EVALUATION AND REVISION OF SELECTED TESTS FOR A BASIC CLOTHING CONSTRUCTION COURSE

By

CYNTHIA JOY KALLENBERGER II

Bachelor of Science in Home Economics

Oklahoma State University

Stillwater, Oklahoma

1980

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
December, 1986

Thesis
1986
K14e
Cop.2

OKLAHOMA STATE UNIVERSITY LIBRARY

EVALUATION AND REVISION OF SELECTED TESTS FOR

A BASIC CLOTHING CONSTRUCTION COURSE

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

1263878

ACKNOWLEDGMENTS

I wish to express my appreciation to all those who have provided help throughout this study. A special thank you goes to my major adviser, Dr. Lynn Sisler, for her continued guidance, advice, and encouragement. Appreciation is extended to committee members, Dr. Tana Stufflebean and Dr. Elaine Jorgenson, for their contributions. Thanks is also extended to the graduate students who reviewed the item banks and made suggestions for improvement and to Mary Lou Wheeler for typing the final draft.

A special thanks goes to my parents, Mr. and Mrs. Robert H. Kallenberger, for their continued support, encouragement and understanding and for their assistance in coding the student tests onto the computer answer sheets.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

At Oklahoma State University all students majoring in clothing, textiles and merchandising and home economics education and community services are required to take CTM 1100, Clothing Construction: Processes and Products which is a modular class consisting of six modules, each worth one credit hour. Modules are as follows:

1. basic construction techniques,
2. pattern selection and garment construction,
3. selecting quality ready-to-wear,
4. pattern alteration and fitting,
5. couture techniques and problem fabrics, and
6. construction of designer garment.

Students enroll in 1 to 3 credit hours per semester and may take a total of six credit hours. To meet the requirements for each degree, students majoring in apparel design are required to complete modules 1, 2, 3, 4, 5, and 6; apparel merchandising majors must complete 1 and 3 and either 2, 4, or 5; and home economics education and community service majors must complete modules 1, 2, and 4.

After the completion of each module a 50 item test is given to evaluate what the student has learned. The test consists of multiple-choice, matching, and true-false items. As with any course, the tests

need to be revised periodically in order to up-date and meet the changing requirements for the course.

Purpose and Objectives

The purpose of this study was to evaluate tests for Modules 1 and 2 of the clothing construction course to determine needed revisions. Specific objectives were:

1. To determine item difficulty by using past tests,
2. To determine item discriminating power by using past tests,
3. To develop a table of specifications for test items for Module 1 and Module 2, and
4. To develop an item bank of test questions to be used in developing a final test for each module.

Definition of Terms

Definitions of terms used throughout the study are listed as follows:

> Evaluation - the systematic documentation of consequences [results or effects] of programs [curriculums] and the determination of their worth [merit] in order to make decisions about them (Green and Stone, 1977, p. 4).

Test - "an instrument or systematic procedure for measuring a sample of behavior" (Gronlund, 1985, p. 5).

> Measurement - the process of obtaining a numerical description of the degree to which an individual possesses a particular characteristic (Gronlund, 1985, p. 5).

Module - a unit of study which includes objectives, instructions, and textbook references necessary for the completion of the unit.

Table of Specifications - a two-way chart which relates

instructional objectives to the course content and specifies the emphasis to be given to each learning outcome (Gronlund, 1985).

CHAPTER II

REVIEW OF LITERATURE

According to Davis (1980, p. 12) the main goal of evaluation is "to make available the best [most accurate, most useful] information for improving understanding and facilitating decision-making." When evaluating, Green and Stone (1977) have suggested that a teacher needs to first determine the purpose for conducting the evaluation. In curriculum evaluation the evaluator must determine whether the program is meeting its intended purposes, goals, and objectives. It should focus on the unexpected, as well as the expected outcomes. Green and Stone (1977) also stated that the key element is the evaluator. The evaluator must know what to look for, where and how to look for it, how to know he/she is seeing it and how to appraise and report the results.

Evaluation can be quantitative (measurements) and/or qualitative (nonmeasurements). Evaluation always includes value judgments of how desirable the results are (Gronlund, 1985).

Gronlund (1985) listed five general principles of evaluation:

1. A priority of evaluation is to determine and clarify what to evaluate.

2. Evaluation techniques should be selected according to the intended purpose.

3. A variety of techniques is required in comprehensive evaluation.

4. An awareness of the limitations and strengths of evaluation techniques is required for proper use.

5. Evaluation is not an end, but a means to an end.

Hopkins and Antes (1978, p. 86) defined a test as "an instrument, device or procedure which proposes a sequence of tasks to which a student is to respond - the results of which are used as measures of a specified trait." They also indicated that by assigning a meaningful number value to the characteristic the test is measuring, that characteristic can be quantified. This is done for each student taking the test and is the primary purpose of a classroom test. Tests may be used to make decisions about pupils, instruction methods or teachers (Nunnally, 1964).

Gronlund (1985) identified four basic types of tests, each with its own purpose. Placement tests are usually pretests to determine if a student has the needed prerequisite skills for a given course or to determine the extent to which a student has previously achieved the objectives of a course. Formative tests are given as a method of monitoring the progress of a student's learning. Usually given periodically, throughout the course of instruction, the tests indicate the strengths and weaknesses of the student's learning. Diagnostic tests help to diagnose areas where a student is having learning difficulty. The summative test is generally given at the end of the course. It is used for assigning grades or evaluating course effectiveness.

Constructing Test Questions

When planning a test Gronlund (1985) listed four factors to be considered:

1. Determine the purpose of the test.
2. Develop test specifications.
3. Select test items that are appropriate.
4. Prepare test items which are relevant.

Another important aspect of preparing a test is knowing how to select and arrange the test items. Hall and Paolucci (1970) gave ten suggestions:

1. To maintain a student's interest and to permit individuals to show how well they can perform on a specific type of item, a major test should be designed with more than one type of item.
2. Restrict a test to no more than three types of items to prevent confusion over directions.
3. Provide encouragement by progressing from the simplest to the most difficult within each group of items.
4. Group all items of the same type together.
5. Arrange test items randomly to avoid responses which fall into a regular sequence.
6. Each item should be independent of the others so as not to allow answers to be obtained from another question.
7. Concise, clear, and complete directions should be included for each section of the test. Students should know what is expected, how to proceed, and where to write responses.
8. All parts of a question should be placed on the same page.
9. Responses should be made simple and convenient. Often this is done through providing blanks beside the item numbers.
10. Prior to giving the test, an answer key and a simple scoring method should be prepared.

Write each test item on an individual note card. Hall and Paolucci (1970) suggested a 5 X 8-inch card since it allows room for writing the objective and the answer, as well as information about the item analysis and notes on the item effectiveness. These cards may be used in developing an effective item bank of possible test questions. Hall and Paolucci (1970) gave suggestions on writing effective test items.

1. Use an item best suited to the content and specific objective being measured.

2. Select items which require the application of what the student has learned, rather than recall or recognition type questions.

3. Rather than taking items directly from the book, provide new situations in which to test the student's ability to apply what he/she has learned.

4. Make the content of the item such that the student must think before answering.

5. The content of an item, not how it is worded, should determine if the answer is correct.

6. Each item should contain only one idea and be short and definite.

7. Simplify items by using language familiar to students.

8. Reduce the possibility of guessing by selecting items with more than two choices.

9. Unless indicated differently in the directions, be sure there is only one correct answer.

10. Use correct grammar. Avoid irrelevant clues such as 'a' or 'an.'

11. Clarify any words with hidden meanings.

True-false, matching, and multiple choice items are referred to as selection-type items. They are self-contained, because they do not allow the students to make a response by going beyond the content of the item (Hopkins and Antes, 1978). These test items all require that the student select an answer from a given number of alternatives (Gronlund, 1985).

