A COMPARISON OF THE EFFECTIVENESS OF USING SLIDES AND NON-VISUALS AS TEST INSTRUMENTS FOR DESIGN UNDERSTANDINGS

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY

Faye L. Brasington

1966

R350

1

ABSTRACT

A COMPARISON OF THE EFFECTIVENESS OF USING SLIDES AND NON-VISUALS AS TEST INSTRUMENTS FOR DESIGN UNDERSTANDINGS

by Faye L. Brasington

Matrix: Design for Living is a course designed to provide students with basic design understandings. Slide examples and illustrations are employed. Testing is accomplished through objective, machine scorable examinations. This research involves comparing the effectiveness of slide items and verbal items in the examinations. The hypothesis states that slide items will more effectively measure a higher level of intellectual ability than will verbal items, and also that slide items will be more discriminating.

Working in connection with an Educational Development Project, a study group wrote, revised and selected sixty slide items and sixty verbal items to be used as the final examination for TRA 140, Spring Term, 1966. The majority of these items were paired in subject matter and difficulty. All of the items had been pretested Winter Term, 1966. By means of item analysis, items on the pretest and final test were given an index of discrimination and index of difficulty by the Scoring Office of the Evaluation Services, Michigan State University. Each item was

assigned a classification level of intellectual ability according to the Taxonomy of Educational Objectives. Flanagan's index of discrimination, not affected by difficulty, was also computed for each item. The study group determined validity and the Kuder-Richardson method determined reliability. The comprehensive picture came from item analysis of total verbal items, total slide items, and the classification levels within the verbal and slide items. Computation of the correlation coefficients helped determine a comparison of reading scores to total verbal and total slide scores, and of CQT scores to total verbal and total slide scores. Student interviews and a student questionnaire revealed attitudes toward slide and verbal items.

The results showed the total slide items more difficult than the total verbal items, but less discriminating and less reliable. At the lower classification levels, verbal items were more discriminating and more reliable, but less difficult than slide items; however, at the highest level of intellectual ability, the opposite was true.

The questionnaire indicated that students believed the slide items more difficult than verbal items, but definitely worthwhile in the testing program. The highest correlation coefficient, .671, was between the slide items and the CQT scores.

A COMPARISON OF THE EFFECTIVENESS OF USING SLIDES AND NON-VISUALS AS TEST INSTRUMENTS FOR DESIGN UNDERSTANDINGS

By

Faye L. Brasington

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Textiles, Clothing and Related Arts

1966

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Mrs. Lorraine Gross and Dr. Mary Alice Burmester for their guidance in directing this research and assistance in writing and reviewing test items. I also gratefully acknowledge the thoughtful suggestions and understanding help given by Dr. Mary Gephart, and the assistance given by Dr. Elinor Nugent, Dr. Gertrude Nygren and Mr. Robert Bullard. To Miss Louisa Starr for her help in writing items, and Miss Bette Guiliani for her assistance with statistical procedures, I am indebted.

TABLE OF CONTENTS

		Page
ACKNOWLEDGMENTS		ii
LIST OF TABLES		iv
LIST OF ILLUSTRATIONS		v
Chapter		
I.	INTRODUCTION	1
	Justification	3
	Focus of the Study	7
II.	REVIEW OF LITERATURE	9
	Comparison of Visual and Verbal Test Items	9
	Related Tests	11
	Related Research	16
III.	METHODOLOGY	19
	Construction of the Test	19
	Classification Levels of Items	28
	Administration of the Test	29
	Data Analysis of Final Items	29
	Determination of the Quality of the Test	30
	Factor Analysis of Possible Influence on Test Results	31
IV.	FINDINGS AND INTERPRETATIONS	34
	Comparison of Total Verbal and Slide Items	36
	Comparison of Slide Items and Verbal Items Concerning Specific Levels of Classification of Intellectual Abilities	37
	Other Factors of Possible Influence on the Test Results	41
V.	SUMMARY AND RECOMMENDATIONS	43
	Summary	43
	Recommendations for Further Research	45
BIBLIOGRAPHY		46
APPENDICES		50

LIST OF TABLES

Table		Page
1.	Summary Data Item Analysis for Indices of Difficulty and Discrimination and Reliability for the TRA 140 Final Examination, Spring Term, 1966	34
2.	Summary Data Item Analysis for Each Classification Level of Items for the TRA 140 Final Examination, Spring, 1966	39
3.	Correlation Coefficients of the Individual Total Verbal Item Scores and Total Slide Item Scores Obtained by Students in TRA 140 with Their Reading Scores and CQT Scores. . .	41

LIST OF ILLUSTRATIONS

Figure Page

1. Slide Item Illustration for the Final Examination of TRA 140, Spring Term, 1966. . . . 22

2. Slide Item Illustration Using a Slide Pair for the Final Examination of TRA 140, Spring Term, 1966. 23

CHAPTER I

INTRODUCTION

The development of discernment and criticism in relation to art objects and general design is important. Modern society consists largely of consumers rather than producers of art. The design judgments and choices made are a part of our daily living.

> Design is an integral part of every man's environment. The design quality of this environment and the satisfaction it affords the individual depend upon his aesthetic sensitivity. As a basis for aesthetic sensitivity he needs to develop an awareness, creative understanding and appreciation of design, to acquire a core of knowledge concerning design, to achieve competence in making design judgments, and to establish a personal design philosophy.[1]

The student needs to understand the various principles, concepts and elements in design. Also, the student should develop an evaluative ability in applying the standards of design, acquire an interest in, and an appreciation of, good composition, and increase his understanding of the use of different media in the design of his environment.

If these understandings and abilities are the

---

[1]Course Outline for TRA 140: Matrix: Design for Living, Textiles, Clothing and Related Arts Department, Michigan State University, 1966 (in the files of the Department).

objectives of the course, an evaluation of student achievement of whether or not these objectives have been accomplished is needed.

The Textiles, Clothing and Related Arts Department offers Matrix: Design for Living, TRA 140, as a basic core course in the College of Home Economics, Michigan State University. The large section lecture course incorporates slides for illustration and discussion purposes. The objectives of the course are to develop in the students:

1. Awareness of the nature of design and its manifestation in life's matrix.
2. Knowledge about the design elements, principles and concepts.
3. Some competence in utilizing the basic design elements, principles and concepts in solving design problems and making design decisions.
4. Formulation of a value system and philosophy related to design.[1]

Student achievement of the second and third objectives is measured by objective tests, while the first and fourth objectives are measured by student essays.

During the past academic year the testing program of TRA 140 has been the subject of an Educational Development Program.[2] The project produced new testing methods and new test items and a revision of existing test items.

