AN INVESTIGATION OF THE FEASIBILITY OF USING COMPUTER GENERATED TESTS IN A CLOTHING SELECTION COURSE AT OKLAHOMA STATE UNIVERSITY

By

WINELLEN E. WILKINS

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1970

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
July, 1971

AN INVESTIGATION OF THE FEASIBILITY OF USING COMPUTER GENERATED TESTS IN A CLOTHING SELECTION COURSE AT OKLAHOMA STATE UNIVERSITY

Thesis Approved:

Brovalyn Sisler
Thesis Adviser

Donice H. Kelly

D. Durham
Dean of the Graduate College

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. Lynn Sisler, Assistant Professor of Clothing, Textiles, and Merchandising for her guidance and encouragement throughout the course of this study; to Dr. Donice Kelly, Head of the Clothing, Textiles, and Merchandising department and Dr. Lora Cacy, Assistant Professor of Home Economics Education for their assistance and guidance in the writing of this report.

Acknowledgement is also extended to Mr. Bill Accola, Mr. David Pickering, and Miss Dena Meenen of the University Computer Center for their assistance in conducting the computer programming portion of this study. The author is also grateful to Dr. James Kirby, Head of the Religion department for the use of the test generation program developed for his classes.

Special appreciation is extended to the author's husband, Mel, for his patience, understanding, encouragement, and help during the entire course of this program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One of the most important aspects of teaching is interaction between the student and the teacher. The student must know how well he is doing and the teacher needs to know how the student is progressing and what help or learning experiences he needs. This interaction has almost vanished in many courses because of increasingly large enrollments. Along with this lack of interaction caused by large classes there is an increase in the amount of record keeping that must be done. Constructing and scoring tests and recording grades can become almost a full time job.

Many programs for individual study have been developed to give the teacher more time to work with those students who need help and to allow those students who do not need help opportunities to work ahead. This type program requires even more record keeping than the traditional methods. Detailed records must be kept up-to-date on the progress of each individual student. The instructor cannot do an adequate job of teaching if his time must be spent in preparing and scoring tests and keeping records on each student.

This problem can be partially overcome with the use of high speed computers. "One of the most important potential uses of computers in schools is their use to individualize the education process" (4). Computers can store information, generate and score tests, record grades

and other information, evaluate student progress and analyze tests more efficiently and accurately than can be done by hand. The computer center at Oklahoma State University has developed several programs for performing these various tasks. The purpose of this study was to determine the feasibility of using the computer for generating tests for the basic clothing selection course at Oklahoma State University.

Objectives

The objectives of the study were:

1. To evaluate one examination used in the clothing selection course.
2. To rewrite the examination for use on the computer.
3. To evaluate the computer generated test and compare the results with those of the original test.

Definition of Terms

1. Item analysis -- a comparison of the performance of students in the upper 27 per cent of the class with the performance of students in the lower 27 per cent of the class.
2. Difficulty level -- the percentage of students who answer an item correctly.
3. Discrimination index -- the ability of a test item to indicate the difference between those students in the upper 27 per cent of the class and those students in the lower 27 per cent of the class.
4. Distractors -- those choices on a multiple-choice item which are incorrect.

5. Computer generated test -- a test composed of items randomly selected from an item pool and printed by a computer.
6. Item pool -- the list of test items from which the computer selects items for a test.
7. Program -- the written instructions fed into the computer to designate the procedure to be followed.

Limitations of the Study

The study was limited to the modification of one examination for use on the computer. The oirignal form of the test was given to 225 students enrolled in the basic clothing selection course for the fall semester of 1970. During the spring semester of 1971, the computer generated form of the test was given to 141 students. No attempt was made to randomize the sample since all students in the course each semester took the test.

CHAPTER II

BACKGROUND FOR THE STUDY

The first step in developing a course is to determine what should be taught. According to Brown (3) no teaching effort will be effective unless it is directed toward a specific goal or objective which is understood and accepted by both the teacher and the students. Likewise, no evaluation will be meaningful unless it is planned according to the objectives of the course.

Objectives should be stated in terms of the behavior changes expected to occur as a result of the learning experience. Tyler (15) states that the most useful objectives are those that state what types of behavioral changes are expected and in what content area these changes should take place.

According to Mager (11):

> A meaningfully stated objective is one that succeeds in communicating your intent; the best statement is one that excludes the greatest number of possible alternatives to your goal . . . . First, identify the terminal behavior by name; we can specify the kind of behavior which will be accepted as evidence that the learner has achieved the objective. Second, try to further define the desired behavior by describing the important conditions under which the behavior will be expected to occur. Third, specify the criteria of acceptable performance by describing how well the learner must be able to perform to be considered acceptable. But though each of these items might help an objective to be more specific, it will not be necessary to include all three in each objective. The object is to write objectives that communicate; the characteristics described above are merely offered as guides to help you know when you have done so.

Construction of Tests

Properly written objectives serve as a guide to determine the type of evaluation or examination to be used since the main purpose of any type of examination is to determine how much the student has learned in relation to the course objectives. The examination helps the teacher and the student determine areas of weakness or problems which the student may be having in certain content areas.

Tests may also motivate.

