding Growth in Handwriting" to evaluate 30 first grade and 29 second grade student's manuscript handwriting samples. Three judges were trained through three practice sessions to evaluate overall quality and to develop consistent standards regarding the emphasis to be placed on the individual aspects of slant,
spacing, clarity of letter, form, and letter size. The three judges evaluated each student's manuscript handwriting sample independently and rated them on a special form.

The statistical method used to analyze the data was that known as the components of variance technique. This methodology yields not only reliability coefficients but also independent estimates of the error variance arising from various sources. Five categories of measurement error were hypothesized for ratings of handwriting quality: (a) momentary variations in the quality of pupil writing, as reflected within the sample taken in a single testing session; (b) undependability or inconsistency arising from the judging process; (c) day-to-day fluctuations in the quality of the writing performance; (d) systematic variation among judges in the emphasis given various aspects of quality; and (e) systematic variation other than error of measurement (Feldt, 1962).

An analysis of variance of the four error components reveals:

that judge-to-judge differences in overall harshness or leniency is by far the most serious. This is true in both grades studied. The error variance arising from the measuring procedure itself is not extremely large considering the magnitude of true subject differences. Day-to-day fluctuations by individual subjects represent about as potent a source of error as the within-day, within-judge component. The judge-to-judge inconsistencies, on the other hand, were rather small. Their size suggests that judges were weighting the several aspects of the writing in much the same way in arriving at their judgments. The error components for Grades 1 and 2 were fairly consistent in size. It appears that judges encountered no greater difficulty in applying one grade scale than the other (Feldt, 1962).
This study by Feldt demonstrated how three people using the same scales and evaluating the same manuscript handwriting samples differ greatly in their evaluations.

Anderson (1965) questioned reliability of handwriting scales when he stated:

Handwriting scales are made of representative specimens of children's handwriting in a series of scales that measure a certain number of degrees of quality (what the scale proposes to measure) at each grade level. These specimens are usually selected by people who have been given training in understanding the basic elements of handwriting; however, evidence shows that at best it is difficult to eliminate subjectivity in the choosing of handwriting specimens.

Askor, Otto and Askor (1971) examined existing scales and stated "scales developed for research purposes must permit fine discrimination among samples: "thus, many quality levels must be provided. When such scales are brought into the classroom, however, ratings by individuals tend to have low reliability."

Feldt (1962) reported that reliability of existing manuscript handwriting scales can be raised by training teachers how to discriminate more closely between manuscript handwriting samples. To date no research has shown that the use of a handwriting scale to be more than a subjective evaluation when measuring handwriting samples.
Individual Factors Affecting the Legibility of Handwriting

Quant (1946) and Craig (1966) each conducted research to measure selected factors of legibility. These two studies used as a criteria the readability of cursive handwriting samples when measuring the effects of certain factors on legibility.

Quant (1946) completed a study to evaluate how various factors modify the legibility of cursive handwriting. In his study he considered legibility as synonymous with readability. Legibility was not considered a unitary characteristic but a composite made up of these factors: (1) letter formation, (2) spacing, (3) alinement, (4) slant, and (5) quality of line. His study considered each factor separately from the other factors. The ease in which cursive handwriting samples could be read as measured by eye movements was used to attempt to answer the following:

1. What effect, if any, do variations in letter formation have on legibility of handwriting?
2. What effect, if any, does variation in spacing have on legibility of handwriting?
3. What effect, if any, does irregularity of alinement have on legibility of handwriting?
4. What effect, if any, does irregularity of slant have on legibility of handwriting?
5. What effect, if any, does the quality of line have on legibility of handwriting? (Quality is used here in a restricted sense, and refers to the weight of the line of writing: Quant, 1946).
The content of the cursive handwriting samples read by each of the 35 subjects remained constant. Eleven selections were prepared for the study with these characteristics:

Selection 1 -- a printed paragraph.
Selection 2 -- a paragraph of normal handwriting in the author's habitual style of writing.
Selection 3 -- handwriting with poor letter formation.
Selection 4 -- normal handwriting, but with reduced spacing between letters and between words.
Selection 5 -- normal handwriting with reduced spacing between letters, but with normal spacing between words.
Selection 6 -- normal handwriting, but with the spacing reduced to 1/4 inch between the lines of writing.
Selection 7 -- normal handwriting, but with increased spacing between letters and between words.
Selection 8 -- a paragraph written with uneven alignment of letters.
Selection 9 -- a paragraph written with irregular slant.
Selection 10 -- a paragraph written with a heavy line.
Selection 11 -- a paragraph written with a light line (Quant, 1946).

Each subject was asked to "read through the paragraph once, silently in their accustomed manner of reading any material that they wished to remember. (Quant, 1946). They were told questions would be asked at the end of each paragraph. These questions were not standardized but given as means to insure the paragraph was read. The study considered legibility synonymous with readability and readability measured by eye movements. Each subject's eye movements were filmed as they read each of the eleven selections."
Three separate, average measures were made from these films: number of words read per fixation, duration of fixations, and number of words read per regression. The average number of words per fixation was computed by the total number of words in the paragraph divided by the number of times the eye moved. The average duration of the fixations was the total time required to read the selection divided by the number of eye movements. The number of words read per regression was defined as the total number of words in the selection divided by the number of times the eyes retraced material already read once. An examination of Table 1 for these three measures when used with each of the eleven characteristics shows marked differences between them. The following data support the theory that good letter formation is the most important characteristic in determining the legibility of cursive handwriting. Conversely, poor letter formation reduced legibility more than any other characteristic under study. Data in Table 1 show that poor letter formation was tied with Selection 7 for having the fewest words read each time the eye was fixed on the reading material. This selection had the second longest average time when measuring how long the eye was fixed on this reading material. Finally the eye was regressing back over the reading material a second time so frequently that poor letter formation allowed the lowest average number of words
TABLE 1

Average Number of Words Read Per Fixation, Average Duration of Fixations, and Average Number of Words Read Per Regression for Selections 1 to 11

<table>
<thead>
<tr>
<th>Selection</th>
<th>Average Number of Words Read Per Fixation</th>
<th>Average Duration of Fixation In Second</th>
<th>Average Number of Words Read Per Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed</td>
<td>1.4</td>
<td>6.5</td>
<td>19.4</td>
</tr>
<tr>
<td>Normal Handwriting</td>
<td>1.0</td>
<td>6.0</td>
<td>10.1</td>
</tr>
<tr>
<td>Poor letter formation</td>
<td>0.8</td>
<td>6.4</td>
<td>4.3</td>
</tr>
<tr>
<td>Reduced spacing between letters and words</td>
<td>1.1</td>
<td>6.0</td>
<td>9.2</td>
</tr>
<tr>
<td>Reduced spacing between words</td>
<td>1.0</td>
<td>6.1</td>
<td>9.7</td>
</tr>
<tr>
<td>Reduced space between lines to 1/4&quot;</td>
<td>1.0</td>
<td>5.8</td>
<td>8.0</td>
</tr>
<tr>
<td>Increased spacing between letters and words</td>
<td>0.8</td>
<td>6.2</td>
<td>4.4</td>
</tr>
<tr>
<td>Uneven alinement of letters</td>
<td>1.1</td>
<td>5.9</td>
<td>11.4</td>
</tr>
<tr>
<td>Irregular slant</td>
<td>1.0</td>
<td>5.9</td>
<td>7.8</td>
</tr>
<tr>
<td>Heavy line</td>
<td>1.0</td>
<td>5.8</td>
<td>8.6</td>
</tr>
<tr>
<td>Light line</td>
<td>1.1</td>
<td>6.0</td>
<td>8.8</td>
</tr>
</tbody>
</table>

(Quant, 1946)

to be read per regression. There was no significant difference in legibility as defined by readability when:
(1) reducing the spacing between words and/or letters;
(2) reducing space between lines; (3) maintaining or changing evenness of letter alinement; and (4) changing the weight of the line when forming letters. The regularity of the letter slant was the second most important characteristic to legibility. The average number of words read per
regression did digress when the slant of the letters did not maintain some uniformity. All five factors (letter formation, spacing, alinement, slant, and quality of line) appear to contribute to handwriting legibility but letter formation and slant seem to contribute the most.

Quant (1946) measured the readability of certain cursive handwriting samples by measuring eye movement while Craig (1966) used the criterion of readability by measuring oral reading of cursive handwriting samples.

Craig (1966) extended the study of Ayres (1912) to study the relationship between selected characteristics of cursive handwriting and the ease in which cursive handwriting samples could be read. Craig examined the relationship between the ease of reading sixth grade cursive handwritten papers by peers and:

A. Height of small letters.
B. Uniformity of size of small letters.
C. Uniformity in size of tall and capital letters.
D. Estimated mean degree of slant.
E. Estimated mean variation in slant.
F. Mean distance between words.
G. Estimated mean variation in distance between words.
H. Alignment.
I. The use of pen or pencil.

A total of 240 samples were utilized from a collection which included three specimens of writing from each student in three elementary schools' sixth grade classes in Kokomo, Indiana. These were rated by their teachers as excellent, acceptable, or non-acceptable and the "ease of reading" each sample was developed as a reading score through oral readings by sixth grade students.
All specimens were analyzed for the following selected handwriting factors: slant, spacing, alignment, size of letters, uniformity of writing, and the use of pen or pencil (Craig, 1966).

The relationships were subjected statistically to a Pearson Product-Moment correlation coefficient and the following characteristics were found to significantly alter the ease of reading at the .01 level.

1. the height of the small letters;
2. the uniformity in writing the small letters;
3. the uniformity in writing the tall and capital letters; and
4. alignment.

A comparable relationship was indicated between reading scores and the teacher ratings (Craig, 1966).

Lehman (1973) was the first to objectively measure length, shape, and spacing of letters. His work is different than Quant's or Craig's since his evaluation considers deviations for certain letter characteristics from a norm rather than using readability as the measuring criteria.

Lehman (1973) conducted his study using as his alphabet the roman, upright, and slanted version italic (Appendix A). His study was designed to look at the characteristics of a person's handwriting that made it more or less legible than someone else's. His study, to measure handwriting legibility, evaluated a single set of objective measures. These measures were specific deviations from prescribed forms of the various versions of the roman alphabet and its slanted
version italic. Four criteria were selected to measure the qualities of size, shape, and position created by each letter and word. These four and the rationale for their choice were:

**Letter Angle.** Without consistent, harmonized alignment, the eye is required to decipher individual letters rather than symbol patterns with resulting greater number of eye sets.

**Space Consistency Between Letters.** Generally, handwriting legibility is improved by even appearing spacing, that is as close as practical. Widely spread handwriting requires greater concentration and confuses perception. Combinations of letters should appear to be equally spaced.

**Vertical Length of Letters.** The idea of length is less ambiguous than height. Length encompasses the possible evaluation of ascenders, letter body, and descenders. The criterion itself is an aspect of scale, shape, and vertical position.

**Counter Space Shape.** This criterion is the direct evaluation of form as space definitions that are simultaneously designed white areas, both inside and outside the strokes of the pen. This one measurement indicates the directional thrust and structural magnitude (Lehman, 1973).

These four criteria when measured "can be used to determine the frequency, kind, and extent of form deviation in handwriting. Considered together, they are the formal cause of legibility in our cultural standard, the roman alphabet and the slanted version italic."

Eighty one first grade students were selected from both public and private schools in September, 1972, and handwriting samples were collected from each. A score sheet and
computer were used for recording measurements. Using the standard sample sentence, "The quick brown fox jumps over the lazy dog," a score sheet was developed listing these thirty five letters down the left hand side of the page. The four criteria were written across the top of the page. Each letter of the subject's sample was evaluated four times: once for each criteria. The four measurements of deviation were recorded opposite the appropriate letter. The four criteria were measured in this manner before subjecting the data to a computer.