True-False Items

True-false items are the most common form of the alternative-response item. Other forms are right-wrong, correct-incorrect, yes-no, fact-opinion, and agree-disagree. The true-false question is most commonly used to measure the ability to identify the correctness of statements of fact or principles and definition of terms. The most useful of these is in distinguishing fact from opinion (Gronlund, 1985).

True-false questions are often criticized because although a student may recognize that a statement is false, he/she may not know what is true. Another disadvantage is that students tend to guess on true-false tests. They have a 50-50 chance of guessing correctly (Gronlund, 1985).

Gronlund (1985) presented eight suggestions to aid in the construction of true-false items. They are:

1. Avoid the use of broad general statements.
2. Avoid trivial statements.
3. Avoid negative statements, especially double negatives.
4. Avoid using two ideas in one statement, unless measuring a cause and effect relationship.
5. Attribute an opinion to some source, unless the ability to identify the opinion is being measured.

6. Avoid complex, lengthy statements.

7. Try to equalize the length of true and false statements.

8. Try to equalize the number of true and false statements.

Matching Items

Matching exercises traditionally consist of two parallel columns with each item in one column being matched to a word, sentence or phrase in the second column. Premises are the items in the column for which a match is being sought, while responses are the items in the column from which the selection is being made. An imperfect match is when there are more responses than premises. In this case directions should be written in a manner to indicate whether the responses may be used once, more than once, or not at all (Gronlund, 1985).

In determining how many items should be in each column Gronlund (1985) stated that four to seven items is best while there should never be more than ten items in either column. Ahmann and Glock (1981) indicated that a good guide to follow is to only exceed ten items if the maturity of the student and the nature of the subject matter permits it. For younger students they suggest restricting the number of items to five.

Matching questions are most appropriate as stated by Gronlund (1985, p. 162) "whenever learning outcomes emphasize the ability to identify the relationship between two things, and a sufficient number of homogeneous premises and responses can be obtained." Gronlund listed six suggestions to follow when constructing matching exercises:

1. In a single exercise, use only homogeneous material.

2. Include an unequal number of responses and premises. Give clear directions for completing.

3. Keep items to be matched brief with the shorter responses on the right.

4. Arrange responses in a logical order (alphabetical or numerical).

5. In the directions, indicate the basis for matching the premises and responses.

6. Place all items in the exercise on one page.

Multiple-Choice Items

The most applicable and useful objective test item is the multiple-choice item. It consists of a problem and a list of solutions. The problem, called the stem, may be a direct statement or an incomplete statement. The solutions, known as alternatives, choices or options, may include words, numbers, phrases, or symbols. The correct alternative is the answer. The remaining alternatives are distracters, decoys, or foils (Gronlund, 1985).

Ebel (1979) stated that it has been common practice to use three or four distracters for each item. When a larger number of good distracters is used the item is likely to be highly discriminating. As more distracters are written, however, they generally become weaker. For this reason Ebel (1979) suggested that three good distracters may be the best, however, there is no reason why all test items must have the same number of alternatives.

Multiple-choice items are used to test student knowledge of terminology, specific facts, principles and methods and procedures (Gronlund, 1985). Gronlund made 13 suggestions when writing multiple-choice items:

1. The stem should be meaningful and present a definite problem.

2. The stem should not include irrelevant material but should include as much of the item as possible.

3. Use a negative stem only when significant learning outcomes require it.

4. The stem and alternatives should be grammatically consistent.

5. Items should contain only one correct or best answer.

6. Items measuring understanding should contain some novelty.

7. Distracters should be plausible.

8. Avoid verbal connections between the stem and the correct answer.

9. A clue to the answer should not be provided by the length of the alternatives.

10. Correct answers should be found in an equal number in each alternative position, but in a random order.

11. Be sparing in the use of alternatives like 'none of the above' or 'all of the above.'

12. When another type is more appropriate, do not use multiple-choice items.

13. When you have a good reason, break any of these rules.

Item Analysis Procedures

After tests have been given an item analysis may be conducted. Item analysis serves two main purposes. The first is that the results provide a better indication of the worth of tests and make it possible to construct better tests. Second, it can be used for diagnosis (Ahmann and Glock, 1981).

Gronlund (1985) outlined a procedure for conducting an item analysis. His sample included 32 test papers. Steps were as follows:

1. Rank the papers from the highest score to the lowest.
2. Select the 10 highest scoring papers and the 10 lowest scoring papers.
3. The remaining 12 papers will not be used in the analysis.
4. For each test item calculate the number of students in the two groups who selected each alternative.
5. Compute the difficulty for each item.
6. Compute the discriminating power for each item.
7. Evaluate the effectiveness of each item's distracters.

Computing Item Difficulty

A test item's difficulty is determined by the percentage of correct answers given for the item. To compute item difficulty the following formula is used:

$$\text{Item Difficulty} = \frac{R}{T} \times 100$$

where, R = the number of students who correctly answered the item and T = the total number of students who tried the item (Gronlund, 1985).

The item difficulty represents the percentage of students who responded to the item correctly. All items with extremely high or low difficulty levels should be carefully analyzed. The discriminating power tends to be lowered if the difficulty level is not in the range of 40 to 70 percent. One use of items to which 100 percent responded correctly is to place one or two of them at the beginning of the test to serve as a simple transition into the test (Ahmann and Glock, 1981).

If all students respond to all items on a four or five alternative multiple-choice test, the expected range of scores is 25 to 100. The ideal difficulty level is 65 (Gay, 1980). Ahmann and Glock (1959) recommended that a midrange level, between 40 and 70 percent be used as a guideline.

Computing Discriminating Power

The degree to which a test item discriminates between students with low and high scores is known as discriminating power. Item discriminating power is obtained by using the formula:

$$\text{Item Discriminating Power (D)} = \frac{R_U - R_L}{\frac{1}{2}T}$$

where, R_U is the number of students in the upper group who have the correct answer, R_L is the number of students in the lower group who have the correct answer and $\frac{1}{2}T$ is half of the total number of students in the item analysis (Gronlund, 1985).

It would be ideal to have all test items with a positive discrimination. An item's discrimination is positive if more students in the upper group answer the item correctly than in the lower group. A positive discrimination shows that the item discrimination is in the same direction as the total score (Gronlund, 1985).

Indices of Discriminating Power

After computing discriminating power, D values above +0.40 may be interpreted as very good, between +0.40 and +0.20 as satisfactory, and between +0.20 and zero as poor. It has been suggested that greater than 50 percent of the test items should be between +0.40 and +0.20, fewer

than ten percent between +0.20 and zero, while none should have a negative value (Ahmann and Glock, 1981). The remaining 40 percent should be above +0.40.

Table of Specifications

A table of specifications is a two-way chart relating instructional objectives to the course content. It specifies the emphasis to be given to each learning outcome. To build a table of specifications you must first determine the instructional objectives, then outline the course content and third, prepare the chart (Gronlund, 1985).

When planning the classroom test all intended learning outcomes need to be considered. The types of performance expected by the student [know, understand, apply] are described in the instructional objectives (Gronlund, 1985). The area in which each type is to be shown is indicated by the course content. The course content outline may be a list of major topics to be covered or a detailed list of topics and subtopics. The purpose of the test, the part of the course that is covered and the interpretation of the test determines how detailed the outline should be (Gronlund, 1985).