> Tests that are well constructed and effectively used can motivate students to develop good study habits, to correct errors, to direct their activities toward the achievement of desired goals . . . . Testing procedures control the learning process to a greater degree perhaps, than any other teaching device. (14)

Hall and Paolucci (8) list the following guidelines for well constructed test items:

1. Select a type of item that is best suited for the content and specific objective you wish to measure.
2. Use items that require the students to apply their learnings not merely to recall or reorganize information.
3. Select items that provide new situations in which the students can test their ability to apply their learnings.
4. Make the entire content of an item homogeneous and plausible, so that the student will have to think before determining an answer.
5. The content of an item should determine the correctness of an answer.
6. Make items short and definite, including only one independent idea in each question.
7. Word the items simply, using language familiar to the students.
8. Whenever possible, select items that include more than two choices to reduce the possibility of guessing.
9. Make sure there is only one correct answer, unless the directions indicate some other procedure.
10. Use correct grammar, and do not give irrelevant clues to the correct answer.
11. Be sure to clarify any words that are qualitative or that have hidden meanings if such words are used.

In his analysis of teacher made tests, Degenhardt (5) states that the objective test is a flexible and valuable type of examination

> since it can be used to test for the retention of facts, the grasp a student has of subject matter in his text book, the relationship between different bodies of knowledge, the application of facts learned to new situations, and the ability to use skills involved in scientific problems.

Two forms of the objective test are the multiple-choice and the true-false test. The following criteria for multiple-choice items are given by Hedges (9).

1. All options should be grammatically consistent.
2. Make the stem long and the distractors or options brief.
3. All extraneous material should be excluded from the stem.
4. The stem of each test question should contain a central problem.
5. Double negatives should be avoided.
6. The optimal number of options is five.
7. All distractors should be plausible.
8. The more homogeneous the options, the higher the level of understanding required.
9. The correct response should not be consistently longer or shorter than the decoys.
10. Only one of the options should be the correct or 'best' answer.
11. Whenever an item reflects a controversial opinion, authority should be cited.
12. State the original multiple-choice item in question form.

Hall and Paolucci (8) suggest the following rules for writing true-false test items.

1. Restrict each statement to one central idea.
2. Make the point of the question clear. Avoid 'trick' questions.
3. A false element should be part of the reason rather than the basic part of a statement.
4. In general, avoid the use of 'specific determiners' which make an answer obvious.
5. Use quantitative rather than qualitative language.
6. Keep the true and false sentences approximately the same length.
7. Approximately half of the statements should be true and half false, with these responses distributed at random so as not to form a regular sequence or pattern.

Examinations are used to determine how well the students are achieving the objectives of the course. An examination can also diagnose weaknesses, provide practice for knowledge and skills, motivate, and control the learning process (14).

Item Analysis

Examinations must be continuously evaluated and improved if they are to function effectively. Various types of evaluations can be used on examinations to determine their effectiveness and any improvements that may be needed. The clearest, most objective type of evaluation is the statistical item analysis (6).

An item analysis compares the performance of students scoring in the upper portion of the class with the performance of students scoring in the lower portion of the class. As scores approach the mean score for the test, the differences between the two groups decrease. If the scores around the mean are discarded, the differences between the two groups become more distinct. Kelly (10) found that using the upper and lower 27 per cent of the scores yields the maximum reliability.

A statistical item analysis yields two values for each item on a test. The item difficulty is found by dividing the total number of students in each group who answered the item correctly by the total number of students in each group attempting the item. An acceptable item has a difficulty level of 0.40 to 0.70 (or 40% to 70%). An item with a difficulty level of 0.70 or above is too easy since a high percentage of the students chose the correct answer. An item is too difficult if it has a difficulty level of 0.40 or below (2).

The discrimination index is an indication of how well the item discriminates between the students in the upper portion of the class and the students in the lower portion of the class. This value is found by using the following formula.

$$DI = \frac{Ur - Lr}{N}$$

DI - discrimination index

Ur - number of students in the upper group answering the item correctly

Lr - number of students in the lower group answering the item correctly

N - number of students in either group

The value for the discrimination index may vary from +1.00 to -1.00. A value of +1.00 to +0.40 is considered good, from +0.40 to +0.20 is fair, and from +0.20 to 0.00 poor. A negative value means that more students in the lower group than in the upper group answered the item correctly (2).

An item analysis also indicates the number of students selecting each incorrect answer for each item. With this information it is possible to determine which test items need to be changed or eliminated and which distractors are ineffective. Weakness in certain areas of subject matter may also be determined from the results of an item analysis.

Computer Generated Tests

Colleges and universities ranging in size from less than 1,000 students to more than 20,000 students are using the computer for many tasks

that were formerly done by hand. One area in which the computer has proven useful is that of testing. The computer can reduce the time consuming task of scoring tests and recording grades to a matter of seconds. Summaries of tests can be obtained to indicate not only the performance of the students, but also the effectiveness of test items in indicating student ability.

Computers are also being used to generate tests. The use of computer generated tests allows each student to take a different form of a test. Cheating is kept to a minimum since every test is different. Test security is improved because the entire file is stored on the computer and only the instructor has access to it.

Testing in courses in which the student is allowed to retake tests is improved by using computer generated tests because the student can take a different form of the test over the same material. Once the test file has been created, time is saved in constructing different tests since only a few seconds are required for the computer to select the correct number of test items and print them.

Computer generated tests can be used in any subject area. The programs may be complex ones in which the computer actually writes the test items or relatively simple ones in which the tests are generated from files of teacher written test items.