---

[1] Ibid.

[2] Educational Development Program, Provost's Office, Michigan State University.

This research is an outgrowth of the EDP project which is concerned with comparing testing methods.

The objective tests developed from new and improved items must prove valid and reliable for effectiveness; "Further research is needed on the development and refinement of tests of aesthetic judgment, especially in regard to the measurement of judgment in specific situations."[1] Elfreda Samuels, in her study concerning the construction of a test of design judgment, states:

> The need then seems to be for an instrument geared to comprehension of contemporary art education, . . . because it needs to be devised to test the types of art activities found in the average art classes of today.[2]

Justification

Evaluation, inevitable in education, normally takes the form of testing.[3] The testing of large numbers of students in the elusive areas of aesthetics and discrimination necessitates a search for valid new methods of evaluation

---

[1]Marilyn Joan Horn, "The Ability of College Students to Apply Principles in Concrete and Abstract Situations and Its Relation to Art Interest" (unpublished Master's thesis, Cornell University, 1953), p. 149.

[2]Elfreda C. Samuels, "The Construction of a Test of Design Judgment" (unpublished Master's thesis, Boston University, 1955), p. 7.

[3]Paul Dressel, Evaluation in Higher Education (Boston: Houghton-Mifflin Company, 1961), p. 160.

in these domains.[1]

Dressel states that conventional evaluation procedures dependent on words alone are inappropriate in attempting to measure intangible reactions.[2] Many art and design educators use primarily written tests based on the instructor's lecture material. They believe "objective" evaluation to be improper in the judgment of art. However, when there are large numbers of students in a class, subjective methods of evaluation become a practical impossibility.

Researchers have for a long time been developing measuring instruments for objective means of evaluation in the area of aesthetic judgment and appreciation.

> Educator, artist, and layman alike hold measurement in the fine arts to be a controversial issue with no scientific basis or truth on either side, or on any of the many sides of the problem.[3]

> The issue which seems basic to all the objections raised against scientific measurement in art stems from the idea that objectivity must necessarily involve an absolute standard, that such a standard measures conformity only and is therefore in contradiction to the true meaning of art.[4]

---

[1]Project Proposal: Course Development of TRA 140--Matrix: Design for Living, Textiles, Clothing and Related Arts Department, Michigan State University (in the files of the Department).

[2]Dressel, op. cit., p. 160.

[3]Julius Heller, "Changes in Art Judgment Resulting from Courses in Art Appreciation" (unpublished Doctoral dissertation, University of Southern California, 1948), p. 1.

[4]Peter A. Carmichael, "The Phantom of Critical Objectivity," Journal of Aesthetics, Vol. 9 (September, 1950), p. 13.

Such a standard could inhibit our individual responses to art and design. It would seem, however, that reliable design judgments could be made concerning the elements, principles and concepts of design and that construction of a basic set of standards for evaluating the use of design principles and elements should be possible. There also must be criteria by which to evaluate the function of a form or object and the techniques and materials used.

The visual method of evaluating students' design understandings has produced a controversy among educators. Munro states a criticism of tests using paired pictures selected by a group of experts:

> The usual effects of such tests is to penalize all deviation from adult, conventional norms of taste in that particular environment, since the student who prefers the "right" examples gets a high grade. The relativity of aesthetic values is ignored, no allowance being made for legitimate differences in taste and style, or for the fact that different art forms may be desirable under different circumstances.[1]

In a study concerning visual testing procedures, Curtis and Knopp reported that this method of test administration can yield a greater coverage of test content in a unit of time than can the normal mode of test presentation.[2]

---

[1]Thomas Munro, "Aesthetics as Science: Its Development in America," Journal of Aesthetics, Vol. 9 (March, 1951), p. 180.

[2]H. A. Curtis and Russell Knopp, "Experimental Analyses of Various Modes of Item Presentation on the Scores and Factorial Content of Tests Administered by Visual and Audio-Visual Means: A Program of Studies Basic to Television

Gropper found the importance of employing testing procedures closely related to teaching methods.[1] Because design theory is taught to the students in large sections of TRA 140 through the use of slides, the course committee believes that knowledge gained by this teaching method should be evaluated by using the same type of stimulus for testing.

For many areas of education, including art, Benjamin[2] lists several reasons for testing with visuals:

1. Dependence upon reading as a sole means of providing test stimuli is reduced.
2. Various parts of questions can be presented almost simultaneously, without the necessity for verbal buildups or descriptions.
3. It is easier to see relationships among various parts of data in questions.
4. Pictorial or graphic representations of things, events, or situations can be fairly lifelike, making it easier for students to see relationships between the posed problem, and actual

---

Testing," Department of Educational Research and Testing, School of Education, Florida State University, National Defense Education Act of 1958, pp. 78-79.

[1]George L. Gropper, "Learning from Visuals," Audio-Visual Communications Review, Department of Audio-Visual Instruction, Washington, D.C., Vol. 14, No. 1 (Spring, 1966), p. 47.

[2]Harold Benjamin, Audio-Visual Instruction Materials and Methods (New York: McGraw-Hill Book Company, 1959), p. 420.

application.

5. Variety is provided through pictorial, recorded, or dramatic elements in testing procedures, improving student attitudes toward testing.
6. Some students believe that evaluation situations which are not completely verbal are easier, thus heightening morale by thinking that students are better able to demonstrate their ability on such a test.
7. Aspects of objectives which cannot be measured at all by strictly verbal means may be able to be measured by employing visual materials.

Little is known, however, about the use of slides for testing. For the students' attainments of the course objectives for TRA 140, an experimental study should help determine the value of slides as a testing medium.

Focus of the Study

This research seeks to determine the relative effectiveness of verbal and slide test items in the evaluation of students' attainments of the course objectives of TRA 140. To compare verbal and slide test items, it was necessary to construct an objective test of design understandings and judgments composed of both verbal and slide items for use as an evaluative device in TRA 140.

The assumptions and hypothesis guiding the research are as follows:

Assumptions:

1. Both slide and verbal questions can be formulated covering the same basic course objectives in TRA 140.
2. The classifications level tested by each item can be determined according to Bloom.[1]

Hypothesis:

Slide questions will provide opportunities to effectively measure a higher level of intellectual skills and abilities, that of qualitative judgments, than will verbal items. Therefore, the use of visuals should prove to be a more discriminating procedure.

---

[1]Bloom, Krathwohl and others, Taxonomy of Educational Objectives, Handbook I: Cognitive Domain (New York: Longmans, Green and Co., 1956).