Developing the two-way chart is the final step in preparing the table of specifications. By relating the objectives to the course content the nature of the test is determined. Prepare the chart by writing the general instructional objectives across the top of the table. Down the left-hand side list the major content areas. Then weight the items by assigning averages for each objective across the bottom row and percentages for each content area down the right-hand

column. Finally, for each of the two-way cells, allot the percentage or number of test items to be used.

A table of specifications may be expanded by adding specific objectives to each general objective and using a detailed course content. This may be done as long as the number remains manageable (Gronlund, 1985).

Past Studies Conducted at Oklahoma State University

Since 1959 various studies have been conducted at Oklahoma State University in an effort to develop, revise, and/or test instruments for use in a basic clothing construction course. The first study was conducted by Walsh in 1959. Walsh's study involved the development of a pretest which could be used as a placement device in the basic clothing construction course. The pretest was used in conjunction with a student questionnaire, which was administered to 135 freshmen. Walsh designed the questionnaire in order to help determine the varying degrees of clothing experiences that freshmen entering college have. Walsh indicated that the best way to develop a better test is to use an existing test, study the results obtained, and then make revisions.

Witt (1961) conducted an item analysis based on the responses of 112 freshmen clothing students to the pretest developed by Walsh. Those items found to be discriminating were revised for the pretest and additional items were added. The test consisted of matching, multiple-choice and true-false items and was designed to evaluate the students' knowledge of clothing construction, care, selection, and the ability to apply principles.

Since it is not possible to evaluate all competencies through a written test, Witt developed a station-to-station test consisting of seven problems. Three of these were designed to evaluate the use of manipulative skills, while the remaining four evaluated the use of judgmental skills. Each problem was set up at a separate table with instructions and materials. During a given time period students moved from station to station to complete the assigned tasks.

Witt concluded that in order to satisfactorily place students in clothing courses, various types of clothing competencies need to be evaluated. Witt recommended that further studies be conducted into the use of evaluation instruments used as exemption devices.

In 1963, Berry and Gould further revised the previously developed tests. Berry worked with the written test, while Gould worked with the performance or station-to-station test. During the fall semester of 1962, Berry conducted a pilot study which included an item analysis providing data which was used as a basis for revising the Witt pretest. The revised pretest was given to 76 beginning clothing students during the spring semester of 1963. Data were obtained and studied and recommendations were made for further revision prior to utilizing the test. Berry further recommended that a variety of evaluative instruments be used in conjunction with the written test in order to maintain validity.

Gould (1963) worked to further revise and develop a clothing construction performance pretest. Twenty-four students participated in a pilot study which consisted of the three manipulative problems in Witt's station-to-station test plus six additional problems. The study indicated that four or five manipulative problems were appropriate

for a one hour test. Results of the study revealed three problems: 1) confusion created by movement between stations, 2) congested traffic at stations with some problems requiring more time than others, and 3) a shortage of supplies due to some students using more than the allotted amount. Additional results showed that students working at the same station influenced each other.

As a result of the pilot study the test was revised to include five problems. The method of administration was changed by placing the needed materials in a large manila envelope for each student. The revised test was given to 77 students during the spring of 1963. Results showed an improvement over the test administered in the pilot study, thus concluding that the test could aid in placing students in college clothing courses.

The focus of Souligny's research (1971) was to evaluate the clothing exemption test. Two groups were given the test. Group I involved 267 students who took the test as an exemption test while the 131 students in Group II took the test as a final examination in the beginning clothing course. An item analysis was conducted on each group of scores and the results were compared. Souligny concluded that the test had a greater discriminating power as an exemption test, than as a final examination. The assumption that students entered college with varying clothing construction knowledge and skills was proven by the excessive range of scores (8 to 94). It was recommended that areas of the test where Group II scores were low be identified and emphasis be placed on these areas in the beginning clothing course.

Two studies have been conducted concerning the use of computers with clothing tests. Wilkins (1971) studied the feasibility of computer-

generated tests for the basic clothing construction course. The test for the Acquisition and Use of Clothing unit was given to 225 students during the fall of 1970. An item analysis was conducted and revisions made where needed. Each test question was punched on a computer card and then stored on a computer tape deck. Thirty different forms of the test and corresponding keys, each having 35 questions, were randomly generated. In the spring of 1971, 141 students took the computer-generated tests. An item analysis was conducted on the computer test and an estimate of time and cost was made. Students were asked to complete a questionnaire concerning their reaction to the test.

Results of the study showed the means, standard deviation, and highest and lowest scores were similar on the tests given in fall 1970 and spring 1971. Most students preferred the easier to read computer written test to the teacher written one. Preference was also given to the answer sheet which could be graded before they left the classroom. The average cost of $1.06 per individual test was too great to justify generating a test for every student. Wilkins recommended that further analysis be made of this and other computer-generated tests.

The second study using computer testing was conducted by Good (1974) to determine the feasibility of a computer-generated test via the cathode ray tube for a basic clothing construction course. The 50 students who participated in the study were divided into two groups. The control group contained 26 students, while the experimental group was given the new computer-generated test which was presented on a cathode ray tube terminal. Thirteen subject matter categories were used for developing test items which were keypunched and stored on a computer disc.

Students in the control group signed up for a convenient time to take the examination. After entering the proper code, the computer generated 100 questions. After completion of the examination, the score was flashed on the screen. The students were then asked to complete a survey which was used to determine the advantages and disadvantages of the computer-generated examination.

The majority of students preferred the computer-generated test to the paper-and-pencil test. They also preferred using the light pen rather than the keyboard when answering questions. Advantages given to the computer-generated test were that it was easier to read, faster to take, and students liked knowing their score immediately. The main disadvantage was the noise in the computer room. After analyzing the examination using a t-test, Good concluded that the computer-generated test is at least as effective as paper-and-pencil testing.

Miller's study (1974) identified competencies to be implemented into a beginning college clothing construction course. The competencies were based on results of a questionnaire mailed to 224 participants in six categories: 1) Clothing Specialists in Extension, 2) State and District Supervisors in Home Economics Education, 3) Secondary Vocational Homemaking Teachers, 4) Clothing Professors in State Colleges, 5) Clothing Professors in Universities, and 6) Home Economists in Business.

The average percent of responses was found by adding the responses for each of the six groups of participants and then dividing by six. Fifty-two competencies of 50 percent and higher were selected for inclusion in a beginning clothing course. Seventeen of the competencies were selected for a course other than beginning clothing construction.

None of the competencies were thought unimportant by all six groups. Miller recommended that further studies be conducted for updating competencies and developing learning packages.

More recently, two studies have been conducted on the development of advanced standing examinations. Lisenby (1979) developed an item pool for an advanced standing examination for a basic clothing construction course. Stufflebean (1982) carried the work further to develop and implement a computerized advanced standing examination for basic clothing construction.

Lisenby's study (1979) was to develop an item pool from which an advanced standing examination could be developed. Data were collected from a comparison of advanced standing scores and experience checklist scores between fall 1972 and fall 1975. Additionally a comparison was conducted of written examination scores and practical assignment scores of students enrolled in basic clothing construction between fall 1975 and fall 1977. An item analysis of written examinations being used in basic clothing construction was used to determine acceptable items for inclusion in an item pool. After completion of the first draft of the item pool, one clothing and textiles professor and eight graduate students enrolled in a graduate level evaluation course critiqued the item pool and made suggestions.

Results indicated that an experience checklist showed no significant indication as to how well the student would perform on the written examination in basic clothing construction. Results indicated that practical assignments may be unnecessary in an advanced standing examination in basic clothing construction. Lisenby recommended that further assessment be made into the use of advanced standing examinations and that the item pool be further developed.