Osburn and Shoemaker (12) conducted an experiment at the University of Houston to determine the operational feasilibity of using computer generated test items in an elementary statistics course. The criteria used for determining the feasibility of using the computer were:

1. The statistical characteristics of the computer generated items as compared with teacher made tests.

2. Student reactions to the computer items.

3. Experience gained in developing and implementing the procedure.

A set of item stems was developed for use on the computer. A group of endings for each stem was also developed. The computer randomly selected a designated number of items. One ending was randomly selected from the group of endings for each stem.

Examinations containing both computer made test items and teacher made test items were administered to an elementary statistics class over a two semester period. The computer generated items were found to be slightly less reliable than the teacher made items, but not unacceptably so. The students considered the computer made items to be more difficult than teacher made items but equally fair. While some students expressed difficulty in reading the computer printout, the overall reaction to the computer generated test was favorable.

A computer program for writing spelling tests was developed at Washington University "to combat difficulties of dictated spelling tests such as unreliable scoring due to illegible writing and the possibility of clues being provided through the enunciation of words by the examiner" (7).

The objectives of the study were (1) to develop a machine scorable test in which randomly ordered items can be answered by a true-false or forced-choice form, (2) to consider the diagnostic function of spelling scales through the use of specific error categories in test construction, and (3) to investigate whether grouping of words, similarly misspelled into items representing distinct error categories, would facilitate discrimination of the correctness or incorrectness of a word's spelling.

Words for the item pool were taken from several books listing frequently misspelled words. Each word was listed on a 3x5 card along with at least one common misspelling, difficult letters underlined, the number of letters in the correct spelling, and the type of error in the misspelled word. This information was punched into cards and fed into the computer.

The following specifications were used to generate tests from this item pool:

1. Each test consisted of 55 items.
2. Each item contained four words randomly selected from the same error category.
3. One item contained either a correct or an incorrect spelling of each of the four words in that item.

During the fall quarter of 1968, three hundred and thirty-five high school students took the computer generated spelling test on the University of Washington campus. Five different forms of the test were used. The tests were graded by words and by items and the scores in each category were compared within each of the five forms of the test.

High correlations among categories for erroneously adding, subtracting, and substituting letters suggest that these categories are similar. Low correlations between inverting letters and the other categories suggest that this is a different process from the others. Inverting letters is the type of error most difficult to discriminate, while substituting letters is the form of error easiest to discriminate.

Some studies have been conducted in which the actual performance of the computer was simulated. However, the results are the same as if the computer had been used. One such study was done by Anastasio (1) in

1969. He conducted an investigation to "identify the properties which characterize 'good' test items for measuring verbal ability and to develop rules for coding words and sentences so that they can be manipulated by the computer."

Sentence completion items were used. This type of item consisted of a sentence from which one or more words were missing. The sentence was completed by choosing the correct word or group of words from a list of five choices. Items used in the study were compound or complex sentences chosen from newspapers, magazines, and books.

Each item was reviewed by seven item writers to determine its appropriateness for use, and the ease of writing good distractors. After the items were accepted by the seven item writers they were used in a pretest and analyzed. Those items having a high degree of discrimination were analyzed for similarity in structure.

One hundred and ten "good" sentences were characterized by the following variables:

1. Sentence length between 18 and 28 words
2. Presence of internal punctuation
3. Use of a subordinating and/or coordinating conjunction
4. At least two prepositional phrases
5. At least one terminal prepositional phrase in a dependent clause

The first four characteristics could easily be incorporated into a computer program. To test the effectiveness of this type of program, a person naive to the research was given a set of the rules mentioned above, a book on conjunctions and prepositions and a randomly selected copy of Saturday Review. From an essay in the magazine, he found 26

sentences that matched four of the five variables.

Sixteen of these items were randomly eliminated and replaced by sixteen items already tested and found to be good. These 26 items were reviewed by seven item writers and ranked as either good or bad. Using the Chi square with one degree of freedom, no significant difference was found between the sentences chosen from the magazine and those already tested and found to be good.

Richards (13) conducted a similar study to develop a computer program for writing the verbal comprehension portion of a college entrance examination.

> Since no computer with a large supplementary memory device was available, the actual operation of the computer was simulated. The simulation was rigorous, however, and the items corresponded exactly to what would be written by a computer.

Item stems were words randomly selected from a book. A list of synonyms for each word was compiled. One synonym was randomly selected from the list of alternatives for each word. Stems were classified according to Roget's Thesaurus classification scheme and distractors for each item were randomly selected from the classifications adjacent to that of the stem. If a stem was more than one part of speech the part of speech to be used was selected before the proper correct alternative and distractors were selected.

A test of 72 items with 4 alternatives each was developed. The test, together with the Wide Range Vocabulary Test, was given to entering Freshmen at the University of Iowa in 1965. At the end of the semester, the grade point average for each of the students was obtained. The means, standard deviations, predictive validities, intercorrelations, and Kuder-Richardson 21 reliabilities for the two tests were determined.

The "computer written" test items were easier and less reliable than those on the Wide Range Vocabulary Test; however, the validities were comparable and the intercorrelations of the two were not far from the limits set by reliability. The predictive validities of the "computer written" test were slightly higher than that of the Wide Range Vocabulary Test.

These studies show that computer generated tests can be used successfully in various types of subject matter. A computer program for generating tests can be developed for almost any type of classroom situation. However, since little has been done with computer generated tests in home economics, more experimentation in this area is needed.