CHAPTER II

REVIEW OF LITERATURE

This chapter reviews the theoretical and pertinent literature and research pertaining to the problem of visual and verbal testing methods. Areas included are: (a) a comparison of visual and verbal test items, (b) related tests, and (c) related research.

Comparison of Visual and Verbal Test Items

As a result of testing by means of both visual and verbal items, Gropper discovered that, provided a visual lesson is suitably programmed, the student can answer both pictorially and verbally stated questions about conceptual phenomena.[1] The conditions permitting this suitable programming to occur appear to be those which facilitate discriminations about similarities and dissimilarities in visual situations.[2] Instructional settings which provide these conditions can aid in the understanding and subsequent

---

[1]George L. Gropper, "Why Is a Picture Worth a Thousand Words?" Audio-Visual Communications Review, Department of Audio-Visual Instruction, Washington, D.C., Vol. 11, No. 14 (July-August, 1963), p. 85.

[2]Ibid.

practice of generalized responses.[1]

Gropper discovered a statistically significant interaction between intelligent quotient and mode of stimulus presentation as measured by verbal test items only. While above average students profited more from verbal presentation, below average students benefited more from the visual presentation.[2]

A general expectation is that the greater the similarity between the learning situation and the testing situation, the greater would be the degree of transfer. As a result of measurement, Gropper found that the solely visual instruction led to superior performance on the visual test items, and the verbal lesson proved a more effective instructional experience for the relatively more abstract verbal items.[3] "While the verbal lesson did lead to successful performance on the verbal test items, it did not prove to be superior to the visual lesson in this regard."[4]

Experience with concrete visual examples in the visual lesson allowed for successful transfer either to concrete visual items or to abstract verbal test items, while concept acquisition based on a programmed verbal lesson

---

[1] Ibid.

[2] Gropper, "Learning from Visuals," op. cit., p. 45.

[3] Ibid., p. 46.

[4] Ibid.

appeared to have facilitated transfer less readily to the visual criterion test than to the verbal test. Gropper concluded on the basis of this difference in findings for visual and verbal lessons that, for transfer to occur, similarity between learning situation and testing situation may be less important when learning is based on visual materials.[1]

Gropper found by achievement testing "(a) that non-significant differences in total test scores between visual and verbal treatments were obtained and (b) that relatively high achievement levels were obtained for both treatments."[2] Total test scores showed no differences in the relative effectiveness of the visual and verbal presentations. However, differences in the relative effectiveness of the visual and verbal presentations were revealed by separate analysis based on scores on either visual or verbal test items.[3]

Related Tests

The McAdory Art Test of art appreciation was published in 1929. It contains pictures of 72 works of art which cover a wide variety of contemporary art forms, ranging from pictures of furniture and other functional objects to works of art in museums. Four versions of each art work

---

[1] Ibid., p. 47.

[2] Ibid., p. 44.

[3] Ibid.

are given, differing in shape, arrangement, shading and use of color. The person being tested is to rank the four versions in terms of his preferences. Its dependence on contemporary art values of 1929 produced a primary weakness of the test.[1] The test was validated by 100 judged ranging from department store workers to competent lay critics and art producers. Meier writes in the Mental Measurements Yearbook:

> . . . save for the possibility that time may outmode some of the prevailing standards on which both the scoring norms and the consensus were based, the test represents a definite achievement in providing a test of general art appreciation.[2]

The Meier Art Judgment Test uses the altered-version type of item for measuring art appreciation. It differs from the McAdory in that only one alternate version is given for each art work, and the examples concern relatively timeless art masterpieces.[3] Meier believes that a work of art can be judged on the basis of its organization through an understanding of the functioning of principles basic in all art.[4] Each example in his test contains some principle or

---

[1]Jum C. Nunally, Educational Measurement and Evaluation (New York: McGraw-Hill Book Company, Inc., 1964), p. 298.

[2]Oscar K. Buros (ed.), The Nineteen Forty Mental Measurements Yearbook (New Jersey: Mental Measurements Yearbook, 1941), p. 146.

[3]Nunally, loc. cit.

[4]Norman Charles Meier, The Meier Art Tests, Examiner's Manual, Bureau of Educational Research and Service (Iowa City: State University of Iowa, 1942), p. 7.

principles which have been singled out for manipulation in one version, so that the two versions presented are nearly identical, but with one having the functioning of the principle impaired. The test was originally known as the Meier-Seashore Art Judgment Test, published in 1929. Revised in 1949, it became the Meier Art Judgment Test.

The Graves Design Judgment Test measured certain components of aptitude for the appreciation or production of art structure. The test measures the degree to which a subject perceives and responds to the basic principles of aesthetic order--unity, dominance, variety, balance, continuity, symmetry, proportion, and rhythm.[1] The items consist entirely of abstract designs in an attempt to be as removed as possible from traditional and contemporary art values. Each item consists of two or three versions of the same basic design, the altered version or versions being constructed to violate a basic aesthetic principle. In a review of this test, Nunally states that the test is a useful measure but adds that only a small amount of empirical work has been done with the instrument.[2]

The Crow Picture Interpretation Test was published in 1926; its purposes were: (a) to measure the ability of students to look at pictures and give aesthetic and

---

[1]Maitland Graves, Design Judgment Test Manual (New York: The Psychological Corporation, 1948).

[2]Nunally, op. cit., p. 300.

thoughtful interpretations of them; (b) to create a wider interest in the study of good pictures in the public schools; (c) to aid teachers in understanding the difficulties of students in looking at pictures; (d) to enable teachers to measure progress of students in picture interpretation by determining standards for the various grades; (e) to enable both teachers and students to see more in pictures and get greater pleasure from them.[1]

The test consists of a booklet of questions and answers, and an envelope of eight copies of masterpieces. The questions are concerned with the pupil's interpretation of details, aesthetic responses to details, the meaning and beauty of the picture, and also points of contact between the pupil's experiences and the experience interpreted in the picture.[2]

Lewerenz's test in the Fundamental Abilities of Visual Art was constructed to enable teachers to measure students' capacities and skills. It is an easily administered and scored group test. The test has nine forms:

1. Recognition of color.
2. Observation of light and shade.

---

[1]Alfred S. Lewerenz, "A Critical Analysis of the Elemental Abilities Required in Art Education with a View to Possible Objective Measurement" (unpublished Master's thesis, University of Southern California, Los Angeles, 1927), p. 36.

[2]Ibid., p. 7.

3. Visual memory of proportion.

4. Originality in line drawing.

5. Recognition of proportion.

6, 7, and 8. Analysis of perspective.