Letter angles are measured with a transparent overlay (Appendix B). The overlay has a horizontal series of lines appearing across it ranging from a 5-degree left tilt at the left side of the page, through vertical zero at the middle, to a 15-degree right tilt at the right of the page. The second and third criteria, letter spacing and length, are also absolute measurements in millimeters. The fourth criterion, counter space shape, requires a double measurement. The intent is to provide numerical evidence of shape deviation from the model. Measurement of the shape's visual axis displacement indicates changes in the structural size and location. The method is to use a transparent overlay that has the model alphabet sentence on it. After positioning the overlay, the evaluator then imposes a second transparent overlay with a cross that has all four branches divided into millimeters. The axis of the cross is placed on the axis of the model letter with the vertical branch aligned with zero degrees letter angle. The measurement than is a task of counting the millimeters of displacement from the optical axis of the model letter to the optical axis of the sample letter (Appendix C). The count is made in both directions on the branches of the cross (Lehman, 1973).
The results of the objective evaluation of student handwriting samples using the four criteria previously listed showed letter angles deviating the greatest from the models. The least deviation of students' letters from the models were counter space shape. A computer printout of the deviations of the students' letters on all four criterial were made available to the classroom teachers. The teachers used this information to prescribe instruction for each student during handwriting instruction.

Legibility of Letter Forms

Studies by Quant (1946), Craig (1966), and Rondinella (1971) showed that incorrect letter forms contributed more to illegible handwriting than any other single factor. Bell (1969) stated "The most important factor which should be considered first for determining legibility is recognition of letters." Two studies by Newland (1969) and Lewis and Lewis (1969) examined number and types of letter errors in handwriting samples. Lewis and Lewis (1969) measured samples of manuscript writing while Newland (1969) measured samples of cursive writing.

Lewis and Lewis (1969) completed a study to determine the relative difficulty of the 52 letter forms of the manuscript style alphabet, the incidence of various types of errors by letter form and the relationship of errors to
certain personal variables of first grade children. 
Manuscript writing samples of the 52 letter forms were 
collected from 354 first grade students two times: once in 
the fall and again six months later. The samples were 
judged for errors in letter construction and then classified 
into the following eleven types:

1. **Reversal**: the mirror image of a letter form.
2. **Partial omission**: any part of a letter form missing.
3. **Addition**: inclusion of a part not shown on the model.
4. **Incorrect relationship of parts**: any letter form in which a part is not correctly oriented.
5. **Incorrect size of letter form or parts of it**: a letter form (or part) that is too large or too small in relation to the guide lines.
6. **Incorrect placement relative to line**: incorrect orientation to the writing line.
7. **Misshapeness**: distortion of all or part of a letter form.
8. **Rotation**: a letter form rotated more than 15 degrees from an imaginary line drawn vertically through its axis.
9. **Retracing**: any letter form (or part) that has been reconstructed after an initial effort.
10. **Inversion**: any letter form upside down.
11. **Total omission**: any letter form not attempted.

(Lewis and Lewis, 1969, pp. 274, 276).

The number and type of errors were analyzed to answer these questions:

1. What is the order of difficulty of the 52 letter forms of the manuscript style alphabet?
2. What is the incidence of various types of errors in each of these 52 letter forms?
3. When the letter forms are grouped into classes according to structural characteristics, do the classes differ in frequency of error? (Lewis and Lewis, 1969, pp. 273, 274).
To determine if the structural characteristics differ in frequency of error, the letters were assigned to classes using the following basis:

I. Those constructed of only vertical or horizontal and vertical lines.

II. Those containing only slanted lines or slanted lines combined with horizontal or vertical lines.

III. Those containing circles or parts of circles that do not merge with vertical lines.

IV. Those in which curves and vertical lines merge.

When the data was evaluated, it showed that: (1) The letters q, g, p, y, and j were written with the greatest number of errors while i, o, L, O, and H were written with the least number of errors; (2) There were 7,285 errors classified because of incorrect size of letter form either in total or parts. The next most frequent error was the incorrect relationship of some part of the letter. These two letter characteristics accounted for 10,413 errors out of a total of 18,010.

There was a significant difference at the .001 level between the number of errors of the four structural classes when randomly selecting nine letter forms from each.
<table>
<thead>
<tr>
<th>Class</th>
<th>Letter Form</th>
<th>Total Number of Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>E F H I L T i l t</td>
<td>1,930</td>
</tr>
<tr>
<td>II</td>
<td>N V W X k w x y</td>
<td>2,946</td>
</tr>
<tr>
<td>III</td>
<td>R S C B D a b o s</td>
<td>2,965</td>
</tr>
<tr>
<td>IV</td>
<td>J V f h j m n r u</td>
<td>3,639</td>
</tr>
</tbody>
</table>

The letters that had the greatest number of errors in form were those that required the use of the descender space and those with the fewest errors were those constructed with vertical and horizontal segments.

Newland (1969) analyzed 1,344,905 letters written by 2,381 different individuals ranging from first graders to adults serving in a variety of occupations. He directed two different people to record each time they encountered something in their reading of the cursive handwriting samples which made them stop and look a second time. Variables such as misspellings, re-writings, crowdings, or erasures were recorded, but not as specific illegibilities. The illegibilities recorded were incorrect letter formations such as "'a like o', 'g like cj', or 'e closed'" (p. 241). A total of 499 different forms of illegibilities were recorded. The letters c, e, j, q, x, a, and z had less than ten different forms in which they appeared illegible. The letters i, n, l, m, and v were found written illegibly between 10 and 19 different ways. The letters p, t and f were found in more than 30 different illegible forms and the rest of the alphabet from 20-29 different illegible forms.
The illegibilities were recorded using percentages by four categories: elementary, high school, adult, and total.

<table>
<thead>
<tr>
<th>Type</th>
<th>Percentages Contributed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>El.</td>
</tr>
<tr>
<td>1. Failure to close letters (a,b,f,g,j,k,o,p,q,s,y,z)</td>
<td>24</td>
</tr>
<tr>
<td>2. Top loops closed (&quot;l like t,&quot; &quot;e like i&quot;)</td>
<td>13</td>
</tr>
<tr>
<td>3. Looping non-looped strokes (&quot;i like e&quot;)</td>
<td>12</td>
</tr>
<tr>
<td>4. Using straight up-strokes (&quot;n like u,&quot; &quot;c like i,&quot; &quot;h like li&quot;)</td>
<td>11</td>
</tr>
<tr>
<td>5. End stroke difficulty (not brought up, not brought down, not left horizontal)</td>
<td>11</td>
</tr>
<tr>
<td>6. Difficulty crossing &quot;t&quot;</td>
<td>5</td>
</tr>
<tr>
<td>7. Difficulty dotting &quot;i&quot;</td>
<td>3</td>
</tr>
<tr>
<td>8. Top short (b,d,f,h,k,l,t)</td>
<td>6</td>
</tr>
<tr>
<td>9. Letters too small</td>
<td>4</td>
</tr>
<tr>
<td>10. Closing c,h,r,u,v,w,y</td>
<td>4</td>
</tr>
<tr>
<td>11. Part of letter omitted</td>
<td>4</td>
</tr>
<tr>
<td>12. Up-stroke too long</td>
<td>2</td>
</tr>
<tr>
<td>13. Letters too large</td>
<td>2</td>
</tr>
<tr>
<td>14. Beginning stroke off line</td>
<td>2</td>
</tr>
<tr>
<td>15. Bottom short (f,g,j,q,y,z)</td>
<td>2</td>
</tr>
<tr>
<td>16. Using rounded up-strokes instead of straight ones (&quot;i like e,&quot; &quot;u like ce&quot;)</td>
<td>1</td>
</tr>
<tr>
<td>17. Down-loop turned incorrectly</td>
<td>1</td>
</tr>
<tr>
<td>18. Excessive flourishes</td>
<td>1</td>
</tr>
<tr>
<td>19. Part added to letter</td>
<td>1</td>
</tr>
<tr>
<td>20. Down-stroke too long</td>
<td>1</td>
</tr>
<tr>
<td>21. Up-loop turned incorrectly</td>
<td>1</td>
</tr>
<tr>
<td>22. Down-loop closed</td>
<td>1</td>
</tr>
<tr>
<td>23. Printing</td>
<td>1</td>
</tr>
<tr>
<td>24. Palmer &quot;r&quot;</td>
<td>2</td>
</tr>
<tr>
<td>25. Unrecognizably recorded</td>
<td>10</td>
</tr>
</tbody>
</table>

*These dashes represent frequencies which accounted for less than one-half per cent of the total (Newland, p. 246).
The illegibilities of the letters a, e, r, and t contributed 45% of the elementary, 46% of the high school, and to 47% of the adult letter errors. Only 16 per cent of the 499 different forms of illegibilities were found common in all three groups while 64 per cent were found peculiar to only one group. Four types of errors in letter formation caused over one-half of the letter illegibilities. These were: "(a) The failure to close letters, (b) closing loop strokes, (c) looping non-looped strokes, and (d) using straight-up strokes rather than rounded strokes (Newland, p. 246)." The gross frequency with which illegibilities appeared tended to increase with age. The elementary students were the least illegible in their writing with the high school students 136 per cent more illegible. The adults were 350 per cent more illegible than the elementary students.

How Perception Effects Legibility

Gordon and Mock (1960) wrote "good and bad writers are to be found in all age groups and at all levels of intelligence, for there is little relation between good writing and intellectual ability." Gordon and Mock's statement was not accepted by all educators, yet little research was available to repute their statement at that time.

Goforth and Hunnicutt (1970) stated that good handwriting seems to be primarily a mental process--an internalization
of legible, external models. A person becomes a legible writer by forming within himself a permanent comprehensive set of images of himself writing legibly.

Harris, Herrick and Rarick (1961, p. 51) posed this question: "Do children of different levels of intelligence perceive given handwriting task in the same way?" The experimental sample was divided into three categories: high intelligence (gifted); average intelligence (normal); and low intelligence (dull). Each of the categories included ten subjects - five boys and five girls. The ten subjects in the group labeled high intelligence had I.Q.'s that ranged from 125-143 with an average of 129 and an average chronological age of nine years and one month. Those in the group with average intelligence I.Q.'s ranged from 94-105 with an average of 100 and an average chronological age of nine years and two months. The group with low intelligence I.Q.'s ranged from 64-78 with an average of 71. That group's average chronological age was nine years and one month. The experiment was designed to look at these two characteristics of normative perceptual and aspirational perception when recognizing, ordering, and producing over intervals of time samples of handwriting "like mine" and my usual model of handwriting. Normative perception was defined in this case as recognition of "how my handwriting actually is." It was assumed that learning a skill like handwriting developed a normative set, a total set of
symbols, connected and non-connected, which formed the precept "this is my handwriting." The aspirational perceptions were defined in this case as handwriting "which I would want my handwriting to look like." This was an aspirational set of symbols connected and non-connected forming what the person would like his handwriting to look like.

A sample of each subject's cursive or manuscript handwriting was collected when he was asked to write his name, address, and school. The size and characteristics of the subject's handwriting was used by the experimenter to select, from the Wisconsin set of 600 handwriting scaled items, a set of five handwriting scale items with similar characteristics. The selected sets of scale items represented five levels of legibility and were shuffled and given to the subject in random order. The subject was allowed to determine his own scale ordering. He ordered the five scales from best to poorest. The subject was asked to choose the one sample which looked like his handwriting (normative) and the one he wanted his to look like (aspired to) without any samples of his handwriting to which he could refer. During the writing phase, the subject who wrote in cursive was asked to write: "The quick brown fox jumps over the lazy dog." The subject who wrote in manuscript was asked to write: "The fox jumps." He wrote this sentence four times: first at his normal speed; once at his fastest speed; then
at his best handwriting; and last at his normal speed a second time. The subject ordered his four handwriting samples from easiest to hardest to read. He then selected his best and poorest handwriting sample.