CHAPTER III

DEVELOPMENT AND ADMINISTRATION OF THE TEST

An investigation was conducted to determine the feasibility of using the computer to generate tests for the beginning clothing selection course at Oklahoma State University.

The steps followed in the study were:

1. To administer the original test and obtain an item analysis by using the computer.
2. To rewrite the test and store it on a computer tape deck.
3. To generate several individual tests.
4. To administer the computer generated tests and obtain an item analysis.
5. To compare the results of the two tests.
6. To administer a questionnaire to all students taking the computer test.
7. To estimate the time and money needed to implement such a program.

The examination over the unit on the Acquisition and Use of Clothing was used for the study. The original teacher made test for this unit consisted of seventy multiple-choice and true-false items. This test was administered to 225 students enrolled in the clothing selection course during the fall semester of 1970. Answers were marked on

separate answer sheets for scoring by machine. An analysis of the test was made by computer.

The graded answer sheets were arranged in order from the highest to the lowest score. One computer card was punched for each test. The card contained an identification number, the answers indicated by the student and the grade for that test.

A frequency count of the number of times each distractor was chosen was made on all of the tests and on the upper and lower 27 per cent of the tests. The computer was also used to calculate the mean and standard deviation for the test and the difficulty level and discrimination index for each item.

The items were ranked in order by difficulty level and by discrimination index. Appendix A, page 31, shows the difficulty level and discrimination index of each item and rates each item as having a good or easy difficulty level and a good, fair, or poor discrimination index. Those items that fell within the acceptable range of difficulty and discrimination were set aside. The remaining items were reviewed to determine the changes needed to improve the test. Changes in either the stem or the distractor were made in 32 of the items. Two items were eliminated, leaving 68 items for the computer test.

The computer program used required that items first be classified according to subject area and then according to objective within a subject area. There were five major subject areas with a total of 13 objectives for this unit. Appendix B, page 36, shows the objectives by subject area for this unit.

The test items were punched into computer cards according to the following format (see Figure 1).

1. The beginning of an item stem was indicated by an asterisk (*) in the first column of the first card for the stem. The correct answer was punched in the second column of the card and the stem was punched in the next 70 columns. The stem may be any length as long as the first column of the first card contains an asterisk.
2. The first column of a distractor contained a pound sign (#) followed by the distractor in the next 71 columns. The distractor may also be any length as long as the first column of the first card contains a pound sign.
3. A code number was punched in columns 73-80 of each card. Columns 73 and 74 indicate the subject area of the item. Columns 75 and 76 indicate the objective tested by the item. Columns 77 and 78 are the item number under that objective, and columns 79 and 80 are the card number within the item.

The test items were transferred by computer from the cards to a tape deck stored in the computer center. As a security check, the test cannot be retrieved from the tape deck until the computer has been fed the proper code word. Only the instructor using the tape knows this code word. The test items look like this when printed by the computer.

```
1)  ACCORDING TO HORN, IF WE WERE TO DRAW A SINGLE CONCLUSION FROM OUR
    STUDY OF FASHION, WHAT WOULD IT BE?
    1  FASHION MOVES THROUGH A COMPLETE CYCLE EVERY 50 YEARS.
    2  FASHION IS CHANGE AND CHANGE IS INEVITABLE.
    3  FASHION WILL INCREASE IN IMPORTANCE DURING THE NEXT DECADE.
    4  SKIRT LENGTHS WILL NEVER AGAIN BE AS SHORT AS IN 1969-1970.
```