Stufflebean (1982) conducted research to determine competencies to be used in the development and implementation of a computer-generated advanced standing examination for a basic clothing construction course. Competencies were determined through the use of a questionnaire completed by 160 selected clothing faculty representing 60 member institutions of the National Association of State Universities and Land Grant Colleges. Seventy-seven competencies were considered essential for use in an item pool for an advanced standing examination.

An item pool was developed with 500 items which were categorized by topic areas and competencies. An item analysis was conducted to determine which items needed revision. After revisions were completed items were entered into the computer using the computer program designed to generate, score, and analyze the examination. During January, 1982, eight students took the computer-generated advanced standing examination.

Stufflebean concluded that the computer-generated advanced standing examination was effective as an evaluation device for awarding credit and/or advancing students. Administration of the examination through the computer was an acceptable method for both students and faculty.

Summary

The development of a classroom test begins with knowing what is to be evaluated. This is necessary in knowing why the test is to be given. In constructing the test items general guidelines should be followed. By following the given procedure for item analysis the effectiveness of a test may be determined. This is generally done by computing item difficulty, computing discriminating power, interpreting

indices of discriminating power and evaluating the effectiveness of distracters. A table of specifications is helpful in determining the specific test items needed for each objective. The use of individual cards for each test item aids the development of an item bank of possible test questions.

Numerous studies have been conducted at Oklahoma State University concerning the evaluation, development, and/or revision of tests or other aspects of basic clothing construction courses. Walsh (1959), Witt (1961), Berry (1963), and Gould (1963) all worked on the development and/or revision of a clothing construction pretest. Souligny's research (1971) was to evaluate a clothing exemption test. Wilkins (1971) and Good (1974) studied the feasibility of computer-generated testing for a basic clothing construction course. Miller (1974) identified competencies to be implemented into a beginning clothing construction course. Most recently Lisenby's (1979) and Stufflebean's (1982) research dealt with an advanced standing examination. Lisenby developed an item pool of questions, while Stufflebean determined competencies for use in the development and implementation of a computer-generated advanced standing examination.

CHAPTER III

METHODS AND PROCEDURES

The problem addressed in the study was to evaluate tests for Modules 1 and 2 of the clothing construction course and to determine needed revisions. The method and procedures used to conduct the study are described in this chapter. Objectives for the study were to determine item difficulty by using past tests, to determine item discriminating power by using past tests, to develop a table of specifications of test items for Module 1 and Module 2, and to develop an item bank of test questions to be used in developing a final test for each module.

Research Design

The study consisted of three phases: item analysis, including determining item difficulty and item discriminating power, developing a table of specifications, and developing an item bank.

Phase I - Item Analysis

The tests used in the study were tests given for Modules 1 and 2 during the semesters of spring 1985, fall 1985, and spring 1986. The students' responses were transferred from the hand graded answer sheets to computer coded answer sheets. All coding was then checked for accuracy. The tests were divided into three groups: 1) Test 1, Form 1;

2) Test 1, Form 2; and 3) Test 2, Form 1. There was only one form of test 2, the test over Module 2. The tests were then taken to the Bureau of Tests and Measurements at Oklahoma State University for data calculation.

The following were calculated for each test:

1. Item difficulty,
2. Item discrimination index,
3. Number of responses to each alternative and the number of omitted responses,
4. Distribution of item difficulties,
5. Distribution of item discrimination indexes, and
6. Test statistics including (actual and recommended) mean, standard deviation, reliability, standard error measure, mean difficulty, and mean discrimination.

Items with a difficulty level between 40 and 70 percent have the best discriminating power. Item discriminating values above +0.40 are very good, between +0.40 and +0.20 are satisfactory, and between +0.20 and zero are poor (Ahmann and Glock, 1981).

Phase II - Developing a Table of Specifications

A table of specifications was needed for each module (1 and 2) in the study. The first step in developing a table is to determine the instructional objectives. The second step is to outline the course content and the third is to prepare the chart (Gronlund, 1985).

Determining Instructional Objectives

The instructional objectives were identified from the packets for Modules 1 and 2. The objectives included 1) defining terms, 2) identifying specific facts, 3) recognizing principles, and 4) interpreting instructions for construction.

Bloom (1971) described the Taxonomy of Educational Objectives as including three domains: cognitive, affective, and psychomotor. Objectives which involve intellectual tasks are classified within the cognitive domain. The behavioral aspects of the objectives are placed within a hierarchy with categories ranging from simple to complex. The six levels within the cognitive domain are: knowledge, comprehension, application, analysis, synthesis, and evaluation. Since the basic clothing construction course is a freshman/sophomore level course, the objectives should be primarily in the lower levels. Objectives 1 and 2 are in level 1, knowledge; objective 3 is in level 3, application; and objective 4 is in level 2, comprehension.

Outlining Course Content

From the objectives the course content for each module was outlined by topical areas. For example, in Module 1 a few of the content areas were: preparing the fabric, pattern layout and cutting, transfer of pattern markings, darts, seams, seam finishes, interfacing, fasteners, hems, and pattern alterations. Content areas were similarly outlined for Module 2.

Preparing the Chart

The chart was first prepared by writing the instructional objectives

across the top of the table. The major content areas were then written down the left hand side. The items were weighted by assigning the number of items for each objective across the bottom row and assigning percentages for each content area down the right hand column. In each of the two-way cells the percentage or number of test items to be used was allotted. Two charts were developed - one for Module 1 and the second for Module 2.

Phase III - Developing an Item Bank

The item analysis data from Phase I of the study and the table of specifications constructed in Phase II were used to develop the item bank of questions. Items were constructed using item difficulty levels and item discriminating power of the present test items and objectives and content areas identified in the table of specifications.

Resources used in revising present items and developing new test items for each module included textbooks, course modules, written examinations and advanced standing examinations for clothing construction which are currently used at Oklahoma State University.

Each test item was listed on a 4 X 6-inch index card. Each card included the course title, the module number and title, the subject matter content, the stem and alternatives and the correct answer. Items included multiple-choice, matching, and true-false. Illustrations were included where needed.

CHAPTER IV

ANALYSIS OF DATA

The purpose of the study was to evaluate the tests used for Modules 1 and 2 of the basic clothing construction course at Oklahoma State University and to determine needed revisions. Data were calculated using the tests taken by students during the spring 1985, fall 1985, and spring 1986 semesters.

The study was conducted in three phases. Phase one consisted of conducting an item analysis to determine the item difficulty and item discriminating power for each item on test 1, form 1; test 1, form 2; and test 2, form 1. Phase two included the development of a table of specifications for each module, 1 and 2. The third phase included the development of an item bank of test questions to be used in the development of a final test for each module.

Phase I - Item Analysis

Data were calculated for each of the three tests - test 1, form 1; test 1, form 2; and test 2, form 1. Item difficulty and item discriminating power were obtained for each test. Table I includes the data for test 1, form 1 and Table II gives the data for test 1, form 2.