After writing and ordering his own handwriting samples, the subject judged the five randomized scale samples again selecting for a second time the one scale sample that looked most like how he viewed his own handwriting. The subject also selected the one representing how he would like his handwriting to look. The subject then judged his own handwriting samples against the scales using the order he placed the scale samples the first time. He then chose the one scale sample which looked like his poorest handwriting sample. The experimenter shuffled the scale items a second time and had the subject rearrange the five scale items again from easiest to read on top to hardest to read on bottom. In case the subject incorrectly ordered the five scale samples the first time, the experimenter placed these samples in the correct order and had the subject choose the samples which he thought matched his best and poorest handwriting.

Harris, Herrick, and Rarick (1961) summarized the results of this experiment by dividing their general findings into these three major areas.

The first was general perception of differences in quality of handwriting. They found the bright and average
group to differ significantly from the dull group in ordering handwriting sample. The bright and average were superior in their ordering by quality, in ordering in accordance with adults and improvement in accuracy in ordering their second handwriting sample. The second area was the consistency of the two successive orderings of "like mine" (normative) and desired (aspired to).

All three groups ordered their handwriting samples more consistently than by chance in these three ways. They demonstrated a difference in their placements of both their two "like mine" and two aspired to handwriting samples. They placed their aspired to handwriting sample equal to or higher in quality than the handwriting sample "like mine."

The bright and average group did better in the relative placement of the best and poorest samples in two orderings than the dull group.

In the third area, they found the following relationships between perception and legibility:

A. Performance on legibility was differentiated by factors of sex (girls consistently better writers); by intelligence groups (bright and average groups better writers than the dull); and by the conditions of the writing task (best samples were better than the fast).

B. The bright and average groups were consistently more accurate than the dull group in ordering their own handwriting samples according to legibility.

C. The bright and average groups and girls differed from the dull group and boys in the range of legibility produced in their writing samples.
D. There was a high correspondence between the high and average intelligence groups both on legibility and perception scores and evidenced a pattern which was different from the dull group.

E. There was no clear-cut difference between the three groups on perception scores but there was a significant one on legibility.

F. High perception scores do not predict either intelligence group or legibility of writing. (Harris et al., 1961, p. 94-95).

Measures of Selected Physiological and Motor Correlates of Handwriting Legibility

The act of constructing letters and words requires fine motor control. The production of legible handwriting is exacting and requires very precise movements of the arm, wrist, and hand and fingers. A study of how different parts of the body function while collecting samples of handwriting under varying conditions could give insight into how the body must function for the student's handwriting to be legible. Harris, Herrick and Rarick (1961) compared some of the physiological and motor correlates of cursive or manuscript handwriting legibility of bright, average, and dull subjects.

Their experiment looked into a comprehensive study of (1) muscular response in the writing arm; (2) body tension, and (3) autonomic responses indicating the affective reactions of the student. While examining each of the reactions individually is important, it is also important to look at
their relationship while the subject is producing legible handwriting. The subjects for this experiment were the same as in their first experiment except for three in the dull group were replaced with three of equivalent I.Q. Two years had elapsed since experiment one and all the subjects have gained two years chronologically. A fifth writing condition was added to the four in experiment one with these verbal instructions: "Relax the muscles of the shoulder and neck and relax your arm while writing in your normal way." Harris, Herrick, and Rarick wanted to examine certain physiological and motor aspects of behavior in relationship to handwriting legibility. This experiment was designed to observe changes (1) in the affective state as measured by the Galvanic Skill Response; (2) muscular tensions as measured by muscle action-potentials; (3) fine motor control as measured by point pressure variability while systematically varying handwriting conditions and to consider the relationship between legibility of handwriting produced under the varying conditions. When the subjects were writing their cursive or manuscript samples under each of the five writing conditions, the subject's galvanic skin response (GSR), electromyogram (EMG), and force variation ratio (FVR) were measured and recorded.

The GSR was used to measure the affective state. The galvanic skin response (GSR) is one of the most frequently used index of autonomic activity associated with affective
states such as apprehension, arousal, and alertness. It is most generally held that the electrical changes in the skin are due to alterations in membrane permeability associated with secretion of these sweat glands. The palm of the hands and soles of the feet are very well supplied with sweat glands and these areas were used to monitor the galvanic skin response.

Muscular tension was measured by electromyogram. Relaxed muscles produce minute or no electrical potentials while muscles being contracted produce very small electrical potentials which can be picked up by electrodes placed correctly on a person. The initiation and duration of muscular response can be accurately recorded by the EMG. Some studies of a lesser nature have suggested that specific patterns of electromyogram activity become characteristic of the task: Daniel (1939); Freeman (1931); Henderson (1952); and, Oldreyd and Moskowitz (1952). Electrodes placed on the arm and neck were used to monitor the subject's electromyogram responses.

Fine motor control was measured using the force variation ratio (FVR). Handwriting, being a perceptual motor activity, would presumably require a high degree of fine motor control in the application of force during the handwriting act. A measure of force variation of the point of the writing instrument as it is being used to form letters and words could serve as an index of the degree of such motor control.
The FVR was a ratio in which the cumulative amount of change in point pressure was divided by time in seconds. A special desk top under the subject's paper was designed to measure the force being applied by the point of the writing instrument.

Legibility of each subject's handwriting sample was measured under each of the five conditions and recorded. GSR, EMG, FVR and legibility rating was recorded for each subject under all five conditions. An analysis of this data showed for these subjects that boys and girls of average intelligence and bright girls performed similarly, but boys of high intelligence wrote more slowly on the average. When an analysis of variance was computed on the five conditions over boys and girls in both intelligence groups, only differences in speed of writing during the five conditions showed any significant differences at the .01 level. Comparing writing speeds by sex or intelligence did not meet significance at even the .05 level. As expected, both the bright and average groups produced their most rapid cursive handwriting under "write your fastest" and their slowest under "write your best."

Looking at the degree of fine motor control when using the force variation ratio as the evaluative instrument showed the least amount of control for all groups under the condition of fast writing and the greatest degree of fine motor control under the relaxed handwriting conditions. An
analysis of variance of the FVR showed that fine motor control was significantly affected by handwriting conditions at the one per cent level. In the majority of cases as legibility increased, the force variation decreased and vice versa. A correlation of .85 was found between the FVR for the three writing conditions, fastest, best and second normal which was a strong indication of the relationship between speed and FVR but not sex or intelligence. An analysis of variance clearly showed that average and bright girls were more legible writers than the boys. It was also shown at the .05 level of significance that legibility was a function of the handwriting condition. A fairly high correlation of -.55 was found between legibility and time, indicating that high legibility tended to be accompanied by writing slowly and vice versa.

Combining the average and bright girls in one group, the average and bright boys in another and the slow learners in a third a Freedman Two-Way Analysis of Variance was run on the handwriting sample data of these six characteristics: legibility, galvanic skin response, neck electromyogram, arm electromyogram, speed, and force variation ratio. This test showed that bright and average girls as a group appear to have developed a more harmonious pattern of psychophysiological mechanisms in accord with their legibility than the boys in this ability grouping.
When comparing the legibility of the four best handwriters with the four poorest, the best writers as a group tended to be superior handwriters under all conditions. These same four best handwriters showed superior fine motor control under most conditions. The superior handwriters seem to have developed a more stable set of motor skills in handwriting which helps them to adapt to different conditions without a serious deterioration of their handwriting.

Summary

The first major attempt to measure the legibility of handwriting was by Thorndike (1910). He developed a set of handwriting scales consisting of samples of cursive handwriting from very bad to very good. His work was quickly followed by two new sets of scales, Ayres (1912) and Freeman (1915), each using a different criterion when selecting cursive handwriting samples. There were six additional scales developed during the seven years following Freeman's but they were duplications or extensions of existing scales. The only major scale development through 1956 was by West (1926) and the revisions by West (1957) and Freeman (1959) of earlier scales. Harris, Herrick and Rarick (1961) and Bezzi (1962) developed the first new major scales in thirty-five years.

Prentner (1914) studied the reliability of Thorndike's and Ayres' scales when using both to measure the same cursive
handwriting samples. He found their agreement to be very high: \( r = .98 \). Feldt (1962), Anderson (1965), and Askor, Otto and Askor (1971) found scales developed through 1959 to be unreliable when evaluating handwriting legibility. Research evaluating the reliability of Harris, Herrick and Rarick (1961) and Bezzi (1962) scales was not available. Quant (1946) and Craig (1966) approached the measurement of legibility by measuring the reliability of cursive handwriting. Quant measured eye fixations while Craig used oral reading as his criterion for reliability. Studies by Quant (1946), Craig (1966), and Rondinella (1971) showed that incorrect letter formation contributed more to illegible handwriting than any other single factor. Lewis and Lewis (1969) and Newland (1969) measured the numbers and types of incorrect letter forms. Lewis and Lewis studies manuscript letter forms while Newland examined cursive letters. Newland found that the letters a, e, r, and t contributed forty-six per cent of the illegibilities measured.

Research to date has not discovered an objective measurement for cursive or manuscript handwriting legibility. Whenever an objective way of showing change in letter formation is discovered and combined with teaching correct letter form, a promising way to increase handwriting legibility may be available to teachers.
Teachers instruct students in the formation of manuscript letters either formally or informally. The amount of time spent learning letter formation during the first two years of a student's life can vary greatly within and between schools. Teachers could be assisted in this task if a procedure was developed for training students to form letters closely approximating the models. One difficulty arising with any assignment requiring the students to form letters from a model is the evaluation of their task. Teachers must study the students' newly formed letters and try to evaluate their approximation of the model. A devise that teachers or students could use to adjudge how close the student's letter approximates the model would increase the objectivity of this judgment.

The purpose of this study was fivefold. One was to examine the feasibility and reliability of using a set of evaluative overlays as an objective measurement of manuscript letter formation. The second was to determine if recognition of letter formation errors by using these evaluative
overlays, combined with self verbalization of errors will improve correct letter formation. The third included the procedure of the second with the addition of writing the letter again with verbalization of errors each time the original copy did not meet criterion. The fourth was the addition of a possible reinforcing consequence to the previous procedure. The fifth purpose was to examine if the change in the formation of trained strokes and letters will generalize to change in untrained letters written independently and/or in sentence writing.

Subjects

Six subjects were used in this study: two each from inner city, urban, and rural schools. These subjects were selected by asking a teacher to identify two children in her classroom who were experiencing difficulty in manuscript letter formation. Their ages ranged between five and nine years old. All subjects attended public schools. Four subjects were selected from a first grade and two from a primary Learning Disabilities classroom.

Setting

The study took place in three elementary schools. The subjects worked at tables in corners of their respective classrooms or in tutorial rooms outside of their regular
classroom. When the experimenters arrived at the classrooms, the teachers told the subjects to go to the designated areas for handwriting instruction. Each subject was seen three sessions per week at various times during the school hours. Depending upon the rate of the subject's letter formations, the length of each session was approximately eight to ten minutes. The experimenters sat to the left of each subject while facing the same direction. The work surface was void of objects except for the subject's writing materials, i.e., pencil, writing paper, a half sheet of blank paper, and model letters.

Materials

Model Letters

Five training sheets of ten model lower case letters were developed using the Zaner-Bloser Creative Growth Manuscript Alphabet (1974). The letters were constructed with a line weight of one millimeter. The letters were printed on paper with two sets of four parallel lines. The spaces between the headline and midline, midline and baseline, and descender space below the baseline were seven-sixteenths of an inch each. The same ten letters were on each sheet only the order of the letters varied. These were the models and arrangement of letters:
Model 1: f h m u j q v x b c
Model 2: b e f h m u j q v x
Model 3: v x b e f h m u j q
Model 4: j q v x b e f h m u
Model 5: m u j q v x b e f h (Appendix D)

Two Probe Sheets were developed using the lower case manuscript letters from the Zaner-Bloser Creative Growth Alphabet. One Probe Sheet contained three sets of four parallel lines with the letters a c d g i k l n o p r s t w y z printed on them. (Appendix E) The second Probe Sheet contained the sentence "the brave foxes quickly jump fences." (Appendix F).