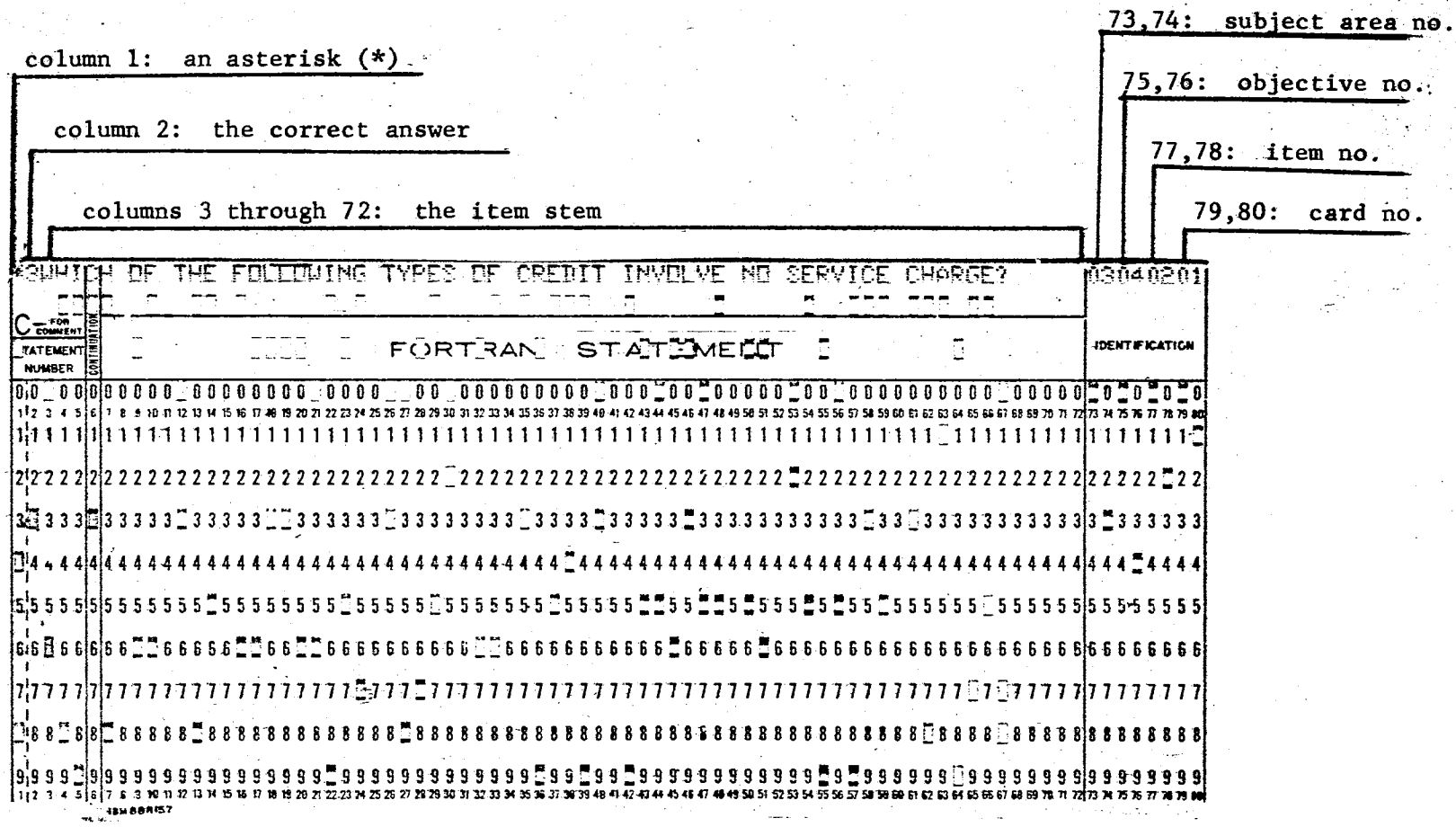


Figure 1. Sample Card for an Item Stem

The test generating program used will print any number of designated test items, any number of randomly selected items, or a combination of both. A set of program cards was punched to indicate the number of tests to be generated, the number of items on each test, which specific items were to appear on all tests, and the number of items to be randomly selected from each objective within each area.

Thirty different tests and their corresponding keys were generated for this study. Each test was composed of three specified items and 32 randomly selected items. Special answer sheets were developed to facilitate hand scoring. The pattern of item selection for the thirty tests is illustrated in Appendix C, page 39. Items 1, 2, and 68 were specified to appear on all of the tests. The number of items randomly selected from the remaining objectives depended upon the total number of items in each objective. Approximately one half of the items in each objective appeared on each test. While several items appeared on almost all of the tests, one item was never selected.

The computer generated tests were administered to 141 students during the spring semester of 1971. After taking the test, each student was given a questionnaire to determine his reaction to the computer test. A sample of the questionnaire is given in Appendix D, page 41. While the student completed the questionnaire, his test was graded. The corrected answer sheet and test were shown to the student for review.

A summary of student responses to the questionnaire is given in Table I. Of the 141 students completing the questionnaire, 123 (87 per cent) of them preferred the computer written test to a teacher written test. Sixty-three per cent indicated that they would prefer marking their answers on the test rather than on a separate answer sheet.

TABLE I

SUMMARY OF RESPONSES TO THE QUESTIONNAIRE

Question	Number	Per Cent*
Which do you prefer?		
- A teacher written test	11	8%
- A computer written test	123	87%
Where would you prefer to mark your answers?		
- On the test	90	64%
- On a separate answer sheet	43	31%
What did you like the most about the test?		
- The test is easy to read and the answer sheet is easy to use	43	31%
- The test was easy to understand	27	19%
- The test can be graded immediately	19	14%
What did you like the least about the test?		
- The length of the computer printout pages	57	40%
- Using a separate answer sheet	14	10%

*Percentages do not equal 100 per cent because many students failed to respond to all of the questions.

In response to the open end questions, 31 per cent of the students indicated that they felt the computer written test was easier to read than teacher written tests and the answer sheets were easier to use than the machine graded answer sheets. The fact that the test could be graded immediately was favored by many students. Forty per cent of the students expressed difficulty in handling the long pages of the computer printout.

The computer generated test was analyzed using the same computer program that was used on the teacher made test. A list of all items and their difficulty level and discrimination index is given in Appendix A, page 33. The computer test did not seem to have been improved by the changes made as a result of the analysis of the original test. However, an accurate comparison of the two tests could not be made. The number of students responding to each item on the computer test varied. Therefore, an item on the computer test answered by only fifty students cannot be compared with an item on the original test answered by 225 students. The means, standard deviations, and highest and lowest scores for the two tests were similar (see Table II).

Although one of the purposes of using the computer is to save time and expense, the initial development of a computer program requires a great deal of both time and money. A program needs to be run several times before it is perfected. Since this particular test generation program was a new one, an estimate was made of the amount of time and money needed to implement the program.