It was discovered that item number 8 on test 1, form 1 had the wrong answer indicated as the correct one. It was therefore necessary to recalculate the difficulty level and discrimination index for that

TABLE I

ITEM ANALYSIS - TEST 1 FORM 1
(N=157)

Item No.	Number Correct Responses[a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
1	144	0.40	Good	91.72	Too easy
2	116	0.17	Poor	73.89	Too easy
3	156	0.04	Poor	99.36	Too easy
4	111	0.38	Satisfactory	70.70	Too easy
5	119	0.34	Satisfactory	75.80	Too easy
6	143	0.36	Satisfactory	91.08	Too easy
7	139	0.35	Satisfactory	88.54	Too easy
8[d]	5	-0.08	Poor	3.18	Too difficult
9	137	0.11	Poor	87.16	Too easy
10	118	0.34	Satisfactory	75.16	Too easy
11	143	0.32	Satisfactory	91.08	Too easy
12	147	0.19	Poor	93.63	Too easy
13	151	0.32	Satisfactory	96.18	Too easy
14	121	0.48	Good	77.07	Too easy
15	103	0.37	Satisfactory	65.61	Good
16	82	0.37	Satisfactory	52.23	Good
17	127	0.36	Satisfactory	80.89	Too easy
18	59	0.27	Satisfactory	37.58	Too difficult
19	115	0.20	Satisfactory	73.25	Too easy
20	131	0.20	Satisfactory	83.44	Too easy
21	126	0.18	Poor	80.25	Too easy
22	148	0.09	Poor	94.27	Too easy
23	68	0.27	Satisfactory	43.31	Too difficult
24	152	0.06	Poor	96.82	Too easy
25	15	-0.09	Poor	9.55	Too difficult
26	140	0.31	Satisfactory	89.17	Too easy
27	88	0.29	Satisfactory	56.05	Good
28	120	0.23	Satisfactory	76.43	Too easy
29	148	0.35	Satisfactory	94.27	Too easy
30	86	0.23	Satisfactory	54.78	Good

TABLE I (Continued)

Item No.	Number Correct Responses [a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
31	153	0.08	Poor	97.45	Too easy
32	153	0.10	Poor	97.45	Too easy
33	147	0.25	Satisfactory	93.63	Too easy
34	77	0.44	Good	49.04	Good
35	137	0.32	Satisfactory	87.26	Too easy
36	141	0.25	Satisfactory	89.81	Too easy
37	126	0.04	Poor	80.25	Too easy
38	89	0.40	Good	56.69	Good
39	98	0.20	Satisfactory	62.42	Good
40	68	0.36	Satisfactory	43.31	Good
41	132	0.40	Good	84.08	Too easy
42	127	0.37	Satisfactory	80.89	Too easy
43	120	0.23	Satisfactory	76.43	Too easy
44	68	0.30	Satisfactory	43.31	Good
45	139	0.17	Poor	88.54	Too easy
46	12	0.10	Poor	7.64	Too difficult
47	113	0.29	Satisfactory	71.97	Too easy
48	136	0.43	Good	86.62	Too easy
49	129	0.17	Poor	82.17	Too easy
50	70	0.35	Satisfactory	44.59	Good

[a]Based on the responses of 157 students who took the examination.

[b]Any discriminating value above +0.40 is good, between +0.40 and +0.20 is satisfactory, and below +0.20 is poor.

[c]Items with a difficulty level of above 70 percent are too easy, between 40-70 percent are good, and below 40 percent are too difficult.

[d]The correct response for item 8 was incorrectly marked on the key.

TABLE II

ITEM ANALYSIS - TEST 1 FORM 2
(N=22)

Item No.	Number Correct Responses[a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
1	18	0.60	Good	81.82	Too easy
2	15	0.22	Satisfactory	68.18	Good
3	11	0.02	Poor	50.00	Good
4	15	0.04	Poor	68.18	Good
5	21	0.47	Good	95.45	Too easy
6	10	0.45	Good	45.45	Good
7	19	0.62	Good	86.36	Too easy
8	20	0.30	Satisfactory	90.91	Too easy
9	18	0.66	Good	81.82	Too easy
10	16	0.53	Good	72.73	Too easy
11	16	0.45	Good	72.73	Too easy
12	16	0.60	Good	72.73	Too easy
13	13	0.10	Poor	59.09	Good
14	15	0.05	Poor	68.18	Good
15	16	0.18	Poor	72.73	Too easy
16	20	0.48	Good	90.91	Too easy
17	19	0.57	Good	86.36	Too easy
18	15	0.71	Good	68.18	Good
19	13	0.09	Poor	59.09	Good
20	20	0.38	Satisfactory	90.91	Too easy
21	19	0.31	Satisfactory	86.36	Too easy
22	17	0.54	Good	77.27	Too easy
23	17	0.58	Good	77.27	Too easy
24	14	0.56	Good	63.64	Good
25	14	0.53	Good	63.64	Good
26	10	0.30	Satisfactory	45.45	Good
27	17	0.01	Poor	77.27	Too easy
28	14	0.53	Good	63.64	Good
29	6	0.32	Satisfactory	27.27	Too difficult
30	18	0.31	Satisfactory	81.82	Too easy

TABLE II (Continued)

Item No.	Number Correct Responses[a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
31	13	0.23	Satisfactory	59.09	Good
32	19	0.42	Good	86.36	Too easy
33	11	0.41	Good	50.00	Good
34	7	0.03	Poor	31.82	Too difficult
35	16	0.59	Good	72.73	Too easy
36	17	0.30	Satisfactory	77.27	Too easy
37	5	0.10	Poor	22.73	Too difficult
38	14	0.43	Good	63.64	Good
39	13	0.30	Satisfactory	59.09	Good
40	9	0.56	Good	40.91	Good
41	21	0.47	Good	95.45	Too easy
42	14	0.61	Good	63.64	Good
43	13	0.76	Good	59.09	Good
44	13	0.51	Good	59.09	Good
45	17	0.56	Good	77.27	Too easy
46	19	0.31	Satisfactory	86.36	Too easy
47	19	0.34	Satisfactory	86.36	Too easy
48	21	0.09	Poor	95.45	Too easy
49	2	0.19	Poor	9.09	Too difficult
50	7	0.49	Good	77.27	Too easy

[a]Based on the responses of 22 students who took the examination.

[b]Any discriminating value above +0.40 is good, between +0.40 and +0.20 is satisfactory, and below +0.20 is poor.

[c]Items with a difficulty level above 70 percent are too easy, between 40-70 percent are good, and below 40 percent are too difficult.

item using the correct answer. The recalculations showed these as difficulty = 90.38 and discrimination = 0.08 instead of difficulty = 3.18 and discrimination = 0.08 as indicated in Table I.

Table III includes the data calculated for test 2, form 1. As stated in Chapter III, there was only one form of test 2.

Ahmann and Glock (1981) stated that more than 50 percent of the test items should be between +0.40 and +0.20, less than 10 percent should be between +0.20 and zero, and none should have negative values. This left 40 percent to be above +0.40. The item difficulty level for each item should be in the midrange level between 40 and 70 percent (Ahmann and Glock, 1959). Table IV indicates the distribution of test items by difficulty and discrimination levels.

In general, on the two forms of test 1, almost one-half of the items had satisfactory discrimination, about one-third had good discrimination, and approximately one-fourth had poor discrimination. On test 2, almost one-half had satisfactory discrimination, a little over one-third had poor discrimination, and less than one-fourth had good discrimination. For both tests, too many of the items had satisfactory or poor discrimination levels.

On the two forms of test 1, in general, slightly less than two-thirds of the items were too easy, approximately one-third had a good difficulty level, and less than one-tenth were too difficult. On test 2, slightly more than three-fourths were too easy, slightly more than one-tenth had a good difficulty level, and less than one-tenth were too difficult. For both test 1 and test 2, too many of the items were too easy.