Evaluative Overlays

A set of transparent overlays were designed to measure three ranges of deviations of student samples from model letters. (Appendix G illustrates one overlay). The first set measured deviations from zero to one millimeter. The second from zero to two millimeters, and third from zero to three millimeters. The three overlays were constructed to form a closed curve around each letter, one with a parameter of three millimeters, one with two millimeters, and the third with one millimeter. When a letter contained a circle stroke a second closed curve was inscribed within the circle with a parameter of three, two, or one millimeter depending
upon the deviation range. For each of the model letters a fine line was drawn so that the model letter was inside a boundary consisting of a set of segments and/or arcs. When any letter contained a circle stroke a fine circle was inscribed within this stroke. The boundaries were constructed by:

1. Drawing lines parallel and three, two, or one millimeter on both sides of all strokes comprised of vertical, horizontal, or oblique segments.

2. Drawing two circles: one inside and one outside of the model circle stroke. One circle was a perpendicular distance of three, two, or one millimeter from a tangent on the circle stroke to a tangent on the circle inscribed within the circle. The second circle was drawn a perpendicular distance of three, two, or one millimeter from a tangent on the circle stroke to the circle surrounding this stroke.

3. Drawing arcs three, two, or one millimeter on both sides of strokes or parts of strokes constructed of an arc.

4. The ends of all strokes comprised of arcs and/or segments had their boundary lines on both sides joined with a semicircle using a radius of three, two, or one millimeter and the end of the stroke as a center.
5. All strokes constructed of arcs and/or segments which did not intersect at an end point had a perpendicular slash mark drawn across the stroke three, two, or one millimeter from their end points.

Construction of Overlays

The boundaries surrounding the inside and outside of each letter were first constructed as a pencil copy using the following equipment: (1) A steelspring bow divider with two fine metal points was used to transfer the one, two, or three millimeter distance from a standard metric ruler. (2) A Brunery Accutrac Drafting machine was used to draw parallel segments to all strokes that were vertical, horizontal, or oblique segments. (A simple adjustable triangle and T-square could have been used.) (3) A steelspring bow divider was used to inscribe a series of points inside and outside all circles and arcs. (4) An ellipse template was used to join the series of points to form circles and arcs. Pencil copies of the letter boundaries were then traced on vellum tracing paper using india ink. The letters drawn with india ink on vellum paper were used to make transparent overlays using the Diazo Process.
Copy Paper

Three sets of paper were developed with the following spacing between lines: seven-sixteenths of an inch between headline, mainline, and baseline with a descender space of seven-sixteenths inch below baseline. The paper used to copy the model letters consisted of two sets of four parallel lines (Appendix I). The paper for the sixteen untrained letters consisted of three sets of four parallel lines (Appendix J).

Behavior Definition and Recording

For the letter stroke to meet criterion:

1. The total strokes were within the confines of the red line of the overlay.
   Example: yes ( ) no ( )

2. Each stroke that was not a complete circle had to begin and end between the small slash mark and in the red line forming the confines of the letter.
   Example: yes ( ) no ( )

3. All circles in the letters a, b, d, g, o, p, q, and the top of the letter e were closed curves.
   Example: yes ( ) no ( )

4. All strokes intersected each successive stroke at one point except for the dot above the i and j.
   Example: yes ( ) no ( )
5. The letter was complete with all strokes present.
   Example: yes (  )  no (  )

6. The horizontal stroke in the t and f intersected
   the other stroke within the confines of the ellipse
   near the center of the verticle stroke.
   Example: yes (  )  no (  )

Recording strokes meeting these criteria required the
   correct usage of evaluative overlays. The overlays had two
sets of parallel broken lines that were drawn to match up
with the guidelines of a subject's handwriting sample. The
overlay was placed over the subject's sample so the broken
lines were directly over the sample's guidelines. The
overlay was moved to the left or right until each stroke,
or as many strokes as possible, of the sample letter were
within the letter boundary on the overlay.

A recording sheet was developed listing the ten model
letters down the left side of the page (Appendix H ). Each
letter included graphical pictures of the strokes used to
construct the letter beside it with blanks to record the
strokes meeting criteria. The evaluation overlay was used
to measure the first letter of the subject's handwriting
sample using all of the six criteria pertaining to this
letter. Each letter stroke meeting criteria was recorded
with a (✔) beside the graphical representation of the
stroke. The evaluator measured the second letter and
recorded the correct strokes and etc. until all ten sample letter strokes were measured and recorded.

To determine which of the three evaluation overlays were to be selected to measure a subject's handwriting samples, the first four baseline samples were measured with all three overlays: three, two, and one millimeter. The results of the measurements of all three overlays were recorded and for each the total correct strokes were divided by forty-two. These quotients were converted to a percentage representing the number of strokes meeting criterion. The selection of the evaluation overlay used in measuring the subject's handwriting sample was the one that demonstrated an average of approximately fifty per cent of the letter strokes meeting criterion.

Design Conditions

Baseline

Each subject was seated to the right of the experimenter with a set of ten model letters, two sheets of line paper, a pencil, and a half sheet of blank paper. When beginning each experimental session, the experimenter used from one to three minutes of some form of verbal behavior that increased the chance the subject felt at ease. For example: "Your hair looks very nice today." "Boy, it's sure cold outside." The student was then given the copy of ten model letters, a
sheet of lined paper and a pencil with these directions: (1) I am going to cover five of the letters with this piece of blank paper. (2) Please use the paper and pencil I have given you and copy the top five letters for me. When the five letters were copied, the experimenter moved the blank paper to cover up the top five letters and asked the subject to then copy the last five letters. The subject's ten sample letters were picked up and one to three minutes were spent in some form of appropriate verbal interaction between the experimenter and subject. For example: "Did you see Walt Disney's t.v. program last night?" "What did you think of that bear?" Then the subject was asked to make a copy of the ten model letters a second time while the experimenter again sat quietly to the subject's left and only moved the blank paper when the first five letters were copied. When the subject had copied the ten letters twice, the experimenter thanked the child for helping him, picked up the materials and left. The experimenter gave no verbal feedback to the student on the quality of his handwriting samples. The only change in baseline sessions was when the experimenter drew from a folder with his eyes closed one of the five models for the first session, then one from the remaining four models for the second session, etc., until all five models had been used. When all models had been used once, the five models were placed back into the folder and the same procedure was repeated.
again. Baseline continued for five sessions or until the number of correctly formed letter strokes had stabilized (i.e., no increasing or decreasing trend) or began a decreasing trend in behavior. Upon the completion of collecting baseline data two Probes were given to each subject during the last session before intervention one. The same setting and type of verbal interaction was used for each Probe, baseline, and all intervention sessions.

Probes

Probe One, sixteen untrained letters (Appendix E), was given to the subject with a sheet of special paper (Appendix J) with the directions to copy each letter one time. When the subject had copied all sixteen letters the experimenter collected all materials and spent two to three minutes in appropriate verbal interaction with the subject. The experimenter gave the subject a copy of Probe Two, the sentence "the brave foxes quickly jump fences" (Appendix F), and a sheet of special lined paper (Appendix K) and asked him to copy the sentence. When the subject completed copying the sentence the experimenter collected all materials, thanked him for helping and left. The two Probes were given after the last session in baseline, each experimental condition, and when changing the overlays deviation range.
**Intervention One**

First the experimenter selected a target letter for intervention one. The letter that generated the greatest percentage of incorrect strokes during baseline was selected as the first target letter. During intervention one the same procedure was followed as was in baseline with the student copying the ten model letters one time. Following the student's ten letter responses, the experimenter placed the evaluation overlay correctly over this letter and asked the subject to look at his letter and the model letter and verbalize how his letter was different. When the subject had difficulty in pointing out the differences in the strokes, the experimenter assisted the subject. For example: "How does the tail of your q differ from the model q?" After the student verbalized the differences, the experimenter removed the subject's handwriting evaluation overlay and spent two to three minutes in appropriate verbal interaction with the subject. The subject was asked again to copy the model letters and again evaluated the target letter chosen by the experimenter. When the subject had verbalized the differences between his letter and the model, the experimenter thanked him for helping and left. This procedure was repeated for at least five sessions or until a significant increase in the number of correct strokes was exhibited by the subject. When the correct number of strokes for the target letter significantly
increased, a second target letter was selected and added to intervention one. This target letter was the letter that generated the second highest percentage of incorrect strokes during baseline. Each time the newly added letter showed a significant increase in the number of correctly formed strokes a new target letter was selected. If, after ten sessions, intervention one did not produce a significant increase in the number of correct letter strokes in the target letter, the subject made copies of both Probes and moved to intervention two.

**Intervention Two**

The same procedure for intervention one was followed through the subject's completion of copying one set of ten letters and self verbalization of errors of the target letter. When the target letter did not meet the criterion of the plastic overlays, the subject was asked to copy the target letter immediately again and then verbalize a second time how his sample letter differed from the model. Two subjects were required to continue copying the target letter and verbalizing their errors until the target letter met criterion. The same procedure was followed after the subject copied the ten letters a second time. When measurement of the target letter showed a significant increase in the number of correct strokes, then a new target letter was chosen and this procedure was continued until all ten letters
had become target letters. When there was no significant increase in correct letter strokes of the target letter(s), after ten sessions, the subject made copies of both Probes and entered intervention three.

Intervention Three

The same procedure was followed as in intervention two except for the addition of potential reinforcement each time the subject's sample target letters met criteria the first time it was written. When the target letter showed significant increase in the number of correct letter strokes a new target letter was selected and so on until all ten letters had been target letters. When ten sessions in intervention three had been completed and the graphs of correct responses for each target letter showed no significant increase, then the subject copied both Probes and stopped the experiment.

Design

The design employed in this study was a combination of two behavioral designs. The study used a multiple baseline design (Cooper, 1974) both across subjects and across behaviors within subjects. One, two, or all three of the following interventions were used: During intervention one, the subject placed the evaluative overlay over his letters
and while he looked at both the model and his own letter verbalized how they differed. Intervention two included the procedures of intervention one with the addition that each time his letter did not meet criterion he immediately wrote it and verbalized how they differed a second time. Intervention three included all the procedures of intervention two with the addition of a potential reinforcement for each letter that met criterion the first time it was written. The multiple baseline design was applied by including several similar behaviors (individual letters) during each condition. A multiple baseline design using two or more similar behaviors of a given subject was applied in this way. All behaviors were measured for change while one target behavior was subjected to treatment and the rest continued on baseline. When the measurement of the target behavior showed a significant change in rate of response, a second target behavior was added to the treatment while the remaining behaviors remained on baseline. Each time the previous target behavior demonstrated a positively increasing trend in correct responses, a new target behavior was introduced to intervention. Figure 1 illustrates a hypothetical example of the design.
Figure 1. Hypothetical example of a multiple baseline design.
Data Collection and Analysis

Data from baseline and intervention procedures were measured and recorded by the researcher using the evaluation overlays and recording sheet. All handwriting samples were evaluated and correct individual letter strokes were recorded. The only letter strokes measured and graphed were from the first letter copies written during each phase of a session. The strokes from the ten letters were graphed by these families: tall, letters f, h; circle, letters b, e; slant line, letters v, x; tail, letters j, q; and hump, letters m, u. The sum of the correct strokes exhibited each session were divided by the total strokes, forty-two, and graphed as a percentage.

The graphs were studied and examined for significant changes in the number of correct strokes for each letter during each condition of the study. A significant change was defined when the subject's letter form consistently approached the criterion of the evaluation overlay. When observing significant results, the multiple baseline design demonstrated to the experimenter a strong inference of causality. This causality was demonstrated when the target behavior, under intervention, showed a significant increase and the other similar behaviors remaining on baseline did not. When a second, third, and fourth target behavior being emitted at a stable or descending rate was added to
the intervention condition, one at a time and a significant increase in the behavior was demonstrated, then it was highly probable that the intervention tactic was the cause for change.