Tables III and IV give a breakdown of the time and expense involved in developing the program and carrying out the study. The time spent developing the program included the time required for a professional

TABLE II

MEAN AND STANDARD DEVIATION OF THE TWO TESTS

Test*	Mean	Standard Deviation	Highest Score	Lowest Score	Range
1	59.21	6.17	70	41	29
2	58.88	6.67	70	34	36

*Test 1 - the teacher made test; Test 2 - the computer generated test

TABLE III

TIME REQUIRED TO DEVELOP AND IMPLEMENT THE TEST GENERATION PROGRAM

Process	Time
Keypunching cards for the item analysis of the teacher made test	12 hours
Keypunching cards for the item analysis of the computer generated test	9 hours
Developing the original program	40 hours
Testing and correcting the program	25 hours
Analyzing test data and creating the test file	27 hours
Analyzing the test data on the computer generated test	6 hours
Total	119 hours

TABLE IV

COST OF DEVELOPING AND IMPLEMENTING THE TEST GENERATION PROGRAM

Process	Cost
Analyzing the teacher made test by computer	$ 5.98
Analyzing the computer generated test by computer	$ 5.63
Storing the test on a tape deck	$13.21
Generating the tests	$44.92
Total	$69.74

programmer to write the test generation program and the time spent by the researcher testing the program and making changes necessary before it could be used successfully. As was mentioned earlier, a program may undergo several unsuccessful trials before it is perfected.

All of the keypunching was done by the researcher. If a professional keypuncher had been employed the time required might have been decreased but the cost would have increased. Once the test file has been created and placed on a tape deck, the only keypunching that must be done is that required for an item analysis of the test each time it is given. The time needed for analyzing the test data decreased also, once the test file had been created, since this included time spent rewriting items in a form acceptable for the computer.

The major expense was that of generating the tests. Part of this expense was involved in perfecting the program. Once the program had been perfected, the cost per test, for generating the test, was about $1.06. This was a greater expense than was expected. Generating tests for every student in the class would require an unjustifiable expense. However, when a few tests are used over and over again as test booklets, as was done in this study, the cost might not be too great. The purpose of having individual tests, however, would be partially defeated.

Some test generation programs have been developed that are less expensive than the one used in this study. The use of one of these programs might alleviate the cost problem in a course based on individual study and testing.

CHAPTER IV

SUMMARY AND RECOMMENDATIONS

A study was conducted at Oklahoma State University to determine the feasibility of using the computer to generate tests for the basic clothing selection course. The study was limited to the modification of one unit examination for generation by the computer.

The original teacher made test for the unit was given to 225 students during the fall semester of 1970. The test was analyzed by computer and items ranking outside the acceptable limits set for difficulty and discrimination were revised.

The test items were stored on a computer tape deck and 30 different forms of the test, having 35 questions each, were generated from items randomly selected from the original 68 items. The computer generated test was given to 141 students during the spring semester of 1971. An item analysis was made on this test as on the original test. All students taking the computer generated test were asked to fill out a questionnaire concerning their reaction to the test. An estimate was made of the time and money required to develop and use the program.

Because of the difference in the number of students responding to each item on the computer test and because of the difference in the number of items given to each student on the two tests, no comparison was made between the two tests. The means, standard deviations, and highest and lowest scores of the two tests were quite similar.

The majority of the students preferred the computer written test to a teacher written test because they felt it was easier to read and because it could be graded before they left the classroom. The students also preferred the special answer sheets that were made over the machine graded answer sheets. Several students expressed a preference for marking their answers on the test rather than on the separate answer sheet. The major complaint about the test was that the long computer pages were difficult to handle.

Some difficulty was encountered in grading the test in the classroom when more than ten students were taking the test at the same time. Since the majority of the students completed the test at about the same time, the area around the grader's desk became congested. Consultation with each student was also difficult when several other students were standing around waiting for their test papers to be graded. This problem was eliminated by requiring students to bring their papers to be graded one at a time.

The average cost per individual test was $1.06. This was the cost of generating the test and did not include the cost of grading and analyzing. The expense is too great to justify generating individual tests for every student. When a few tests are generated and used as test booklets, as was done in this study, the cost is decreased but part of the advantage to having individual tests is defeated.

Recommendations

1. If this particular computer generated test is to be used in the future, further analysis should be done each time the test is administered.

2. The possibilities of allowing students to take a test at any time rather than on designated examination days as is done now should be investigated.
3. Other test generation programs should be investigated and a cost comparison made to develop a less expensive program for use with this course.

BIBLIOGRAPHY

(1) Anastasio, Ernest J. "Computer Assisted Item Writing." (paper presented to the National Council on Measurement in Education, Los Angeles, California) February, 1969.

(2) Ahmann, J. Stanley and Marvin D. Glock. Evaluating Pupil Growth. Boston: Allyn & Bacon, Inc., 1959.

(3) Brown, Clara M. Evaluation and Investigation in Home Economics. New York: F. S. Crofts and Co., 1941.

(4) Cooley, William W. and Robert Glasser. "The Computer and Individualized Instruction." Science, 166 (October, 1969), 574-582.

(5) Degenhardt, Edward F. "Practical Analysis of Teacher-Made Objective Test Questions." American Biology Teacher, 27 (December, 1965), 755-758.

(6) Ebel, Robert L. "Procedures for the Analysis of Classroom Tests." Educational and Psychological Measurement, 14 (Summer, 1954), 352-364.

(7) Gibson, Susann E. "Computer-Aided Construction of Categorized Spelling Tests." (research report by candidate for degree in Psychology at Washington University, Seattle, Washington, Bureau of Testing.) June, 1969.