9. Knowledge of subject matter.

The test was validated on the basis of what was taught in art courses in the Los Angeles schools. The student was to choose the best of four bowls, cornices, curves, composition of landscapes, or other design examples, and, in addition, he was to make ten original line drawings.

In the Nineteen Forty Mental Measurements Yearbook, Faulkner says this about the Lewerenz tests:

> It is of little value to those who believe that art is an integrated activity rather than a series of separate skills, nor is it of great value to those who believe that an approach to art through such general and abstract art elements as light and shade, color, and proportion is less desirable than through such specific fields as architecture, industrial art, and the like. Thus its value is highly dependent on one's philosophy and psychology of art.[1]

A Test for Art Appreciation by Karwoski and Christensen, published in 1926, includes 28 questions of three different forms. One form is the comparison of two examples, one good and the other poor. Five reasons are provided for choice under the paired pictures; the subject is to choose one reason. In the second form the subject judges a single picture and checks one of the five reasons for the preference.

---

[1]Buros, op. cit., p. 149.

The third form is concerned with selecting the best of four examples of similar subjects. The authors believe art appreciation can be tested by forcing the subject to give an opinion of why one art form is preferred over another. The pictures are in the areas of painting, architecture, sculpture, industrial arts, abstract design, and color.[1] A revised version in 1933 included the areas of automotive, flatware, furniture, and costume design.

Related Research

Johnson constructed and evaluated a test designed to determine the degree of intellectual and aesthetic response to painting. He was concerned with reactions to content, composition, color, line, form, and tone. The test consisted of 140 verbal items cast in multiple-choice form which referred to one of seven pictures selected for visualization of the factors being tested. The test proved a reliable measure of the concepts being measured and a valid instrument of the aptitude of art appreciation in a verbalized situation.[2]

Heller conducted an investigation to evaluate art

---

[1]Theodore Karwoski and Erwin Christensen, "A Test for Art Appreciation," Journal of Educational Psychology, Vol. 17 (March, 1926), pp. 187-194.

[2]Dana D. Johnson, "The Construction and Evaluation of a Test of Aesthetic Reactions and Understandings--Paintings" (unpublished Master's thesis, School of Education, Boston University, 1954).

judgment in courses of art appreciation at the university level with contemporary evaluative materials. He wished to discover to what extent art judgment can be measured at the university level, and to what extent it can be changed by instruction. A test instrument in art judgment composed of thirty-nine pairs of pictures was constructed. The pictures were selected according to the availability of the items pictured, their suitability for projection, and the actual situations in life which demand certain judgments to be made. The test form was one based on pairs of pictures which could be projected simultaneously on a large screen. The student was directed to choose the picture he preferred. Heller felt that his investigation indicated that art judgments can be measured, and that they can be changed by instruction.[1]

Horn was concerned with the relationships of abstract art principles, and their application in specific areas of design. She hypothesized that an understanding of the principle or abstraction does not assure an understanding of its application to the concrete or specific situation. In addition to support for her hypothesis, she also found that, conversely, an understanding of the specific does not lead to the development of an abstract art concept which can be applied equally well in a number of situations. Two visual

---

[1]Heller, op. cit.

art tests and an art interest inventory were given to 275 college students. One of the visual tests used was the Graves Design Judgment Test for measuring application of design principles to abstract forms. The other test was a combination of the Brief Form of the McAdory Art Test with some plates from other selected art tests. The latter was designed to measure the application of design principles to concrete forms. Horn found that the ability to make art judgments can be improved with training; however, training appears to be more successful in increasing knowledge and understanding of art principles in regard to abstract forms than to concrete forms in the specific areas of design.[1]

Samuels constructed a test of design judgment for junior high school pupils in Framingham, Massachusetts. One purpose of the test was to measure the students' abilities to exercise good art judgment in the field of design. The test included 22 abstract designs, each illustrating one or more basic art principles and the use of art elements. The test was thought to be valid, but was very low in reliability. It was felt that the test should be lengthened and administered to a higher age group. Each test item consisted of two designs, one better than the other. The student was to select the better of the two.[2]

---

[1]Horn, op. cit.

[2]Samuels, op. cit.

CHAPTER III

METHODOLOGY

Construction of the Test

Development and sources of the items

When the experimental design for this study was developed, it was assumed that both verbal and slide items could be formulated covering the same basic understandings taught in TRA 140. The first step was to devise machine scorable items to test student understanding of all areas of design theory included in the course. The subject matter areas included in this course are:

A. Design elements
    1. Form
    2. Color
    3. Texture
B. Design principles
    1. Balance
    2. Emphasis
    3. Proportion and scale
    4. Rhythm
    5. Harmony--unity
C. Design concepts
    1. Criteria for design judgment

2. Design and culture

3. Timelessness and obsolescence

4. Materials and techniques

5. Form and function

6. Enrichment

7. Originality, expression and beauty

A study group[1] was formed to improve test items and methods for teaching TRA 140. This group was responsible for writing, reviewing, revising, and rejecting test items for the course.

In order to compare verbal items and slide items, it was necessary to construct items in pairs of one of each type of item. In writing paired slide and verbal items, the general procedure was to determine the kinds of questions which could be used to test the different course objectives. The next step was to select pictures or objects for photographing about which an item or items could be formulated concerning the course objective or subject matter area under consideration. All slide possibilities were presented to the study group for discussion and analysis. If a slide item was accepted by the group, a verbal item was written which would be parallel to the particular slide item in subject matter and difficulty.

---

[1]Mrs. Lorraine Gross, instructor of TRA 140, Dr. Mary Alice Burmester, representative for the Educational Development Project, Miss Louise Starr, graduate teaching assistant for TRA 140, and the writer.

The majority of slide items were cast in key form, the answer being selected from a key list of possible answers. The key for the paired verbal items was generally the same as its corresponding slide item. An illustration of such paired items is:

Verbal item:

"Yellow and orange."

Slide item (see Figure 1, page 22):

a picture of a yellow and orange object.

The answer for each item was to be selected from the following key:

1. Monochromatic
2. Complementary
3. Analogous
4. Triadic
5. None of the above

Occasionally the paired slide and verbal items were not cast in the same form. This occurred with some slide items which required the student to make a judgment. An example of such a pair is:

Verbal item:

Which of the following most violates the principle of emphasis?

1. Pale yellow wall with a red-orange and blue chair.
2. Green walls with white cabinets.
3. Large black stars on a white floor in a bathroom.
4. Bright pillows on a white couch.

Slide item (see Figure 2, page 23):

Figure 1.--Slide item illustration for the final examination of TRA 140, spring term, 1966.