Interobserver Reliability

Interobserver reliability agreements divided by the sum of agreements and disagreements, multiplied by 100, was obtained in the following manner:

(1) Three evaluators were trained to use the evaluative overlays and record correct responses when measuring subject handwriting samples. These evaluators were given a subject's handwriting sample, an evaluation overlay, a recording sheet and a letter criterion sheet. The researcher explained the letter criterion sheet by demonstrating all type of errors letter strokes exhibited when not meeting the predetermined criterion. Each evaluator was shown the correct procedure when placing the evaluation overlay over a subject's sample and how to shift it horizontally to bring the letter that was being evaluated within the parameters of the overlay. The use of the recording sheet was demonstrated by the researcher evaluating letter strokes of
one subject's sample and recording all correct responses. The three evaluators practiced using the evaluative overlay and recording correct responses until each met an eighty-five per cent agreement with the researcher. Once the responses, recorded by all three evaluators, met an agreement criteria of eighty-five per cent or better actual interobserver reliability counts began.

(2) Reliability checks were taken during each condition of the study. Approximately one out of each four samples evaluated were re-evaluated a second time by one of the trained evaluators.

(3) A measure of interobserver reliability was obtained by having the second evaluator measure and record the number of correct letter strokes on a subject's sample of twenty letters. An average of eighty per cent criteria of interobserver agreement was required throughout the study.
CHAPTER IV

PRESENTATION, ANALYSIS, AND DISCUSSION OF DATA

Effects of treatment were measured by an instrument developed to measure changes in lower case manuscript letter formations. Three instruments were developed to measure maximum deviations of three, two and one millimeter from all twenty-six letter models. Manuscript handwriting samples from six subjects were evaluated using these instruments in conjunction with a set of behavior definitions. During baseline and interventions one, two and three, each subject was asked to copy the letters b, e, f, h, m, u, j, q, v two times. A special recording sheet was used to record the part or all of the forty-two strokes meeting criterion. Each subject was asked to copy two probes at the end of each experimental condition. Probe one consisted of sixteen untrained letters (a, c, d, g, i, k, l, n, o, p, r, s, t, w, y, z) and Probe two, the sentence: "the brave foxes quickly jump fences." Two recording sheets were used to record the part or all of the eighty-nine letter strokes meeting criterion.
Interobserver Agreement

Interobserver agreement of data recording was determined during each of the four experimental conditions and ten of the probe sessions. One of four trained independent observers measured a subject's manuscript handwriting sample and recorded the strokes meeting criterion. The two recording sheets used to measure the subject's writing sample were compared to determine number of responses scored the same (agreements) and number of responses scored differently (disagreements). Inter-observer agreement was then calculated using the formula given below.

\[
\text{Percentage of Agreement} = \frac{\text{No. of agreements}}{\text{No. of agreements} + \text{No. of disagreements}} \times 100
\]

The condition in which inter-observer agreement measured were gathered and the percentage of agreement are given in Table 2.

**TABLE 2**

Condition in Which Inter-Observer Observations Occurred
Range and Average Percentage of Agreement

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of Sessions with Inter-Observer</th>
<th>Highest Agreement</th>
<th>Lowest Agreement</th>
<th>Average Percentage of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>40</td>
<td>98%</td>
<td>79%</td>
<td>89%</td>
</tr>
<tr>
<td>Phase I</td>
<td>6</td>
<td>100%</td>
<td>86%</td>
<td>92%</td>
</tr>
<tr>
<td>Phase II</td>
<td>5</td>
<td>98%</td>
<td>83%</td>
<td>90%</td>
</tr>
<tr>
<td>Phase III</td>
<td>4</td>
<td>98%</td>
<td>90%</td>
<td>93%</td>
</tr>
<tr>
<td>Probes</td>
<td>10</td>
<td>100%</td>
<td>78%</td>
<td>91%</td>
</tr>
</tbody>
</table>
A total of sixty-five inter-observer sessions measured three thousand two hundred letter strokes. There were a total of two thousand eight hundred eighty-six agreements and three hundred fourteen disagreements. The percentage of agreement ranged from a low of 79 per cent to a high of 100 per cent with an average mean of 90 per cent.

Presentation and Analysis of Data

The first data of manuscript letter strokes meeting criterion were analyzed to answer the question: "Will the measurement procedures be sensitive enough to show changes in manuscript letter formation?" To determine the sensitivity of these measurements, two variables were analyzed. First, as the subject's manuscript handwriting samples varied from day to day, would the instrument measure these changes. Secondly, could two independent observers measure a subject's manuscript handwriting sample and agree on the letter strokes meeting criterion. Six subject's manuscript handwriting samples of twenty letters were measured, converted to per cents and graphed (Figure 2.). Table 3 is a summary of the percentage of letter strokes meeting criterion.

Four independent observers each measured ten samples of forty-two letter strokes to determine inter-observer agreement. This agreement was measured while comparing 1,680
PERCENTAGE OF CORRECT MANUSCRIPT LETTER STROKES

CHAD

LISA

BRUCE

KISSA

JOHN

BILL

* Measurement Criterion = deviation of less than two millimeters from model letters

** Deviation less than three millimeters

Fig. 2. Percentage of correct manuscript letter strokes of six subjects during baseline conditions.
### TABLE 3

Percentage of Correct Manuscript Letter Strokes

<table>
<thead>
<tr>
<th>Subject</th>
<th>No. of Sessions</th>
<th>Lowest Per Cent Meeting Criterion</th>
<th>Highest Per Cent Meeting Criterion</th>
<th>Mean Per Cent Meeting Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>50%</td>
<td>81%</td>
<td>72%</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>57%</td>
<td>81%</td>
<td>72%</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>33%</td>
<td>50%</td>
<td>44%</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>48%</td>
<td>64%</td>
<td>57%</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>40%</td>
<td>74%</td>
<td>62%</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>36%</td>
<td>67%</td>
<td>54%</td>
</tr>
<tr>
<td>Samples (42 letter strokes per sample)</td>
<td>Number of Agreements</td>
<td>Percentage of Agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>36</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>35</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>35</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>35</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>41</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>34</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>41</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>40</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>41</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>39</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>33</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>36</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>40</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>36</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>38</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>41</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>37</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>34</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>34</td>
<td>81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
letter strokes of the letters f, h, m, u, j, q, v, x, b, e (Table 3). The agreement ranged from a low of 79 per cent to a high of 98 per cent with a mean of 89 per cent.

Individual Findings

Subject One

Figure 3 represents the percentage of trained and untrained letter strokes meeting criterion by sessions. Figure 4 represents total number of trained letter strokes meeting criterion during each session.

Baseline: During three days of baseline, 62 per cent of the subject's letter strokes met criterion. The mean number of strokes was 26.0 out of a possible mean of 42.

Intervention I: The subject's percentage of correct strokes increase from 62 per cent during baseline to 74 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 30.9 out of a possible 42. Two target letters, j and q each with a possible four strokes, were introduced to the intervention. The letter j had a mean of 1.0 correct responses during baseline and 3.1 after intervention. The letter q had a mean of 1.9 correct
SUBJECT ONE
Evaluation = Two Millimeter Deviation

![Graph showing percent of trained and untrained letter strokes meeting criterion by sessions](image)

**Intervention I:** Verbalization of errors

**Intervention II:** Verbalization of errors and copying again with verbalization of errors for incorrect response

**Intervention III:** Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

**Probe:** Untrained letters

**Probe Sentence writing trained letters**

**Probe:** Sentence writing untrained letters

**Fig 3** Percent of trained and untrained letter strokes meeting criterion by sessions
Fig 4. Total number of trained letter strokes meeting criterion during each session. The solid line (---) represents baseline while dashed line (----) represents intervention.

**SUSPECT ONE**

**Evaluation = Two Millimeter Deviation**

- **Intervention I**: Verbalization of error
- **Intervention II**: Verbalization of errors and copying again with verbalization of error for incorrect response
- **Intervention III**: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response
response the eleven sessions before intervention and 2.8 after eleven intervention sessions.

**Intervention II:** The subject's percentage of correct responses decreased from 74 per cent during intervention I to 70 per cent. During this intervention, whenever a subject's target letter did not meet criterion, he verbalized how they differed from the model, copied them again, and verbalized his errors. The mean number of correct responses was 29.4 out of a possible 42. Three target letters, f, j, and q each with a possible four strokes, were introduced to intervention. During intervention I target letter j had a mean of 3.1 correct responses and 3.8 after intervention II. The letter q had a mean of 2.4 correct responses during intervention I and 2.6 after intervention II. The letter f had a mean of 2.0 correct responses the two sessions before intervention II and 2.7 after five intervention sessions.

**Intervention III:** When potential reinforcement was added contingent on letters meeting criterion the first time, to verbalization of errors and copying again with verbalization for incorrect responses, the percentage of correct responses increased from 70 per cent during intervention II to 72 per cent. The mean number of correct responses was 30.3 out of a possible 42. Five target letters f, h, b, j, and q were chosen and introduced to intervention III during the
first session. The letters v, u, x, m, and e were added to the intervention in that order. A letter was added each successive session starting with the second session of intervention III. There was not a significant increase in the mean number of correct responses of letters f, h, b, j, q, v, u, x, m, and e when introduced to intervention III.

Subject Two

Figure 5 represents the percentage of trained and untrained letter strokes meeting criterion by sessions. Figure 6 represents total number of trained letter strokes meeting criterion during each session.

Baseline: During four days of baseline, 68 per cent of the subject's letter strokes met criterion. The mean number of correct responses was 28.9 out of a possible 42.

Intervention I: The subject's percentage of correct responses increased from 68 per cent during baseline to 75 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 31.7 out of a possible 42. Two target letters, j and q each with a possible four strokes, were introduced to the intervention. The letter q had a mean of 1.0 correct responses during baseline and 2.6
SUBJECT TWO

Evaluation - Two Millimeter Deviation

Baseline Intervention I Intervention II Intervention III

100 80 60 40 20 0

Percent of Correct Responses

0 10 20 30 40 50 60 70 80 90 100

1 5 10 15 20 25 30 35 Sessions

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig 5 Percent of trained and untrained letter strokes meeting criterion by session
SUBJECT TWO

A Evaluation - Two Millimeter Deviation
B Evaluation - One Millimeter Deviation

Intervention I: Verbalization of error
Intervention II: Verbalization of errors and copying again with verbalization of error for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig 6 Total number of trained letter strokes meeting criterion during each session. The solid line (-----) represents baseline while dashed line (----) represents intervention.
after intervention. The letter j had a mean of 2.0 correct responses the thirteen sessions before intervention and 3.2 after five intervention sessions.

**Intervention II**: The subject's percentage of correct responses increased from 75 per cent during intervention I to 76 per cent. During this intervention, whenever a subject's target letter did not meet criterion, he verbalized how they differed from the model, copied them again, and verbalized his errors. The mean number of correct responses was 31.9 out of a possible 42. Three target letters, b, j, and q each with a possible four strokes, were introduced to intervention. The target letter q had a mean of 2.6 responses during intervention I and 3.30 after intervention II. The letter j had a mean of 2.3 correct responses during intervention I and 3.3 after intervention II. The letter b had a mean of 2.3 correct responses the seven sessions before intervention and 4.0 after three intervention sessions.

**Intervention III**: When using the instrument with two millimeter deviation, the subject's response rate was 90 per cent or better during the last three sessions of intervention II. An instrument with one millimeter deviation was chosen to evaluate the subject's letter formation during intervention
When potential reinforcement was added contingent on letters meeting criterion the first time, to verbalization of errors and copying again with verbalization for incorrect responses, the per cent of correct responses was 71 per cent. The mean number of correct responses was 30.0 out of a possible 42. The five letters f, b, j, q, and u were chosen as target letters and introduced to intervention III. The letters v, h, m, x, and e were introduced in that order. A letter was added each successive session starting with the second session of intervention III. There was an increase in the mean number of correct responses for each of the five letters, v, h, m, x, and e, after they were introduced to intervention III.

Subject Three

Figure 7 represents the percentage of trained and untrained letter strokes meeting criterion by sessions. Figure 8 represents total number of trained letter strokes meeting criterion during each session.