TABLE III

ITEM ANALYSIS - TEST 2 FORM 1
(N=109)

Item No.	Number Correct Responses[a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
1	96	-0.04	Poor	88.07	Too easy
2	93	0.10	Poor	85.32	Too easy
3	70	0.27	Satisfactory	64.22	Good
4	105	0.26	Satisfactory	96.33	Too easy
5	90	0.42	Good	82.57	Too easy
6	84	0.32	Satisfactory	77.06	Too easy
7	83	0.27	Satisfactory	76.15	Too easy
8	80	0.13	Poor	73.39	Too easy
9	96	0.01	Poor	88.07	Too easy
10	92	0.28	Satisfactory	84.40	Too easy
11	87	0.48	Good	79.82	Too easy
12	97	0.44	Good	88.99	Too easy
13	106	0.07	Poor	97.25	Too easy
14	95	0.28	Satisfactory	87.16	Too easy
15	98	0.18	Poor	89.91	Too easy
16	99	0.09	Poor	90.83	Too easy
17	73	0.40	Good	66.97	Good
18	84	0.20	Satisfactory	77.06	Too easy
19	103	0.25	Satisfactory	94.50	Too easy
20	108	0.35	Satisfactory	99.08	Too easy
21	104	0.31	Satisfactory	95.41	Too easy
22	22	0.11	Poor	20.18	Too difficult
23	107	0.25	Satisfactory	98.17	Too easy
24	109	0.00	Poor	100.00	Too easy
25	106	0.08	Poor	97.25	Too easy
26	94	0.27	Satisfactory	86.24	Too easy
27	74	0.13	Poor	67.89	Good
28	94	0.47	Good	86.24	Too easy
29	78	0.42	Good	71.56	Too easy
30	91	0.23	Satisfactory	83.49	Too easy

TABLE III (Continued)

Item No.	Number Correct Responses[a]	Discrimination Index	Rating[b]	Difficulty (Percent)	Rating[c]
31	28	0.17	Poor	25.69	Too difficult
32	99	0.18	Poor	90.83	Too easy
33	103	0.28	Satisfactory	94.50	Too easy
34	81	0.57	Good	74.31	Too easy
35	91	0.30	Satisfactory	83.49	Too easy
36	80	0.32	Satisfactory	73.39	Too easy
37	70	0.17	Poor	64.22	Good
38	26	-0.03	Poor	23.58	Too difficult
39	102	0.13	Poor	93.58	Too easy
41	94	0.44	Good	86.24	Too easy
42	98	0.38	Satisfactory	89.91	Too easy
43	69	0.23	Satisfactory	63.30	Good
44	90	0.26	Satisfactory	82.57	Too easy
45	13	0.05	Poor	11.93	Too difficult
46	94	0.19	Poor	86.24	Too easy
47	84	0.16	Poor	77.06	Too easy
48	56	0.76	Good	51.38	Good
49	75	0.30	Satisfactory	68.81	Good
50	98	0.51	Good	89.91	Too easy

[a]Based on the responses of 109 students who took the examination

[b]Any discrimination value above +0.40 is good, between +0.40 and +0.20 is satisfactory, and below +0.20 is poor.

[c]Items with a difficulty level above 70 percent are too easy, between 40-70 percent are good, and below 40 percent are too difficult.

TABLE IV

DISTRIBUTION OF TEST ITEMS BY DIFFICULTY AND DISCRIMINATION LEVELS

	Test 1, Form 1		Test 1, Form 2		Test 2, Form 1	
	Number of Items	Percent	Number of Items	Percent	Number of Items	Percent
Discrimination						
Above +0.40	6	12	27	54	10	20
+0.40 - +0.20	29	58	12	24	21	42
+0.20 - 0	13	26	11	22	17	34
Below 0	2	4	0	0	2	4
Difficulty						
Above 70 percent	35	70	26	52	39	78
40 - 70 percent	11	22	20	40	7	14
Below 40 percent	4	8	4	8	4	8

As a part of the item analysis, additional statistics were calculated for each test. Table V shows both the recommended value and the actual value for each of the three tests. The following are explanations of the statistics calculated.

Mean - the average score, based on a possible score of 50.

Standard deviation - a measure indicating how the dispersion of a distribution of scores varies.

Reliability - the extent to which a test is consistent in measuring what it is intended to measure.

Standard error measurement - based on the standard deviation and reliability of the test it is the estimate of the possible error involved in each student's score.

Mean difficulty - the average difficulty level.

Mean discrimination - the average discrimination power.

All three tests were within the acceptable recommendation for mean and reliability. On test 1, form 2, standard deviation, mean difficulty, and mean discrimination were within the acceptable recommendation. Test 2, form 1 was within the acceptable range for mean difficulty. All other values were below the recommendations.

Table X, XI, and XII (Appendix) show the distribution of student responses for test 1, form 1; test 1, form 2; and test 2, form 1; respectively. This data can be used to estimate the effectiveness of the various distracters.

Phase II - Table of Specifications

Tables VI and VII are tables of specifications developed for Modules 1 and 2, respectively. The tables indicate the recommended

TABLE V

INDIVIDUAL STATISTICAL MEASURES FOR THE VARIOUS TESTS

	Test 1, Form 1		Test 1, Form 2		Test 2, Form 1	
	Actual	Recommended	Actual	Recommended	Actual	Recommended
Mean	36.26	32.10	34.18	31.49	39.17	31.63
Standard Deviation	4.99	5.97+	7.98	6.17+	4.43	6.12+
Reliability	0.75	0.70+	0.89	0.70+	0.71	0.70+
Standard Error Measure	2.50	-	2.65	-	2.37	-
Mean Difficulty	72.52	64.20	68.36	62.98	78.35	63.27
Mean Discrimination	0.26	0.29+	0.38	0.30+	0.25	0.30+

TABLE VI

TABLE OF SPECIFICATIONS
MODULE 1

Content Area	Type of Objective				
	Define Terms	Identify Specific Facts	Recognize Principles	Interpret Instructions for Construction	Total Percent
Small sewing equipment	5				5
Fabric preparation	1		3	1	5
Fabric layout and cutting		3	3		6
Markings	2	2			4
Hand basting	2			1	3
Sewing machine parts	5				5
Machine stitching		4	3	2	9
Seams	2	2		3	7
Knit seams		3			3
Seam finishes	2	3		3	8
Seam techniques	2	3		3	8
Darts				5	5
Interfacings		2	1	1	4
Bias				2	2
Buttonholes		1		2	3
Buttons		1		2	3
Zippers		1		1	2
Fasteners			2	3	5
Hems		3		3	6
Knit hems		1		2	3
Alterations		1	3		4
Total	21	30	15	34	100

TABLE VII

TABLE OF SPECIFICATIONS
MODULE 2

Content Area	Type of Objective				
	Define Terms	Identify Specific Facts	Recognize Principles	Interpret Instructions for Construction	Total Percent
Take body measurements			9		9
Pattern selection		5	3		8
Select fabric and notions		7			7
Fit		5			5
Prepare fabric			2		2
Pattern layout and cutting		4	4		8
Transfer markings		6			6
Seams and/or finish	1	1		3	5
Hem finish and stitch	1	1		3	5
Machine stitching			3		3
Facings and interfacings	2			3	5
Closures		2		2	4
Darts, gathers, pleats		2		3	5
Set-on collar	2	2		4	8
Set-in sleeves	2	2		4	8
Waistline seam		1		3	4
Pressing		3		5	8
Total	8	41	21	30	100

distribution of test items according to the emphasis placed on each type of objective and in each content area.

Phase III - Development of an Item Bank

Seven items on test 1, form 1 and test 1, form 2 were duplicates. Those items were reviewed to determine if there were any significant differences in the difficulty level and the discrimination index obtained. Table VIII shows the comparison of this. It was decided to recalculate the item difficulty and item discriminating power for each item based on student responses to the item on both tests - forms 1 and 2. The recalculated values are indicated in Table IX. These new values were then used for these particular items in developing the item bank.