(8) Hall, Olive and Beatrice Paolucci. Teaching Home Economics. New York: John Wiley and Sons, Inc., 1967.

(9) Hedges, William D. "How to Construct a Good Multiple-Choice Test." Clearing House, 139 (September, 1964), 8-11.

(10) Kelly, Truman L. "The Selection of Upper and Lower Groups for the Validation of Test Items." The Journal of Educational Psychology, 30 (January, 1939), 17-24.

(11) Mager, Robert F. Preparing Objectives for Programmed Instruction. San Francisco: Fearon Pub., 1962.

(12) Osburn, H. G. and D. M. Shoemaker. "Pilot Project on Computer Generated Test Items." (research conducted at the University of Houston, sponsored by the Office of Education, Department of Health, Education, and Welfare) June, 1968.

(13) Richards, James M. Jr. "Can Computers Write College Admissions Tests?" Journal of Applied Psychology, 51 (June, 1967), 211-215.

(14) Thorndike, Robert L. and Elizabeth Hagen. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1961.

(15) Tyler, John Ralph. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press, 1950.

APPENDIX A

ITEM ANALYSIS DATA FOR THE TEACHER MADE TEST AND THE COMPUTER GENERATED TEST

DIFFICULTY LEVEL AND DISCRIMINATION INDEX
OF THE TEACHER MADE TEST

Item No.	Difficulty Level	Rate	Discrimination Index	Rate
No. 1	58.15%	good	0.2456	fair
No. 2*	89.43%	easy	0.1475	poor
No. 3	85.99%	easy	0.3770	fair
No. 4	61.23%	good	0.3442	fair
No. 5*	94.27%	easy	0.1147	poor
No. 6*	95.59%	easy	0.0819	poor
No. 7*	89.82%	easy	0.1967	poor
No. 8	85.90%	easy	0.3114	fair
No. 9	60.35%	good	0.5409	good
No. 10	53.30%	good	0.5409	good
No. 11	83.26%	easy	0.2131	fair
No. 12*	95.15%	easy	0.0983	poor
No. 13*	88.11%	easy	0.1475	poor
No. 14*	94.15%	easy	0.0163	poor
No. 15	82.38%	easy	0.2457	fair
No. 16	56.39%	good	0.5081	good
No. 17	79.74%	easy	0.2622	fair
No. 18	83.70%	easy	0.2786	fair
No. 19	55.51%	good	0.3278	fair
No. 20*	89.43%	easy	0.1967	poor
No. 21	44.93%	good	0.2131	fair
No. 22	76.11%	easy	0.4590	good
No. 23*	99.12%	easy	0.0163	poor
No. 24*	97.80%	easy	0.0655	poor
No. 25	89.87%	easy	0.2457	fair
No. 26*	95.59%	easy	0.0655	poor
No. 27	82.82%	easy	0.2457	fair
No. 28*	98.68%	easy	0.0000	poor
No. 29*	100.00%	easy	0.0000	poor
No. 30*	96.04%	easy	0.0491	poor
No. 31*	99.56%	easy	0.0163	poor
No. 32*	96.48%	easy	0.0819	poor
No. 33*	98.14%	easy	0.0327	poor
No. 34*	94.27%	easy	0.0983	poor
No. 35	84.14%	easy	0.4262	good
No. 36	86.78%	easy	0.4754	good
No. 37	82.82%	easy	0.3114	fair
No. 38	79.30%	easy	0.2950	fair
No. 39	88.99%	easy	0.2131	fair
No. 40*	98.68%	easy	0.3027	poor
No. 41	85.90%	easy	0.2950	fair
No. 42	82.38%	easy	0.2786	fair
No. 43*	99.56%	easy	0.0163	poor
No. 44	67.84%	good	0.2295	fair
No. 45	51.54%	good	0.1475	poor
No. 46	74.01%	easy	0.5409	good
No. 47*	96.46%	easy	0.1147	poor

(Continued)

Item No.	Difficulty Level	Rate	Discrimination Index	Rate
No. 48	83.26%	easy	0.2622	fair
No. 49	74.89%	easy	0.4426	good
No. 50*	92.95%	easy	0.1475	poor
No. 51	88.99%	easy	0.2457	fair
No. 52*	87.67%	easy	0.1803	poor
No. 53*	98.24%	easy	0.0327	poor
No. 54	81.50%	easy	0.2622	fair
No. 55	89.38%	easy	0.3278	fair
No. 56	77.09%	easy	0.4590	good
No. 57*	99.12%	easy	0.0327	poor
No. 58*	96.48%	easy	0.0491	poor
No. 59*	100.00%	easy	0.0000	poor
No. 60*	86.78%	easy	0.1639	poor
No. 61*	99.56%	easy	0.0163	poor
No. 62*	96.04%	easy	0.0491	poor
No. 63	69.60%	good	0.5573	good
No. 64	65.64%	good	0.3114	fair
No. 65*	95.59%	easy	0.1311	poor
No. 66	63.11%	good	0.3770	fair
No. 67*	99.56%	easy	0.0491	poor
No. 68*	95.11%	easy	0.1475	poor
No. 69	65.18%	good	0.5409	good
No. 70*	91.96%	easy	0.1803	poor