Baseline: During six days of baseline, 42 per cent of the subject's letter strokes met criterion. The mean number of correct responses was 17.5 out of a possible mean of 42.
SUBJECT THREE
Evaluation = Two Millimeter Deviation

Baseline

Intervention I

Intervention II

Intervention III

Percent of Correct Responses

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig 7: Percent of trained and untrained letter strokes meeting criterion by sessions
SUBJECT THREE
Evaluation = Two Millimeter Deviation

Intervention I: Verbalization of errors

Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response

Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig B Total number of trained letter strokes meeting criterion during each session. The solid line (-----) represents baseline while dashed line (----) represents intervention
**Intervention I:** The subject's percentage of correct responses increased from 42 per cent during baseline to 60 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 25.4 out of a possible 42. Three target letters e, j, and q each with a possible four strokes were introduced to the intervention. The letter q had a mean of 0.7 correct responses during baseline and 2.3 after intervention. The letter j had a mean of 2.2 correct responses the six sessions before intervention and 3.1 after twenty-one intervention sessions. The letter e had a mean of 0.8 correct responses the eighteen sessions before intervention and 1.3 after nine intervention sessions.

**Intervention II:** The subject's percentage of correct responses increased from 60 per cent during intervention I to 70 per cent. During this intervention when a subject's target letters did not meet criterion, he verbalized how they differed from the model, copied them again and verbalized his errors. The mean number of correct responses was 29.5 out of a possible 42. Three target letters j, q, and m were introduced to intervention. Letters j and q had a possible four strokes and letter m six strokes. The letter q had a mean of 2.3 correct responses during intervention I and 3.5 after intervention II. The letter j had a mean of 3.0 correct responses the five sessions before intervention.
and 3.7 after six intervention sessions. The letter m had a mean of 2.9 correct responses the seven sessions before intervention and 4.8 after four intervention sessions.

**Intervention III**: When potential reinforcement was added contingent on letters meeting criterion the first time, to verbalization of errors and copying again with verbalization for incorrect responses, the per cent of correct responses increased from 70 per cent during intervention II to 87 per cent. The mean number of correct responses was 36.4 out of 42. The five letters f, u, j, q, and m were chosen as target letters and introduced to intervention III during the first session. The letters e, v, b, x, and h were introduced to intervention in that order. A letter was added each successive session starting with the second session of intervention III. There was an increase in the mean number of correct responses of letters f, h, b, e, x, j, q, m, and u, when introduced to intervention III.

**Subject Four**

Figure 9 represents the percentage of trained and untrained letter strokes meeting criterion by sessions. Figure 10 represents total number of trained letter strokes meeting criterion during each session.
SUBJECT FOUR

Evaluation - Three Millimeter Deviation  Evaluation - Two Millimeter Deviation

Baseline Intervention IA Intervention IB Intervention II

<table>
<thead>
<tr>
<th>Sessions</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of Correct Responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Probe: Untrained letters
Sentence writing trained letters
Sentence writing untrained letters

Fig 9: Percent of trained and untrained letter strokes meeting criterion by sessions
Number of Correct Responses

Baseline

Intervention

SUBJECT FOUR
A Evaluation - Three Millimeter Deviation
B Evaluation - Two Millimeter Deviation

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig 10 Total number of trained letter strokes meeting criterion during each session. The solid line (-----) represents baseline while dashed line (----) represents intervention.
Baseline: During five days of baseline, 57 per cent of subject's letter strokes met criterion. The mean number of correct responses was 23.8 out of a possible mean of 42.

Intervention I A: The subject's percentage of correct responses increased from 57 per cent during baseline to 80 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 33.6 out of a possible 42. Three target letters f, j, and q each with a possible four strokes were introduced to the intervention. The letter f had a mean of 1.2 correct responses during baseline and 3.7 after intervention. The letter j had a mean of 1.9 correct responses the fourteen sessions before intervention and 3.6 after five intervention sessions. The letter q had a mean of 1.2 correct responses the five sessions before intervention and 3.0 after fourteen intervention sessions.

When the subject's sample letters met 100 per cent criterion using the evaluative instrument with a 3 millimeter deviation, intervention II was again introduced to the subject using an evaluative instrument with a deviation of 2 millimeters.

Intervention I B: When the evaluative instrument allowing only 2 millimeters of deviation from the model letter was
introduced the percentage of correct response dropped to 52 per cent. During the sixteen sessions of intervention I B there was an increasing trend in the percentage of correct responses with a mean 76 per cent correct responses. The mean number of correct responses was 32.0. Five target letters, f, e, x, j, and m, were introduced to this intervention. Letter f maintained a mean of 3.75 out of 4. The letter e had a mean of 2.0 correct responses the fourteen sessions before intervention and 3.5 after two intervention sessions. The letter x had a mean of 3.0 correct responses the twelve sessions before intervention and 3.5 after four intervention sessions. The letter j had a mean of 2.6 correct responses the eight sessions before intervention and 3.5 after eight intervention sessions. The letter m had a mean of 2.9 correct responses the nine sessions before intervention and 5.0 after seven intervention sessions.

Intervention III: When potential reinforcement was added contingent on letters meeting criterion the first time, to verbalization of errors and copying again with verbalization for incorrect responses, the per cent of correct responses increased from 76 per cent during intervention I B to 89 per cent. The mean number of correct responses was 37.5. The five letters f, e, x, j, and m were chosen as target letters and introduced to intervention III during the first session. The letters q, b, h, v, and u were introduced to
intervention in that order. A letter was added each successive session starting with the second session of Intervention III. There was an increase in the number of correct responses for the letters f, b, e, j, q, and m when introduced to intervention III.

Subject Five

Figure 11 represents the percentage of trained and untrained letters meeting criterion by sessions. Figure 12 represents total number of trained letter strokes meeting criterion during each session.

Baseline: During six days of baseline 63 per cent of the subject's letter strokes met criterion. The mean number of correct responses was 26.6 out of a possible mean of 42.

Intervention I: The subject's percentage of correct responses increased from 63 per cent during baseline to 70 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 29.3 out of a possible 42. Four target letters x, j, q, and m were introduced to the intervention. Letters x, j, and q had a possible four strokes and m a possible six strokes. The letter x had a mean of 2.2 correct responses the five sessions before intervention and
SUBJECT FIVE
Evaluation = Two Millimeter Deviation

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig. II. Percent of trained and untrained letter strokes meeting criterion by sessions
SUBJECT FIVE
Evaluation = Two Millimeter Deviation

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response
Intervention III: Verbalization of errors, potential reinforcer for correct response and copying again plus verbalization of errors for incorrect response

Fig. 12 Total number of trained letter strokes meeting criterion during each session. The solid line (-----) represents baseline while dashed line (----) represents intervention.
4.0 after three intervention sessions. The letter j had a mean of 0.9 correct responses during baseline and 2.1 after intervention. The letter q had a mean of 1.9 correct responses during baseline and 3.5 after intervention. The letter m had a mean of 4.3 correct responses during baseline and 4.4 after intervention.

**Intervention II:** The subject's percentage of correct responses increased from 70 per cent during intervention I to 88 per cent. During the intervention, whenever a subject's target letter did not meet criterion, he verbalized how they differed from the model, copied them again and verbalized his errors. The mean number of correct responses was 36.8 out of a possible 42. Three target letters, x, j, and m, were introduced to intervention. Letters x and j had a possible four strokes and the letter m six strokes. The letter x had a mean of 2.9 strokes during intervention I and 3.9 after intervention II. The letter j had a mean of 2.1 strokes during intervention I and 3.9 after intervention II. The letter m had a mean of 4.0 correct responses the two sessions before intervention and 5.6 after seven intervention sessions.
Subject Six

Figure 13 represents the percentage of trained and untrained letters meeting criterion by sessions. Figure 14 represents the total number of trained letter strokes meeting criterion during each session.

**Baseline:** During five days of baseline, 57 per cent of the subject's letter strokes met criterion. The mean number of correct responses was 23.8 out of a possible 42.

**Intervention I:** The subject's percentage of correct responses increased from 57 per cent during baseline to 71 per cent when selected target letters were measured and the subject verbalized his errors in letter formation. The mean number of correct responses was 29.9 out of a possible 42. Three target letters x, j, q, each with a possible four strokes, were introduced to the intervention. The letter j had a mean of 0.4 correct responses during baseline and 3.1 after intervention. The letter x had a mean of 1.6 correct responses the six sessions before intervention and 3.4 after eight intervention sessions. The letter q had a mean of 1.6 strokes the eight sessions before intervention and 3.0 after six intervention sessions.
SUBJECT SIX
Evaluation = Two Millimeter Deviation

Fig 13 Percent of trained and untrained letter strokes meeting criterion by sessions
SUBJECT SIX
Evaluation = Two Millimeter Deviation

Intervention I: Verbalization of errors
Intervention II: Verbalization of errors and copying again with verbalization of errors for incorrect response

Fig 14. Total number of trained letter strokes meeting criterion during each session. The solid line (---) represents baseline while dashed line (-----) represents intervention.
**Intervention II:** The subject's percentage of correct responses increased from 71 per cent during intervention I to 85 per cent. During this intervention whenever a subject's target letters did not meet criterion, he verbalized how they differed from the model, copied them again and verbalized his errors. The mean number of correct responses was 35.7 out of a possible 42. The four letters, x, j, q, and m, were chosen as target letters and introduced to intervention. Letters x, j, and q had a possible four strokes and the letter m six strokes. The target letter j had a mean of 3.1 correct responses during intervention I and 3.3 correct responses after intervention II. The letter q had a mean of 2.2 responses during intervention I and 3.7 after intervention II. The letter m had a mean of 3.3 correct responses the three sessions before intervention and 5.3 after three intervention sessions. The letter x had a mean of 3.3 correct responses the three sessions before intervention and 3.7 the three sessions after intervention.

Table 5 is a summary of the mean number of correct responses out of a possible forty-two. Subject three has two means for intervention I, letter A and B. During intervention I A the evaluation instrument had a deviation of three millimeters and during I B, two millimeters. Subject three did not copy any letters during intervention II. Subjects five and six did not copy any letters during intervention III.
### TABLE 5
Mean Number of Correct Responses
Out of a Possible Forty-Two

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Subjects</th>
<th>Baseline</th>
<th>Intervention I</th>
<th>Intervention II</th>
<th>Intervention III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>26.0</td>
<td>30.9</td>
<td>29.4</td>
<td>30.3</td>
</tr>
<tr>
<td></td>
<td>Two</td>
<td>28.9</td>
<td>31.7</td>
<td>31.9</td>
<td>30.0*</td>
</tr>
<tr>
<td></td>
<td>Three</td>
<td>17.5</td>
<td>25.4</td>
<td>29.5</td>
<td>36.4</td>
</tr>
<tr>
<td></td>
<td>Four</td>
<td>23.8</td>
<td>33.6(A)32.0(B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Five</td>
<td>26.6</td>
<td>29.3</td>
<td>36.8</td>
<td>37.5**</td>
</tr>
<tr>
<td></td>
<td>Six</td>
<td>23.8</td>
<td>29.9</td>
<td>35.7</td>
<td></td>
</tr>
</tbody>
</table>

* Evaluation changed from two millimeters to one millimeter deviation.

** Evaluation changed from three millimeters to two millimeters deviation.

Summary of Probes: Two Probes, I and II, were given to each subject at the completion of the last days of baseline, intervention I, intervention II, and intervention III. The percentage of correct responses on Probe I and Probe II, for each of the six subjects, are graphically represented by figures 3, 5, 7, 9, 11, and 13. Probe I consisted of the sixteen untrained letters a, c, d, g, i, k, l, n, o, p, r, s, t, w, y, and z. Probe II consisted of the sentence: "the brave foxes quickly jump fences." The three tables 6, 7, and 8 are a summary of these Probes. Table 6 gives the subject's per cent of correct response for Probe I. Table
Table 7 gives the subject's per cent of correct responses for the trained letters in the sentence: "the brave foxes quickly jump fences." Table 8 gives the subject's per cent of correct responses for the untrained letters in the sentence: "the brave foxes quickly jump fences."