In an effort to develop an effective item bank it had to be determined which items from the original tests should be kept, revised, and omitted. Item difficulty level and item discriminating power were used in making this determination. The items which were kept and revised were identified by content area and compared to the table of specifications in order to determine the content areas of the items to be added. After this determination was made the number of items in each of the discrimination categories was counted. On the Module 1 test, of those items with good discrimination, 27 were kept, zero were revised, and four were omitted due to the lack of importance as related to the content areas of the course. Of those items rated satisfactory, 25 were kept, six were revised, and seven were omitted. From those items rated poor, zero items were kept, seven items were revised, and 17 items were omitted. Thirty-five new items were added to complete the 100 question item bank for Module 1.

TABLE VIII

ORIGINAL VALUES OF DUPLICATE ITEMS

Test 1, Form 1			Test 1, Form 2		
Item No.	Difficulty	Discrimination	Item No.	Difficulty	Discrimination
5	75.80	0.34	2	68.18	0.22
11	91.08	0.32	22	77.27	0.54
20	83.44	0.20	46	86.36	0.31
21	80.25	0.18	47	86.36	0.34
24	96.82	0.06	48	95.45	0.09
25	9.55	-0.09	49	9.09	0.19
26	89.17	0.31	50	77.27	0.49

TABLE IX

RECALCULATED VALUES OF DUPLICATE ITEMS

Item No.		Difficulty	Discrimination
Form 1	Form 2		
5	2	77.12	0.25
11	22	88.98	0.22
20	46	83.05	0.24
21	47	83.90	0.12
24	48	97.46	0.05
25	49	11.02	-0.02
26	50	85.59	0.29

On the Module 2 test the number of items with good discrimination which were kept was ten. No good items were revised or omitted. Of the items rated satisfactory, 15 items were kept, three items revised, and two items omitted. In the poor category zero items were kept, nine items were revised and 11 items were omitted. Sixty-three additional items were needed in order to complete the 100 question item bank for Module 2.

After revisions were made and new items added the item banks were critiqued by one professor and three graduate students who are currently teaching or have previously taught basic clothing construction. Their suggestions were reviewed and additional revisions were made as needed.

Copies of the item banks were not included in the study since these items will be used in the future for the basic clothing construction course.

CHAPTER V

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

The purpose of this study was to evaluate the tests for Modules 1 and 2 of the clothing construction course and to determine needed revisions. An item analysis was conducted consisting of item difficulty and item discrimination power, number of responses to each alternative and the number of omitted responses, distribution of item difficulties, distribution of item discriminating power, and actual and recommended test statistics including mean, standard deviation, reliability, standard error measure, mean difficulty and mean discrimination. The data obtained were analyzed and used to determine those test items which were good, needed to be revised, or should be omitted.

A table of specifications was developed for each module - 1 and 2. Objectives were determined and content areas specified. The number of items for each type of objective and the percentage or emphasis placed on each content area were established.

The item banks were then developed using the item analysis data and the tables of specifications. Three graduate students and one professor who were currently teaching or had previously taught basic clothing construction critiqued the item banks. Their suggestions were used in making final revisions.

Conclusion

Results of the study indicated that with using the item difficulty and item discrimination values for each of the tests some of the items were good and should be kept, some needed revisions, and some should be omitted. Using the data in Table V, all tests - test 1, form 1; test 1, form 2; and test 2, form 1 had mean scores, reliability, and mean difficulty values which were higher than the recommended values. The standard deviation and mean discrimination values for both test 1, form 1 and test 2, form 1 were lower than recommended while the same values were higher on test 1, form 2. These data indicated that, all of the tests were generally good tests, however, some revisions were indicated on individual items.

Recommendations

The study was conducted to evaluate and make needed revisions of the tests for Modules 1 and 2 of the basic clothing construction course. Recommendations for further research and test development include:

1. Use the item banks developed for Modules 1 and 2 to construct two 50-item tests for each module representing the areas identified on the table of specifications in the established percentage. The tests should then be administered to students and an item analysis conducted to determine item difficulty and item discrimination for the items on the new tests.

2. Replicate the study for the tests used for Modules 3, 4, 5, and 6.

3. Develop additional items for the item banks for each module so that different forms of the test can be constructed.

4. Develop and implement advanced standing exams for Modules 3, 4, 5, and 6 of the basic clothing construction course.

SELECTED BIBLIOGRAPHY

Ahmann, J. S., & Glock, M. D. (1959). Evaluating pupil growth. Boston: Allyn and Bacon.

Ahmann, J. S., & Glock, M. D. (1981). Evaluating pupil growth. Boston: Allyn and Bacon.

Berry, J. C. (1963). The revision and development of a clothing pretest for appraising competencies of first year clothing students. Unpublished master's thesis, Oklahoma State University, Stillwater.

Bloom, B. S. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw Hill Book Co.

Davis, E. (1980). Teachers as curriculum evaluators. Sidney, Australia: George Allen and Unwin.

Ebel, R. L. (1979). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Furst, E. J. (1958). Measurement and evaluation in teaching. New York: Logmans, Green and Co.

Gay, L. R. (1980). Educational evaluation and measurement: Competencies for analysis and application. Columbus, OH: Charles E. Merrill Publishing.

Good, L. K. (1974). Feasibility of computer-generated testing via the cathode ray tube in a basic clothing construction course at Oklahoma State University. Unpublished master's thesis, Oklahoma State University, Stillwater.

Green, J. L., & Stone, J. C. (1977). Curriculum evaluation: Theory and practice. New York: Springer Publishing.

Gronlund, N. E. (1985). Measurement and evaluation in teaching. New York: MacMillan.

Gould, G. F. (1963). A performance pretest for placement of college students in beginning clothing courses. Unpublished master's thesis, Oklahoma State University, Stillwater.

Hall, O. A., & Paolucci, B. (1970). Teaching home economics. New York: John Wiley and Sons.

Hopkins, C. D., & Antes, R. L. (1978). Classroom measurement and evaluation. Itasca, IL: F. E. Peacock Publishers.

Lewy, A. (1977). Handbook of curriculum evaluation. Paris: Unesco.

Lisenby, D. S. (1979). Development of an item pool for advanced standing examinations in basic clothing construction. Unpublished master's thesis, Oklahoma State University, Stillwater.

Marshall, J. C., & Hales, L. W. (1971). Classroom test construction. Reading, MA: Addison-Wesley Publishing.

Miller, K. P. (1974). An identification of competencies in beginning clothing construction for college level. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.

Nunnally, J. C. (1964). Educational measurement and evaluation. New York: McGraw-Hill Book Co.

Posner, G. J., & Rudnitsky, A. N. (1978). Course design: A guide to curriculum development for teachers. New York: Longman.

Souligny, D. M. (1971). An evaluation of the clothing exemption test at Oklahoma State University. Unpublished master's thesis, Oklahoma State University, Stillwater.

Stufflebean, T. W. (1982). Development and implementation of a computerized advanced standing examination in basic clothing construction. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.

Walsh, G. M. (1959). The development of a pencil and paper pretest for placement of college students in first course in clothing, textiles, and merchandising at Oklahoma State University. Unpublished report, Oklahoma State University, Stillwater.

Webster's new collegiate dictionary. (1977). Springfield, MA: G. & C. Merriam Co.

Wilkins, W. E. (1971). An investigation of the feasibility of using computer generated tests in a clothing selection course at Oklahoma State University. Unpublished master's thesis, Oklahoma State University, Stillwater.

Witt, M. R. (1961). The revision and development of selected evaluation devices for appraising certain clothing competencies of college freshmen. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.