(Note that the colors in this reproduction are not comparable to the slide used which showed a much brighter yellow and orange, and an almost non-existent blue.)

Figure 2.--Slide item illustration using a slide pair for the final examination of TRA 140, spring term, 1966.

A

B

Key A: Which is a more successful design?	Key B: Which principle was the major influence on your decision?
1. A	1. Rhythm
2. B	2. Balance
3. A and B are equal	3. Emphasis
	4. Proportion
	5. Harmony

This slide item consisted of a pair of slides. Slide A displayed a small kitchen with simple lines, rather plain but with large black stars on a white floor. Slide B displayed a plain modern kitchen of more acceptable design.

The paired slides used for comparison in this study are the result of an attempt to allow only one element, principle, or concept to vary while all others remain constant. It should also be noted that the slides chosen for paired comparison are thought to be of relative artistic value, and should not be considered examples of good and bad design.

A series of nine verbal items was included which concerned knowledge about factors valued in judging the successful achievement of a principle or concept. These items could each constitute a pair with any slide item requiring the student to make a judgment concerning the same principle or concept. The following is a verbal item of this type:

> Which of the following factors is NOT important in judging the achievement of a successful relationship between form and function?
>
> 1. Is the form the best fulfillment of the function?

2. Is the form designed to express primarily the aesthetic function?
3. Is the form objectively designed to express the function?
4. Does the form express the beauty inherent in a truly functional object?
5. All of the above are important.

Preparation of slides

Pictures and objects to be used for slide items were photographed and developed into 2" x 2" slides. Special care was taken when photographing materials to be used in slide pairs to see that the objects in question would be the same size and location in relation to the total picture.

Pretest of items

Items were pretested during Winter Term, 1966. Because TRA 140 includes two one-hour mid-term examinations and a two-hour final examination, the writing and pretesting of items was accomplished in three parts. A total of 91 slide items and 75 verbal items were pretested with some of each type appearing on two mid-term examinations and the final examination. During this term a total of 300 students were enrolled in two sections of TRA 140. Two forms of each examination were prepared so that students seated next to each other had alternate forms.

Item analysis

The data analysis for each item was tabulated by machine at the Scoring Office of the Evaluation Services, Michigan State University.

Because two test forms were used, two item analyses were provided for each item from the Scoring Office. These two analyses were averaged together to obtain one number for the index of difficulty and one number for the index of discrimination for each item. The information obtained from item analysis is helpful in indicating which items should be retained in their original version, discarded, or revised before further use. Subsequent tests comprising items selected on the basis of item analysis can be improved even more substantially.[1]

The index of difficulty is an item statistic which gives the proportion of total students taking the test who missed the item. Thus a low index indicates an easy item. For achievement tests, most test constructors desire items with difficulty indices from 20 to 80, with an average index of 50 to 60. Actually, opinions of test theorists are varied as to whether in general the items should all be of about 50 per cent difficulty or whether there should be a fairly wide range of difficulty, with an average of near 50 per cent.[2] The index of difficulty helps the test constructor determine whether the difficulty level of the items is suited to the group for which the test is intended.

---

[1]Dorothy Adkins Wood, Test Construction, Development and Interpretation of Achievement Tests (Columbus, Ohio: Charles E. Merrill Books, Inc., 1960), p. 92.

[2]Ibid., p. 82.

A discrimination index concerns the degree to which a single item separates the superior from the inferior individuals in the quality being measured.[1] The index of discrimination is the difference between the proportion of the upper 27 per cent of students who answered the item correctly and the proportion of the lower 27 per cent who answered the item correctly. This index depends upon the difficulty of an item. "Most methods of discrimination are of little practical value because they do not control the difficulty level effect."[2]

An index of discrimination for each item was computed according to the Flanagan method. He has developed a chart by which an index of discrimination may be computed which is not affected by the index of difficulty.

> Items of 50 per cent difficulty tend to provide the most valid test. Therefore a method which combines a rough measure of validity or discrimination value with a device which will favor items of 50 per cent difficulty will tend to appear to be superior to a method which provides a more valid index of item validity unaffected by difficulty.[3]

As a practical rule, the higher the discrimination,

---

[1]Kenneth L. Bean, Construction of Educational and Personnel Tests (New York: McGraw-Hill Book Company, 1953), p. 153.

[2]John C. Flanagan, "General Considerations in the Selection of Test Items and a Short Method of Estimating the Product-Moment Coefficient from Data at the Tails of the Distribution," The Journal of Educational Psychology, Vol. 30 (December, 1939), p. 675.

[3]Ibid., p. 676.

the better the test item.

Selection of final test items

The items, accompanied by the data from the pretest, were again presented to the study group to be analyzed and discussed. Items which appeared to have been ambiguous to students were revised. Some items contained foils which did not function; these options were also revised. Certain items were rejected for a number of reasons; for example, there may have been too many questions on a particular area of the subject matter, or an item may not have had a functioning paired item. Some items were rejected because analysis showed a very low index of difficulty and index of discrimination. However, items were not rejected on the basis of these two factors alone. Some items known to be relatively easy for the group were retained if they were discriminating so that the final group of items would test all subject matter areas of TRA 140. A total of sixty slide items and sixty verbal items were retained for the final test with which this research is concerned.

Classification Levels of Items

Because it was hypothesized that slide items would more effectively measure a higher level of intellectual skills and abilities than verbal items, each item for use on the final examination was assigned a classification level

according to the Taxonomy of Educational Objectives.[1] Bloom and his co-workers have attempted to take the vagueness out of statements of educational objectives at the college level by including illustrative test items to accompany each description of a subclass of objectives. The taxonomy provides no directive guidance as to objectives of higher or lower priority. It does arrange the classifications from the simple to the more complex behaviors, and from the concrete to the more abstract.[2]

Administration of the Test

The test was administered as the final examination in TRA 140, Spring Term, 1966. Ninety-four students took the two-hour examination. The slides were shown near the middle of the total time period, with approximately one minute allowed for each slide item.

Data Analysis of Final Items

Data analysis was computed by the Scoring Office for the total verbal items, the total slide items, and the total verbal items and total slide items of each classification level.

---

[1] Bloom, op. cit.

[2] Ibid., p. 31.

Determination of the Quality of the Test

Validity and reliability

A test is valid insofar as it measures the qualities we wish to measure. It is reliable insofar as it measures with precision.

> Only to the extent that a test measures something accurately can it measure it validly. Reliability is important only as a necessary condition for a measure to have validity.
>
> However, the converse is not true. Validity is something over and beyond mere accuracy of measurement. A test can measure with the greatest precision and still not be valid for our purposes.[1]

Validity of this test was first based on the agreement of the study group concerning the selection of items which would test for student achievement of the objectives of TRA 140. Further validity was determined upon consideration of the Item Analysis obtained from the Scoring Office.