**TABLE 6**

Percentage of Correct Responses for Untrained Letters

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Subjects</th>
<th>Baseline</th>
<th>Intervention I</th>
<th>Intervention II</th>
<th>Intervention III</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>82</td>
<td>56</td>
<td>63</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>40</td>
<td>78</td>
<td>75</td>
<td>56*</td>
<td></td>
</tr>
<tr>
<td>Three</td>
<td>53</td>
<td>44</td>
<td>39</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Four</td>
<td>59</td>
<td>94</td>
<td>56**</td>
<td>68**</td>
<td></td>
</tr>
<tr>
<td>Five</td>
<td>64</td>
<td>76</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Six</td>
<td>68</td>
<td>69</td>
<td>54</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Evaluation changed from two millimeters to one millimeter deviation.

** Evaluation changed from three millimeters to one millimeter deviation.
### TABLE 7

Percentage of Correct Responses for Trained Letters in Sentence Writing

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Subjects</th>
<th>Baseline</th>
<th>Intervention I</th>
<th>Intervention II</th>
<th>Intervention III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>64</td>
<td>52</td>
<td>73</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Two</td>
<td>73</td>
<td>42</td>
<td>85</td>
<td>55*</td>
</tr>
<tr>
<td></td>
<td>Three</td>
<td>39</td>
<td>36</td>
<td>37</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Four</td>
<td>56</td>
<td>80</td>
<td>66**</td>
<td>73**</td>
</tr>
<tr>
<td></td>
<td>Five</td>
<td>66</td>
<td>97</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Six</td>
<td>67</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 8

Percentage of Correct Responses for Untrained Letters in Sentence Writing

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Subjects</th>
<th>Baseline</th>
<th>Intervention I</th>
<th>Intervention II</th>
<th>Intervention III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One</td>
<td>65</td>
<td>70</td>
<td>78</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Two</td>
<td>52</td>
<td>22</td>
<td>74</td>
<td>48*</td>
</tr>
<tr>
<td></td>
<td>Three</td>
<td>57</td>
<td>43</td>
<td>28</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Four</td>
<td>52</td>
<td>92</td>
<td>71**</td>
<td>70**</td>
</tr>
<tr>
<td></td>
<td>Five</td>
<td>49</td>
<td>78</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Six</td>
<td>52</td>
<td>64</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

* Evaluation changed from two millimeters to one millimeter deviation.

** Evaluation changed from three millimeters to one millimeter deviation.
CHAPTER V

DISCUSSION, LIMITATIONS, IMPLICATIONS, AND SUMMARY

Discussion

Research Questions:

The primary purpose of this study was to examine seven questions concerned with teaching manuscript letter formation. The first question dealt with the measuring instrument developed for this study. Conclusions to the last six research questions depended heavily on the conclusion drawn for question one. Separately the seven research questions will follow with a discussion of recorded data which pertained to each one.

1. Will measurement procedures be sensitive enough to show changes in manuscript letter formation?

To determine the sensitivity of these measurements two variables, percentage of correct response ranges and interobserver agreement, were analyzed. The range of per cent of correct responses (Figures 3, 5, 7, 9, 11, and 13) varied significantly for each subject. The ranges by
subjects were: subject one, 50 per cent - 88 per cent; subject two, 56 per cent - 94 per cent; subject three, 33 per cent - 98 per cent; subject four, 48 per cent - 100 per cent; subject five, 40 per cent - 92 per cent; and subject six, 36 per cent - 88 per cent. The mean range was 44 per cent - 93 per cent with an average difference of 49 per cent of correct responses. The interobserver agreement consisted of sixty-five sessions while using the evaluative instrument to measure a total of three thousand two hundred letter strokes from the four experimental conditions. A summary of these results (Table 2) showed a low of 79 per cent agreement, a high of 100 per cent agreement, with a significant mean agreement of 90 per cent. A mean range difference of 49 per cent with a mean interobserver agreement of 90 per cent, supports the conclusion that the measurement procedure is sensitive to show changes in manuscript letter formation.

2. Will student recognition of errors as indicated by the measurement procedure and verbalization of the error increase or decrease the occurrence of the error?

All letters did not receive treatment as target letters during intervention I; self verbalization of errors (Figures 4, 6, 8, 10, 12 and 14). A comparison of mean number of correct responses for baseline letters and those
introduced to intervention (target letters) and those letters not treated during intervention I were calculated. The number of baseline and intervention I sessions varied for each subject. A procedure was developed to equalize the number of sessions for each subject. All letters consisted of a possible four strokes, except letter m which has a possible six strokes. During each condition the total correct responses for the letter m was multiplied by the fraction 2/3 to reduce it proportionally to a possible four strokes. (2/3 • 6 = 4). Data from intervention I B for subject four were not included because the deviation of the measuring device was changed from three millimeters during baseline and intervention I A to two millimeters. The mean number of correct responses for intervention I target letters increased from a baseline mean of 1.46 correct responses to 2.89 when the subjects verbalized their errors. The mean number of correct responses for the letters not introduced to intervention I increased from a mean of 2.57 correct responses during baseline to 2.84. The target letters demonstrated an increase in mean correct responses of 98 per cent from baseline to intervention I. The non-target letters demonstrated an increase in mean correct responses of 11 per cent from baseline to intervention I. An analysis of the data demonstrates: When the subjects verbalized their errors in letter formation there was a significant decrease in letter formation errors.
3. If student verbalization of manuscript errors generates no change in accuracy of the letter formation, will immediately copying the letter a second time increase or decrease accuracy?

All letters did not receive treatment as target letters during intervention II: self verbalization of errors and immediately copying the letter again (Figures 4, 6, 8, 10, 12 and 14). Two comparisons of mean number of correct responses for target letters introduced to intervention II were calculated. The mean number of correct responses for letters introduced to intervention II were compared to the mean number of correct responses for letters not introduced to intervention. The mean number of correct responses for target letters introduced to intervention II were compared to the same letters introduced to intervention I. The number of intervention I and intervention II sessions varied for each subject. A procedure was developed to equalize the number of sessions for each subject. All letters consisted of a possible four strokes except the letter m which had a possible six strokes. During each condition the total correct responses for the letter m were multiplied by the fraction 2/3 to reduce it proportionally to a possible four strokes. The conclusions drawn are from data analyzed from subjects one, two, three, five, and six, since subject four did not copy any letters during intervention II.
The target letters introduced to intervention II had a mean of 3.51 correct responses and those not introduced to intervention II, a mean of 3.00 correct responses. When analyzing that comparison a compounding effect of intervention I and the addition of recopying and verbalization of errors a second time both could have been variables effecting the change. To look at the effect of recopying a target letter a second time and verbalizing the errors on mean correct responses a comparison of target letters introduced to both intervention I and II were made. Mean number of correct strokes were calculated for only letters that were introduced to intervention I and then introduced to intervention II. The target letters introduced to intervention I had a mean of 2.73 correct responses while those introduced to intervention II had a mean of 3.37 correct responses. The letters not introduced to intervention II demonstrated an increase in mean of correct responses from 2.85 for all letters during intervention I to 3.00 during intervention II. The target letters introduced to intervention II demonstrated an increase in mean of correct responses from 2.85 for all letters during intervention I to 3.51 during intervention II. The non-target letters demonstrated a 5 per cent increase in mean correct responses and the target letters introduced to intervention a 23 per cent increase in mean correct responses over intervention I. An analysis of the
data demonstrates that when the subjects immediately copied a letter and verbalized his errors a second time, the accuracy of his letter formation increased as demonstrated by increases in mean number of correct responses.

4. If student verbalization of manuscript errors and copying the letter a second time generates no change in accuracy of letter formation, will potential reinforcing consequences increase or decrease manuscript accuracy?

All letters did not receive treatment as target letters during intervention III (Figures 4, 6, 8, 10, 12, and 14). A subject verbalized the target letter's errors in letter formation and received a potential reinforcer when the letter met criterion. When the target letter did not meet criterion, the subjects immediately copied the letter again and verbalized their errors. Two comparisons of mean number of correct responses for target letters introduced to intervention III were calculated. The mean number of correct responses for target letters introduced to intervention III were compared to those letters which were not introduced to intervention III. The mean number of correct responses to target letters introduced to intervention II were compared to the mean number of correct responses to the same letters when introduced to intervention III. The number of intervention II and intervention III sessions varied for each.
subject. A procedure was developed to equalize the number of sessions for each subject. All letters consisted of a possible four strokes except the letter m which had a possible six strokes. During each condition the total correct responses for the letter m were multiplied by the fraction 2/3 to reduce it proportionally to a possible four strokes. The conclusions drawn are from data analyzed from subject one and three. Subject two sample letters during intervention II were evaluated with a two millimeter deviation and one millimeter, during intervention III. Subject four did not copy any letters during intervention II and subjects five and six did not copy any letters during intervention III. The target letters introduced to intervention III had a mean of 3.42 correct responses and those not introduced a mean of 2.90 correct responses. Six target letters introduced to intervention II demonstrated a mean of 3.32 correct responses. When these same six letters were introduced to intervention III they demonstrated a mean of 3.13 correct responses. The mean number of correct responses of these six target letters did decrease from their introduction to intervention II to intervention III. To determine if intervention III was effective in increasing manuscript accuracy, intervention II data for subjects one and three were compared to that from intervention III. The mean number of correct responses increased from 2.70 to 2.90 for the letters not introduced to
intervention II or III. The target letters introduced to intervention II demonstrated a mean of 2.63 correct responses and those to intervention III a mean of 3.42 correct responses. The non-target letters demonstrated a 7 per cent increase in correct responses from intervention II to intervention III. The target letters introduced to intervention III, when compared to the target letters introduced to intervention II, demonstrated a 30 per cent mean increase in correct responses. When analyzing the data from subjects one and three the addition of a potential reinforcer for letters meeting criterion to verbalization of errors and recopying those letters not meeting criterion significantly increased manuscript accuracy.

5. If a change occurs in one stroke used to construct a letter, will this change generalize to other non-trained letters using the same stroke?

The percentage of trained letter strokes meeting criterion increased from baseline to intervention I and each succeeding intervention (Figures 3, 5, 7, 9, 11, and 13) except for subject one. Subject one's percentage of correct responses for each intervention was greater than that of baseline (Figure 3). The percentage of correct responses for the sixteen untrained letters (Table 6) was analyzed to determine if a change in the percentage of untrained letter
strokes meeting criterion was proportional to the changes in percentage of trained letter strokes meeting criterion. The percentages of untrained letter strokes meeting criterion showed a decreasing trend for subjects one and six; increasing trend for subjects two, four, and five; and both a decreasing and increasing trend for subject three. An analysis of the data shows an increasing trend in the percentage of trained letter strokes meeting criterion and no clear trend in untrained letter strokes. This study has not demonstrated a generalization of changes in trained letter stroke accuracy to untrained letter strokes.

6. If a change occurs in one letter of a letter-family, will that change generalize to other non-trained letters of that family?

The percentage of trained letters meeting criterion increased from baseline to intervention I and each succeeding intervention (Figures 3, 5, 7, 9, 11, and 13) except for subject one. Subject one's percentage of correct responses for each intervention was greater than that of baseline (Figure 3). The percentage of correct responses for the sixteen untrained letters (Table 6) was analyzed to determine if a change in the percentage of untrained letters meeting criterion was proportional to the changes in percentage of trained letters meeting criterion. The percentage
of untrained letters meeting criterion showed a decreasing trend for subjects one and six; increasing trend for subjects two, four, and five; and both a decreasing and increasing trend for subject three. An analysis of the data shows an increasing trend in the percentage of trained letters meeting criterion and no clear trend in percentage of untrained letters meeting criterion. This study has not demonstrated a generalization of increases in trained letters meeting criterion to untrained letters meeting criterion.