APPENDIX

DISTRIBUTION OF STUDENT RESPONSES

TABLE X

DISTRIBUTION OF STUDENT RESPONSES ON TEST 1 FORM 1 (N=157)

Item No.	Number of Responses[a] A	B	C	D	E	Omits	Item No.	Number of Responses[a] A	B	C	D	E	Omits
1	144	6	3	4	0	0	26	10	2	4	140	0	1
2	14	23	4	116	0	0	27	2	66	88	1	0	0
3	0	0	1	156	0	0	28	120	37	0	0	0	0
4	23	111	3	19	0	1	29	2	2	148	5	0	0
5	14	119	19	4	0	1	30	13	54	86	4	0	0
6	143	14	0	0	0	0	31	153	4	0	0	0	0
7	7	2	9	139	0	0	32	0	2	2	153	0	0
8[b]	3	145	5	4	0	0	33	147	3	1	5	0	1
9	137	14	3	3	0	0	34	49	12	77	19	0	0
10	18	20	1	118	0	0	35	8	6	6	137	0	0
11	7	2	5	143	0	0	36	141	8	3	5	0	0
12	5	147	3	2	0	0	37	126	14	5	12	0	0
13	4	1	1	151	0	0	38	17	39	89	12	0	0
14	19	121	5	12	0	0	39	9	11	38	98	0	1
15	4	103	35	15	0	0	40	18	18	53	68	0	0
16	12	82	45	17	0	1	41	13	7	132	5	0	0
17	30	127	0	0	0	0	42	30	127	0	0	0	0
18	23	59	41	33	0	1	43	120	4	1	32	0	0
19	15	11	16	115	0	0	44	88	68	0	0	0	1
20	9	4	131	13	0	0	45	2	2	14	139	0	0
21	5	126	12	14	0	0	46	12	123	14	7	0	1
22	6	148	0	3	0	0	47	113	24	18	2	0	0
23	11	52	26	68	0	0	48	3	1	136	17	0	0
24	1	2	152	1	0	1	49	28	129	0	0	0	0
25	0	14	15	127	1	0	50	12	73	2	70	0	0

[a]Correct responses are underlined.

[b]The correct response for item 8 was incorrectly marked on the key.

TABLE XI

DISTRIBUTION OF STUDENT RESPONSES ON TEST 1 FORM 2 (N=22)

Item No.	Number of Responses[a] A	B	C	D	E	Omits	Item No.	Number of Responses[a] A	B	C	D	E	Omits
1	2	2	0	18	0	0	26	2	7	3	10	0	0
2	4	15	2	1	0	0	27	2	17	3	0	0	0
3	7	4	0	11	0	0	28	1	0	14	1	6	0
4	0	4	2	15	0	1	29	2	6	2	2	10	0
5	0	0	1	21	0	0	30	1	18	1	1	1	0
6	12	10	0	0	0	0	31	13	0	3	6	0	0
7	2	0	19	1	0	0	32	2	1	0	0	19	0
8	20	1	1	0	0	0	33	11	0	4	7	0	0
9	1	1	18	2	0	0	34	8	7	7	0	0	0
10	0	16	2	4	0	0	35	4	16	2	0	0	0
11	1	3	2	16	0	0	36	3	0	17	2	0	0
12	0	16	4	2	0	0	37	5	1	0	9	7	0
13	1	1	7	13	0	0	38	14	5	2	1	0	0
14	15	2	5	0	0	0	39	2	2	3	13	2	0
15	5	16	1	0	0	0	40	6	0	3	9	4	0
16	0	2	20	0	0	0	41	21	0	1	0	0	0
17	0	1	19	2	0	0	42	5	14	1	1	1	0
18	15	4	0	3	0	0	43	13	3	2	1	3	0
19	1	8	0	13	0	0	44	5	0	13	3	1	0
20	0	1	1	20	0	0	45	2	2	0	17	1	0
21	0	2	1	19	0	0	46	0	2	19	1	0	0
22	2	2	1	17	0	0	47	1	19	0	2	0	0
23	0	17	2	3	0	0	48	0	1	21	0	0	0
24	14	7	1	0	0	0	49	1	2	2	17	0	0
25	14	7	0	1	0	0	50	3	0	1	17	1	0

[a]Correct responses are underlined.

TABLE XII

DISTRIBUTION OF STUDENT RESPONSES ON TEST 2 FORM 1 (N=109)

Item No.	Number of Responses[a] A	B	C	D	E	Omits	Item No.	Number of Responses[a] A	B	C	D	E	Omits
1	13	**96**	0	0	0	0	26	5	7	3	**94**	0	0
2	5	10	**93**	1	0	0	27	**74**	27	6	2	0	0
3	4	23	12	**70**	0	0	28	10	3	2	**94**	0	0
4	0	2	1	**105**	0	1	29	5	**78**	6	20	0	0
5	4	8	**90**	7	0	0	30	**91**	15	2	1	0	0
6	25	**84**	0	0	0	0	31	20	61	0	**28**	0	0
7	5	5	16	**83**	0	0	32	2	2	1	5	**99**	0
8	26	2	0	**80**	0	1	33	4	1	**103**	1	0	0
9	2	1	9	**96**	0	1	34	10	11	**81**	7	0	0
10	**92**	1	10	5	0	1	35	3	0	9	**91**	4	2
11	6	14	**87**	2	0	0	36	27	0	2	**80**	0	0
12	1	7	4	**97**	0	0	37	4	19	**70**	16	0	0
13	2	**106**	1	0	0	0	38	**26**	1	0	82	0	0
14	5	**95**	1	8	0	0	39	4	1	1	**102**	0	1
15	0	1	9	**98**	0	1	40	1	5	**101**	1	1	0
16	**99**	9	0	1	0	0	41	4	7	**94**	4	0	0
17	**73**	36	0	0	0	0	42	5	5	**98**	1	0	0
18	24	**84**	1	0	0	0	43	**69**	2	9	28	0	1
19	5	1	0	**103**	0	0	44	3	10	**90**	6	0	0
20	0	**108**	0	1	0	0	45	0	**13**	66	30	0	0
21	2	1	**104**	2	0	0	46	4	**94**	10	1	0	0
22	3	11	**22**	73	0	0	47	5	12	7	**84**	0	1
23	0	0	**107**	2	0	0	48	6	**56**	30	17	0	0
24	0	0	0	**109**	0	0	49	14	2	**75**	18	0	0
25	1	0	**106**	1	0	1	50	3	**98**	5	3	0	0

[a]Correct responses are underlined.

VITA

Cynthia Joy Kallenberger

Candidate for the Degree of

Master of Science

Thesis: EVALUATION AND REVISION OF SELECTED TESTS FOR A BASIC CLOTHING CONSTRUCTION COURSE

Major Field: Clothing, Textiles and Merchandising

Biographical:

Personal Data: Born in Borger, Texas, January 20, 1958; the daughter of Robert H. and Corinne J. Kallenberger.

Education: Graduated from Sooner High School, Bartlesville, Oklahoma, in May, 1976; received Bachelor of Science Degree in Vocational Home Economics Education from Oklahoma State University in July, 1980; completed requirements for the Master of Science degree at Oklahoma State University in December, 1986.

Professional Experience: Vocational Home Economics and Science teacher, Caney Valley Junior High School, Ochelata, Oklahoma, 1980-1983; Vocational Home Economics teacher, Stuart Public Schools, Stuart, Oklahoma, 1983-1985; graduate teaching assistant, Clothing, Textiles and Merchandising Department, Oklahoma State University, August, 1985 to present.

Professional Organizations: American Home Economics Association; Graduate Student Home Economics Association, Vice-President.