Although an absolute minimum for the reliability of a measurement procedure cannot be set, an indication of the level of reliability that is required to achieve specified levels of accuracy in describing an individual or a group is possible.[2] "A test with relatively low reliability will permit us to make useful studies of and draw accurate conclusions about groups."[3] All things being equal,

---

[1]Robert Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education (New York: John Wiley and Sons, 1961), p. 185.

[2]Ibid., p. 189.

[3]Ibid., p. 190.

a test with a higher reliability coefficient is preferable.

Reliability for the test was determined by the Kuder-Richardson reliability procedure, which provides a measure of inter-item consistency from a single administration of the test. It is based on an analysis of the performance of each item, rather than requiring a split-half analysis. Inter-item consistency or test reliability is concerned with the consistency of the subjects' responses to all items in the test. It provides a measure of both equivalence and homogeneity. The assumption in the procedure is that the items within one form of a test have as much in common with one another as do the items in that one form with the corresponding items in a parallel or equivalent form. This means that the items are homogeneous in the sense that every item measures the same general factors of ability as do the others. Thus there tends to be a limiting factor in the analysis as one purpose of the items in the TRA 140 examination is to measure student achievement on many levels of ability.

Factor Analysis of Possible Influence on Test Results

Correlation coefficients were computed to determine the relationship of verbal items and slide items with both reading scores and CQT scores. A correlation coefficient is useful in describing the accuracy with which a test score

predicts some other factor.[1] This statistic may range in value from 1.00 to -1.00. A correlation of 1.00 indicates a perfect positive relationship between two variables; 0 indicates no relationship whatsoever; and -1.00 indicates a perfect negative relationship.[2] The product-moment method for computing a correlation coefficient was used.[3]

Reading scores

Because it was anticipated that those students who were more proficient readers might have an advantage over the other students concerning the verbal items, correlation coefficients were computed to find the relationship between verbal scores and reading scores, and also between slide scores and reading scores. The reading scores were obtained from performances on a reading test given to students upon entering Michigan State University.[4]

CQT scores

College Qualification Tests[5] are a series of ability

---

[1]Thorndike, op. cit., p. 121.

[2]Allen L. Edwards, Statistical Methods for the Behavioral Sciences (New York: Holt, Rinehart and Winston, 1964), p. 145.

[3]Thorndike, op. cit., p. 567.

[4]Reading Test Form A62, Office of Evaluation Services, Michigan State University.

[5]College Qualification Tests Manual, The Psychological Corporation, New York, 1957.

tests for college admission, placement and guidance. They were developed as predictors of success in college courses. The CQT have been found to be favorably valid and reliable.[1] Correlation coefficients were computed to determine the relationship between both the CQT scores and verbal scores and the CQT scores and slide scores.

Student interviews and questionnaire

It was anticipated that students' attitudes toward slide items in particular might have an influence on the results of the examination. Twelve students who had taken TRA 140, Winter Term, 1966, were selected for interviews. The method of selection was based on obtaining a sampling of students from all levels of achievement on the final examination, Winter Term, 1966. The results from these student interviews were analyzed and from the student attitudes expressed, statements were formed and arranged in a questionnaire administered to 83 students enrolled in TRA 140, Spring Term, 1966. Care was taken that the statements appearing on the questionnaire did not favor either verbal or slide items. The purpose of the questionnaire was only to enlighten the writer on student attitudes, and the results reported in this thesis concern those statements which were felt to reveal particularly salient attitudes.[2]

---

[1] Ibid., pp. 25 and 28.

[2] A copy of the interview questions and the questionnaire may be found in Appendices A and B.

CHAPTER IV

FINDINGS AND INTERPRETATIONS

The findings reported in this chapter are concerned primarily with the results of the test constructed and administered as the final examination in TRA 140, Spring Term, 1966. The examination was composed of 120 items of which 60 were verbal items and 60 were slide items. The results of both the total verbal items and the total slide items in the examination are summarized in Table 1.

Table 1. Summary data item analysis for indices of difficulty and discrimination and reliability for the TRA 140 final examination, Spring Term, 1966

	Mean Index of Difficulty	Mean Index of Discrimination	Mean Flanagan Discrimination	Kuder-Richardson Reliability
Total items on the examination (120 items)	32	18	26	.5056
Total verbal items (60 items)	29	19	28	.6064
Total slide items (60 items)	35	17	24	.4048

The total examination showed a mean index of

difficulty of 32. As stated in the preceding chapter, most test constructors prefer a mean difficulty index of 50 to 60. If this standard is accepted, then there is evidence that the examination as a whole was not sufficiently difficult for the group tested. It must also be recognized that the low difficulty must in part be a reflection of the fact that many of the very easy questions were paired verbal and slide items and thus appeared twice on the examination. Had the primary purpose of this test been a higher level of difficulty, these items would have been disregarded as a result of pretesting. However, the questions were retained for three reasons: (a) they were testing student attainment of various objectives of TRA 140, (b) they were acceptable in helping to fulfill the need to compare verbal and slide testing methods, and (c) the study group felt that easy items which were discriminating were valid testing tools.

The mean discrimination index for the total examination was 18. The index of discrimination represents the difference between the percentage of high achieving students who marked the items correctly and the percentage of low achieving students who marked the items correctly, and when determined by the process used by the Scoring Office, a maximum discrimination index is 100 with a minimum index of -100. It is, however, unusual for this index to exceed 70 with the range generally between -20 and 50.[1] Thus the

---

[1]University of Minnesota Classroom Teaching Bulletin (mimeographed).

mean index of discrimination of 18 is average. When computed by the Flanagan method, which is not affected by the index of difficulty, it rises somewhat to an average discrimination index of 26.

The Kuder-Richardson reliability coefficient for the total examination was computed to be .5056. Coefficients generally desirable for reliability usually fall in the .80's and .90's.[1] Thorndike advises that a minimum cannot be set for the reliability of a measuring instrument.[2] The primary interest is in obtaining as high a coefficient as possible. Therefore, if a choice of instruments meeting the needs is available, the one with the higher reliability coefficient will be selected. This coefficient is therefore a relative measure and it is difficult to assess the value of a single coefficient.

Comparison of Total Verbal and Slide Items

Table 1 indicates that the total slide items were harder for the students than were the verbal items. This result is consistent with student attitude in TRA 140 as assessed in a student questionnaire. The responses indicated that 82 per cent agreed with the statement "slide

---

[1]Anne Anastasi, Psychological Testing (New York: Macmillan Book Company, 1954), p. 105.