7. If a change occurs in one letter of a letter-family will that change generalize to other non-trained letters being used in writing a sentence?

The Probe: "the brave foxes quickly jump fences" was copied by subjects one, two, three and four, four times, and by subjects five and six three times. The trained letters were measured and the percentage of correct responses recorded (Table 7). The untrained letters were also measured and the percentage of correct responses recorded (Table 8). An analysis of Table 7 demonstrates an increasing trend in the percentage of correct letter strokes for trained letters as each subject proceeds from baseline to his last intervention. An analysis of Table 8 demonstrates an increasing trend in the percentage of correct letter strokes for the sixteen untrained letters for all six subjects. No
conclusion is drawn but to state a slight increasing trend in the percentage of trained letter strokes used in sentence writing is matched by an increasing trend in the percentage of untrained letter strokes meeting criterion in sentence writing.

Limitations

The evaluative instruments developed for this study were designed to measure deviations of three, two, and one millimeters from lower-case manuscript letters. The criterion described in Chapter Three, to be used in manufacturing the three and two millimeter deviation evaluative instruments, were not stringent enough for the letter e. Decisions made when measuring the letter e with the three and two millimeter evaluation instruments were sometimes subjective.

The choice of target letters for introduction during each intervention phase was determined by the letters exhibiting the lowest mean correct response rate from the previous experimental condition. The choice of letters exhibiting the greatest per cent of errors also increases the probability they would show the greatest increase in the mean number of correct responses.

Comparisons between the effects of intervention I and II, on changing manuscript accuracy, were justified since data were available for six subjects during intervention I
and five subjects for intervention II. Any implications drawn from the data, when intervention III was introduced, were made from comparisons of two subjects' data to five subjects' data.

A fourth limitation of this study is that it cannot be said which of these three would have produced the largest in the mean number of correct responses if each subject was introduced to only one: (1) self verbalization of errors in letter formation, (2) self verbalization of errors combined with immediate recopying and verbalization of errors when the first letter did not meet criterion, (3) the addition of possible reinforcing consequence for correct letter formation to verbalization of errors. All three of these procedures produced an increase in the mean number of correct responses.

Implications

An objective procedure for measuring changes in manuscript letter formation is now available. The evaluation instruments are economically feasible so each classroom teacher can have them available for their students. Copies of the evaluative instrument were made on heavy plastic at a unit cost of 40¢. A complete set of three measuring instruments, three, two, and one millimeter deviations could be made for $1.20 per student. These instruments showed
very little wear during the six months they were used and would last a classroom teacher several years. The use of these instruments required a minimal amount of time. The set of twenty-six lower case manuscript letters are constructed from sixty-three letter strokes. Four trained evaluators measured the time it required them to measure forty ten-letter samples of manuscript handwriting and record their evaluations. They found they could evaluate a mean of 12.89 letter strokes per minute. A classroom teacher, at the rate of 12.89 strokes per minute, could measure and record a student's manuscript sample of the total alphabet in less than five minutes. Six naive people, ranging in age from ten to thirty-eight years, were selected to measure manuscript samples of ten letters. The behavior definition and recording procedure was read and explained to each of the six people. They were given one practice session to measure and record twenty-one letter strokes and to have any questions answered regarding the use of the evaluative overlays. The time required for each of the six to measure four handwriting samples, of twenty-one strokes, was measured (Table 9). The six naive evaluators measured and recorded with a mean of 7.46 strokes per minute. These six evaluators were trained to use the evaluative instruments in less than ten minutes. The six naive evaluators had an interobserver agreement of 88 per cent with the trained evaluators. Teachers, or students, may be taught with a
minimal amount of time to be accurate evaluators of their own, or others, manuscript letter formation. During the study a subject verbalized his errors to the experimenter as he placed the evaluative instrument over his letter to compare how his letters differed from the model. This method of a visual feedback of his letters deviation from the model may be a procedure for students to self-evaluate their manuscript letter formation.

Many teachers of exceptional children are faced with the task of preparing their students for regular class placement. One behavior required of the exceptional child, as soon as he becomes settled in his new regular class, is written responses to academic tasks. Many exceptional children do not demonstrate legible letter formation. The use of the clear plastic evaluation instrument as a visual feedback to the student about his manuscript letters, could increase his letter formation accuracy.

Recommendations for Further Research

A series of studies could be proposed from results of the present study. The three intervention procedures, as shown in the graphic presentations of results (Figures 4, 6, 8, 10, 12, and 14), demonstrated the mean number of correct responses increased during each procedure. A future study could evaluate subject changes in letter formation when
TABLE 9
Summary of Strokes Measured and Recorded Per Minute By Six Naive Evaluators

<table>
<thead>
<tr>
<th>Naive Observers</th>
<th>Evaluation Session</th>
<th>Strokes Measured and Recorded Per Minute</th>
<th>Mean Strokes Per Minute</th>
<th>Per cent of Interobserver Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>1</td>
<td>4.06</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.46</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.67</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.74</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>1</td>
<td>4.76</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.81</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.00</td>
<td>7.66</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.08</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>#3</td>
<td>1</td>
<td>9.54</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.63</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10.70</td>
<td>9.31</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8.00</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>#4</td>
<td>1</td>
<td>8.71</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12.00</td>
<td>9.60</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8.40</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.00</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td>#5</td>
<td>1</td>
<td>5.60</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.00</td>
<td>6.32</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.26</td>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#6</td>
<td>1</td>
<td>6.00</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>6.00</td>
<td>6.00</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.00</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.70</td>
<td></td>
<td>86</td>
</tr>
</tbody>
</table>
copying model letters and using the evaluative instruments as self-visual feedback. Another study could evaluate the changes in letter formation as a subject copies a model letter, evaluates his copy with the evaluative instrument, and receives a potential reinforcer for each letter meeting criterion. A replication of this study showing students as effective as adults, when listening to subjects verbalize their errors in letter formation, could reduce teachers work load. Further studies on generalization of letter strokes need to be conducted. Do combinations of strokes in trained letters showing increases in correct formation generalize to untrained letters using the same combinations of strokes (b to d, g to q)? Will changes in letter formation, during manuscript letter formation training classes, generalize to written academic assignments later the same school day?

Out of the above recommended studies would grow more recommendations for further studies. These studies would furnish more empirical data on the reliability of the evaluative instruments and visual feedback on changes in letter formation.

Summary

Handwriting is a common skill taught to all children receiving a formal education. It is a set of symbols used to communicate with others. An objective procedure for
evaluating these symbols which was feasible in teacher time and cost to schools has eluded educators.

The study examined three areas in letter formation: measurement procedures, training procedures, and generalizations. A set of ten lower-case manuscript letters were chosen as model letters for six subjects, ages six to nine years, to copy. A set of transparent overlays were designed to measure deviation ranges of student's samples from model letters. The first set measured deviations from zero to one millimeter. The second from zero to two millimeters and the third from zero to three millimeters.

The design employed in the study was a combination of two behavioral designs. The study used a multiple baseline design when introducing subjects to experimental conditions and when introducing target letters within experimental conditions. The study consisted of four experimental conditions: baseline, self-verbalization of errors in letter formation (intervention I), self-verbalization of errors in letter formation and immediately recopying the letter again with verbalization of errors when not meeting criterion (intervention II), and self-verbalization of errors in letter formation paired with a potential reinforcer for meeting criterion or immediately recopying and verbalization of errors if the letter did not meet criterion (intervention III). The subjects' baseline samples of forty-two letter
strokes were evaluated with the evaluative instruments of three, two and one millimeter deviations. An evaluation instrument was selected for each subject that demonstrated approximately 50 per cent of the subject's letter strokes meeting criterion during baseline.

Three experimenters conducted experimental sessions with two subjects in the following manner. They met with one subject at a table either inside of the subject's classroom or in an adjoining room. The subject copied the ten model letters once and, after three to five minutes, copied the ten letters again. This procedure was used with each subject during each experimental condition. Two Probes of the sixteen remaining lower-case manuscript letters and the sentence: "the brave foxes quickly jump fences," were copied by subjects at the end of each experimental session.

Following each experimental session the experimenter used the evaluative instrument with the correct deviation range to measure and record each manuscript letter stroke meeting criterion. The set of three evaluative instruments were examined to determine if they could measure change in letter formation and if two people could agree upon these measurements. Sixty-five experimental sessions were selected from all experimental conditions and three thousand two hundred letter strokes were evaluated by a second observer. The six subjects' manuscript handwriting samples
demonstrated a range of 33 per cent to 100 per cent of the strokes meeting criterion. The interobserver agreement for these 65 sessions ranged from a low of 79 per cent to a high of 100 per cent with a mean agreement of 90 per cent.

Selected target letters introduced to intervention I, self-verbalization of errors, demonstrated a 98 per cent increase in mean correct responses over baseline while the non-target letters demonstrated an 11 per cent increase. During intervention II each subject self-verbalized his errors and immediately recopied and verbalized again his letters not meeting criterion. The target letters demonstrated a 23 per cent increase in mean correct responses over the mean for intervention I. The non-target letters demonstrated a 5 per cent increase in mean correct responses over the mean for intervention I. During intervention III each subject verbalized his errors in letter formation and received a potential reinforcer when the letter met criterion. When the letter did not meet criterion he immediately recopied it and verbalized his errors. The target letter introduced to intervention III demonstrated a 30 per cent mean increase in correct responses over the mean for the target letters introduced to intervention II. The non-target letters demonstrated a mean increase of 7 per cent in correct responses over the mean for non-target letters during intervention II.
The Probes given to each subject were analyzed to see if increases in correct responses of the ten model letters generalized to the sixteen letters of the Probes. The data demonstrated an increase in the mean number of letter strokes for the ten model letters but not generalization to an increase in the mean number of letter strokes for the other sixteen untrained letters. There was a small increase in the mean number of correct responses for the ten trained letters used in sentence writing and a similar increase in the correct responses for untrained letters.

It could be concluded that: (a) The evaluative instruments were sensitive enough to measure changes in manuscript letter formation. (b) Self-verbalization of errors paired with visual feedback from the plastic overlay significantly increased accuracy in letter formation. (c) Self-verbalization of errors in letter formation using the plastic overlay and recopying the letter again with verbalizations increase the accuracy of letter formation. (d) Self-verbalization of errors in letter formation using the plastic overlay paired with a potential reinforcer for correct formation and/or recopying and verbalization again for incorrect formation increased letter form accuracy. (e) Improvements in letter stroke formation of trained letters demonstrates little or no generalization to improvement in non-trained letter stroke formation.
APPENDIX A

Roman, Upright and Slanted
Italic Alphabet
Upper-case italic in width groups as taught in grade one and all subsequent grades.

Same lower-case italic as joined for full cursive hand development during grade two.
Lower-case italic (in basic teaching "families") as taught in grade one and all subsequent grades.
APPENDIX B

Transparent Overlay Used in the Measurement of Letter Angles
Design for the transparent overlay used in the measurement of shape dislocation. The dotted line indicates a student sample; the black line, a model letter.
APPENDIX C

Measurement of Letter Shape Dislocation
Design for the transparent overlay used in the measurement of letter angles.
APPENDIX D

Set of Ten Model Letters
APPENDIX E

First Probe Sheet
APPENDIX F

Second Probe Sheet
the brave foxes
quickly jump fences.
APPENDIX G

Transparent Overlay
APPENDIX H

Recording Sheet
Recording of Letter Evaluation ( = Met Criterion)

f

h

m

u

j

SUBJECT =

MODEL NO. =

TOLERANCE =

EXP. CONDITION =
APPENDIX I

Copy Paper (Model Letters)
APPENDIX J

Copy Paper (Probe One)
APPENDIX K

Copy Paper (Probe Two)
REFERENCES


West, P. V. Chart for Diagnosing Elements in Handwriting, Bloomington, Ill.: Public School Publishing Co., 1926.

West, P. V. Handwriting: Elements of Diagnosis and Judgment of Handwriting, Bloomington, Ill.: Public School Publishing Co., 1926.