[2]Thorndike, op. cit., p. 189.

questions are harder than verbal questions."[1] However, negative to the hypothesis of this study, verbal items attained a slightly higher mean index of discrimination.

As a group, verbal items were considerably more reliable than slide items. However, although reliability is of importance to the testing program of TRA 140, item discrimination is the primary consideration. In item discrimination the results of the total verbal items proved preferable.

Comparison of Slide Items and Verbal Items Concerning Specific Levels of Classification of Intellectual Abilities

Paired items did not necessarily fall into the same classification level. Each item was classified at its highest possible level.

Classification Level 1.00--involves the recall of specifics and universals. This classification refers to the mere bringing to mind of the appropriate material; the process of remembering.[2]

The test included only verbal items within this level. Because it was felt to be unwise to use the same slides for testing as were used for the lectures, this necessitated some alteration of the material from the original

---

[1]Student questionnaire. See Appendix B.

[2]Bloom, op. cit., p. 201.

form in lecture, and therefore such items were not to be classified at the 1.00 level.

Table 2 indicates that those 17 verbal items occurring at the 1.00 level obtained an index of difficulty of 26 and an index of discrimination of 30. It is interesting to observe that these items show the highest reliability of all levels tested, each level testing homogeneous knowledge.

Classification Level 2.00--is known as Comprehension.[1] This level refers to the lowest level of understanding, that level at which the individual knows what is being communicated and can use this knowledge without relating it to other material.[2] Seventeen slide items and fourteen verbal items dealt with the comprehension level of understanding. The slide items ranked at a considerably higher level of difficulty than the verbal items; however, the verbal items ranked a higher index of discrimination and were much more reliable.

Classification Level 3.00--is concerned with application of abstractions. These may be in the form of principles, ideas, theories, or methods, which must be remembered and

---

[1] Ibid., p. 204.

[2] Ibid.

Table 2. Summary data item analysis for each classification level of items for the TRA 140 final examination, spring, 1966

	Verbal Items				Slide Items			
Classification Level	Number of Items	Mean Index of Difficulty	Mean Index of Discrimination	Kuder-Richardson Reliability	Number of Items	Mean Index of Difficulty	Mean Index of Discrimination	Kuder-Richardson Reliability
1.00	17	26	30	.4694				
2.00	14	22	33	.4543	17	35	26	.1830
3.00	13	29	31	.3069	23	35	25	.2060
6.00	16	39	23	.0131	20	37	29	.4396

applied.[1] Thirteen verbal items and twenty-three slide items fell within this level. Again the slide items were more difficult, but less discriminating and less reliable than the verbal items, although the differences were not as great as at the 2.00 level.

Classification Level 6.00--involves judgments concerning the value of material and methods for specific purposes; the use of a standard of appraisal.[2] The results of this level are opposite to those at levels 2.00 and 3.00. The twenty slide items concerned were slightly less difficult than the sixteen verbal items, but the slide items were more discriminating and considerably more reliable. The hypothesis for this research is here supported in finding a higher discrimination with the slide items at this higher classification level. It was also found that at this level the writing of slide items was less difficult than the writing of verbal items. The students agreed in the questionnaire that, were slide questions eliminated, it would not be possible to test all information taught in TRA 140 (see Appendix B). The writer feels that this attitude would be reflected to the greatest extent at this classification level.

---

[1] Ibid., p. 205.

[2] Ibid., p. 207.

Other Factors of Possible Influence on the Test Results

Reading scores

The correlation coefficients indicating the relationship of reading scores with both verbal and slide items are .343 and .397 respectively, as can be seen in Table 3.

Although a positive relationship is present in each case, the coefficients are relatively low. There is no significant difference between the two correlations and thus it cannot be predicted that a student with a high score on the reading test will obtain a higher score on either the verbal or slide items.

Table 3. Correlation coefficients of the individual total verbal item scores and total slide item scores obtained by students in TRA 140 with their reading scores and CQT scores

Variables	Correlation Coefficients
Verbal scores and reading scores	.343
Slide scores and reading scores	.397
Verbal scores and CQT scores	.392
Slide scores and CQT scores	.671

CQT scores

The correlation coefficients between CQT scores and verbal and slide scores are .392 and .671, respectively. This would indicate that it may be predicted that a student obtaining a high score on the CQT would be more apt to

consistently obtain a high score on the slide items than on the verbal items. The hypothesis is in agreement with this result, especially at the 6.00 classification level as it was anticipated that slide items would more effectively measure learning of a higher ability level.

Student questionnaire

Certain attitudes expressed by the students in answer to the questionnaire have been included in this thesis where pertinent. However, additional results of interest may be noted.

Although the data analysis results indicated that verbal items are favorable except at the highest classification level, student responses on the questionnaire would indicate that slide items do indeed occupy a significant role. One such response was that, were slide items eliminated, it would not be possible to test all information taught in TRA 140.

In addition, nearly all of the students felt that the expectation of slide items on an examination caused them to be more attentive in class. They did, however, feel that there may be more than one correct answer to a slide question, and that personal opinion was a factor. These attitudes, although not universal, do perhaps indicate that slide items should be used more than just at the highest classification level. They also are indicative of the fact that great care must be taken in the structuring of slide questions.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

Summary

One method of evaluation used for the TRA 140 course at Michigan State University is objective examination. Because the course is taught with the aid of slides, it has been of interest to know whether objective testing is more effective with the use of slide items. Therefore, the purpose of this research was to determine whether slide items or verbal items are more effective as testing tools for the TRA 140 course. The hypothesis that slide items would more effectively measure a higher level of intellectual skill and would be more discriminating than verbal items was supported in that slide items were more discriminating at the highest classification level.

A study group was formed in connection with an Educational Development Project. The group wrote, analyzed and revised test items. The items were pretested, and many were eliminated so that sixty verbal items and sixty slide items remained on the final examination. The majority of these items were paired in subject matter and difficulty.

Validity of the examination was based on the face validity of the items, and also on the results of the item

analysis. The Kuder-Richardson method of inter-item consistency was used as a measure of reliability. An item by item analysis including difficulty and discrimination was used for interpreting the results of the examination. In addition, the Flanagan method was used to compute another index of discrimination for each item. The results were tabulated so that information was available on the total verbal items, the total slide items, and the total verbal and total slide items at each classification level.

Because the effect of some other factors was anticipated, reading scores and CQT scores were obtained for each student taking the examination. A coefficient of correlation was then computed between each of these scores and the total verbal scores and the total slide scores for each student. Twelve interviews with students who had taken TRA 140 the preceding term were used as a basis for the preparation of a questionnaire answered by the students taking the course Spring Term, 1966. The questionnaire was designed to determine student attitudes toward verbal and slide items.

The findings revealed that the total slide items were more difficult for the student than the total verbal items; however, the verbal items were slightly more discriminating and more reliable than the slide items. Therefore, it would appear preferable to use verbal items whenever possible.

An analysis of the items of each classification

level indicated that verbal items are probably better to use at the lower classification levels because they were more reliable and more discriminating than slide items. However, these results were reversed at the highest classification level and thus slide items appear preferable at this level.

Because the questionnaire revealed strong student attitudes that slide items occupy a position of value, slide items should possibly be used more than only at the highest classification level.

The highest correlation coefficient was between the CQT scores and the total slide scores. The indication is that those students obtaining the higher CQT scores should score above the others on the slide items. Thus, to this extent, slide items give the advantage to the students with greater ability.

Recommendations for Further Research

Repeat the testing using two matched groups of students and administering the slide items to one group and the verbal items to the other group.

Administer the total examination to two groups, alternating the sequence of verbal and slide items.

Conduct further research to discover which students excel on verbal items and which students excel on slide items and the causes for the excellence.

BIBLIOGRAPHY

Anastasi, Anne. Psychological Testing. New York: Macmillan Book Co., 1954.

Bean, Kenneth L. Construction of Educational and Personnel Tests. New York: McGraw-Hill Book Company, Inc., 1953.

Benjamin, Harold. A V Instruction Materials and Methods. New York: McGraw-Hill Book Company, Inc., 1959.

Bloom, Krathwohl, et al. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: Longmans, Green and Co., 1956.

Buros, Oscar K., Editor. The Nineteen Forty Mental Measurements Yearbook. New Jersey: Mental Measurements Yearbook, 1941.

Carmichael, Peter A. "The Phantom of Critical Objectivity," Journal of Aesthetics, Vol. 9, Sept., 1950.

College Qualification Tests Manual. New York: The Psychological Corporation, 1957.

Course Outline, for TRA 140: Matrix: Design for Living, Textiles, Clothing and Related Arts Department, Michigan State University, 1966 (in the files of the Department).

Curtis, H. A., and Knopp, Russell. "Experimental Analyses of Various Modes of Item Presentation on the Scores and Factorial Content of Tests Administered by Visual and Audio-Visual Means: A Program of Studies Basic to Television Testing." Department of Educational Research and Testing, School of Education, Florida State University, National Defense Education Act of 1958.

Dressel, Paul. Evaluation in Higher Education. Boston: Houghton-Mifflin Company, 1961.

Educational Development Program, Provost's Office, Michigan State University.

Flanagan, John C. "General Considerations in the Selection of Test Items and a Short Method of Estimating the Distribution," The Journal of Educational Psychology, Dec., 1939.

Graves, Maitland. Design Judgment Test Manual. New York: The Psychological Corporation, 1948.

Gropper, George L. "Learning from Visuals," A V Communications Review, Department of Audio-Visual Instruction, Washington, D.C., Vol. 14, No. 1, Spring, 1966.

Gropper, George L. "Why Is a Picture Worth a Thousand Words?" A V Communications Review, Department of Audio-Visual Instruction, Washington, D.C., Vol. 11, No. 14, July-August, 1963.

Heller, Julius. "Changes in Art Judgment Resulting from Courses in Art Appreciation," Dissertation, University of Southern California, 1948.

Horn, Marilyn Joan. "The Ability of College Students to Apply Principles in Concrete and Abstract Situations and Its Relation to Art Interest," Thesis, Cornell University, 1953.

Johnson, Dana D. "The Construction and Evaluation of a Test of Aesthetic Reactions and Understandings--Paintings," Thesis, School of Education, Boston University, 1954.

Karwoski, Theodore, and Christensen, Erwin. "A Test for Art Appreciation," Journal of Educational Psychology, Vol. 17, March, 1926, 187-194.

Lewerenz, Alfred S. "A Critical Analysis of the Elemental Abilities Required in Art Education with a View to Possible Objective Measurement," Thesis, University of Southern California, Los Angeles, California, 1927.

Meier, Norman Charles. The Meier Art Tests, Examiner's Manual. Iowa City: Bureau of Educational Research and Service, State University of Iowa, 1942.

Munro, Thomas. "Aesthetics as Science: Its Development in America," Journal of Aesthetics, Vol. 9, March, 1951.

Nunally, Jum C. Educational Measurement and Evaluation. New York: McGraw-Hill Book Company, Inc., 1964.

Project Proposal: Course Development of TRA 140--Matrix: Design for Living, Textiles, Clothing and Related Arts Department, Michigan State University (in the files of the Department).

Reading Test Form A62. Office of Evaluation Services, Michigan State University.

Samuels, Elfreda C. "The Construction of a Test of Design Judgment," Thesis, Boston University, 1955.

Thorndike, Robert, and Hagen, Elizabeth. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, 1961.

University of Minnesota Classroom Teaching Bulletin.

Wood, Dorothy Adkins. Test Construction, Development and Interpretation of Achievement Tests. Columbus, Ohio: Charles E. Merrill Books, Inc., 1960.

APPENDICES

APPENDIX A

INTERVIEW SCHEDULE

(Questions Presented to Students in Interviews)

1. Were slide questions easier or harder than verbal questions? Why?

2. Did you feel that the slide questions were objectively selected?

3. Did you feel able to make a definite choice when asked to choose which was the better of a pair of slide examples?

APPENDIX B

Name_____ Student No._____

Key: 1. Strongly agreed
2. Agreed
3. Disagreed
4. Strongly disagreed

1. I think that slide questions are harder than verbal questions.

2. When judging paired slides, I find it difficult to choose which one is better.

3. It is difficult for me to picture what is being asked in a verbal question.

4. I like slide questions better because I can actually see what the question is asking.

5. I think that there may be more than one correct answer to a slide question because people see things differently.

6. I think that slide questions are ambiguous.

7. Slide questions are of value to me because I feel that it is more important that I can see the design qualities than to talk about them abstractly.

8. I think that verbal questions are harder than slide questions.

9. I feel that slide questions are expressive of the instructor's values and opinions.

10. I find that I must study differently for an exam when I know that slide questions will be asked.

11. I find that expecting slide questions on an exam causes me to be more attentive to the slides shown in lectures.

12. If slide questions were eliminated, I think that it would still be possible to test all information taught in TRA 140.

13. I feel that the correct answers to the slide questions are based on design criteria and would be selected by other people trained in design.