*Those items that were rewritten or replaced

DIFFICULTY LEVEL AND DISCRIMINATION INDEX
OF THE COMPUTER GENERATED TEST

Item No.	Difficulty Level	Rate	Discrimination Index	Rate
No. 1	75.89%	easy	0.4054	good
No. 2	63.83%	good	0.4054	good
No. 3	00.00%	****	0.0000	****
No. 4	89.00%	easy	0.0541	poor
No. 5	73.33%	easy	0.1622	poor
No. 6	68.75%	good	0.2432	fair
No. 7	71.05%	easy	0.2432	fair
No. 8	73.33%	easy	0.1892	poor
No. 9	89.00%	easy	0.1892	poor
No. 10	84.09%	easy	0.1351	poor
No. 11	94.52%	easy	0.0811	poor
No. 12	75.00%	easy	0.1081	poor
No. 13	83.52%	easy	0.1351	poor
No. 14	52.17%	good	0.0270	poor
No. 15	85.71%	easy	0.2162	fair
No. 16	43.02%	good	0.2432	fair
No. 17	90.00%	easy	-0.0811	poor
No. 18	98.25%	easy	-0.0270	poor
No. 19	96.00%	easy	0.0541	poor
No. 20	97.87%	easy	0.2701	fair
No. 21	80.65%	easy	0.0270	poor
No. 22	89.71%	easy	0.1351	poor
No. 23	96.84%	easy	0.1622	poor
No. 24	97.56%	easy	0.0811	poor
No. 25	87.18%	easy	-0.0270	poor
No. 26	94.52%	easy	0.0270	poor
No. 27	97.78%	easy	-0.1081	poor
No. 28	97.12%	easy	-0.0541	poor
No. 29	95.00%	easy	0.1892	poor
No. 30	55.56%	good	0.1622	poor
No. 31	88.46%	easy	0.0541	poor
No. 32	69.44%	good	0.1081	poor
No. 33	40.30%	good	0.1892	poor
No. 34	91.21%	easy	0.0270	poor
No. 35	90.38%	easy	0.3243	fair
No. 36	82.80%	easy	0.1351	poor
No. 37	66.67%	good	0.1081	poor
No. 38	77.03%	easy	0.2432	fair
No. 39	76.79%	easy	0.2432	fair
No. 40	79.12%	easy	-0.0270	poor
No. 41	93.44%	easy	0.1892	poor
No. 42	91.04%	easy	0.0541	poor
No. 43	91.76%	easy	0.0541	poor
No. 44	92.00%	easy	0.1892	poor
No. 45	87.50%	easy	0.0541	poor
No. 46	81.58%	easy	0.1081	poor
No. 47	66.10%	good	0.2701	fair

(Continued)

Item No.	Difficulty Level	Rate	Discrimination Index	Rate
No. 48	77.27%	easy	0.4054	good
No. 49	100.00%	easy	-0.0541	poor
No. 50	86.99%	easy	0.1081	poor
No. 51	77.03%	easy	0.1351	poor
No. 52	80.60%	easy	0.1341	poor
No. 53	81.36%	easy	0.0811	poor
No. 54	83.58%	easy	-0.0541	poor
No. 55	96.30%	easy	0.0811	poor
No. 56	93.14%	easy	0.0541	poor
No. 57	97.22%	easy	-0.0270	poor
No. 58	66.67%	good	-0.0270	poor
No. 59	98.84%	easy	-0.0541	poor
No. 60	77.78%	easy	0.1081	poor
No. 61	88.54%	easy	0.3874	fair
No. 62	83.91%	easy	0.1081	poor
No. 63	82.73%	easy	0.1892	poor
No. 64	82.98%	easy	0.1081	poor
No. 65	66.22%	good	0.3514	fair
No. 66	82.22%	easy	-0.1081	poor
No. 67	86.44%	easy	-0.0541	poor
No. 68	97.14%	easy	0.0270	poor

**** Items which did not appear on any of the computer generated tests

APPENDIX B

OBJECTIVES FOR THE UNIT

OBJECTIVES ARRANGED BY SUBJECT AREA

Wardrobe Planning

Objectives:

1. The student will be able to make an inventory of his present wardrobe according to the form provided.

2. The student will be able to analyze his present wardrobe in regard to (1) adequacy for his activities, (2) condition of his clothes, and (3) color coordination.

3. The student will be able to make a plan for future clothing purchases, based on a realistic budget, incorporating ideas presented in this unit with present wardrobe.

Fit of Clothes

Objectives:

1. The student will be able to differentiate between garments which fit properly and those which fit improperly and to recognize various fitting problems.

Responsibilities of the Consumer

Objectives:

1. The student will be able to recognize responsibilities of the clothing consumer.

2. The student will be able to list factors to consider when buying ready-made garments.

3. The student will be able to apply criteria for effective purchases of clothing at sales.

4. The student will be able to recognize the advantages and disadvantages of using various types of credit for clothing purchases.

Characteristics of Fibers, Fabrics, and Finishes

Objectives:

1. The student will be able to recognize major characteristics, performance expected, and care required of the following fibers: cotton, silk, linen, wool, polyester, acrylic, nylon, and rayon.

2. The student will be able to describe the characteristics and care required of the following blends: cotton-polyester, wool-nylon, wool-rayon, and rayon-acetate.

3. The student will be able to differentiate among the following types of fabric construction and to identify characteristics of each: plain weave, twill weave, satin weave, single knit, and double knit.

4. The student will be able to differentiate among various types of finishes and to recognize characteristics of fabrics with the following finishes: permanent press, soil release, sizing, and shrink resistant.

Care of Clothing

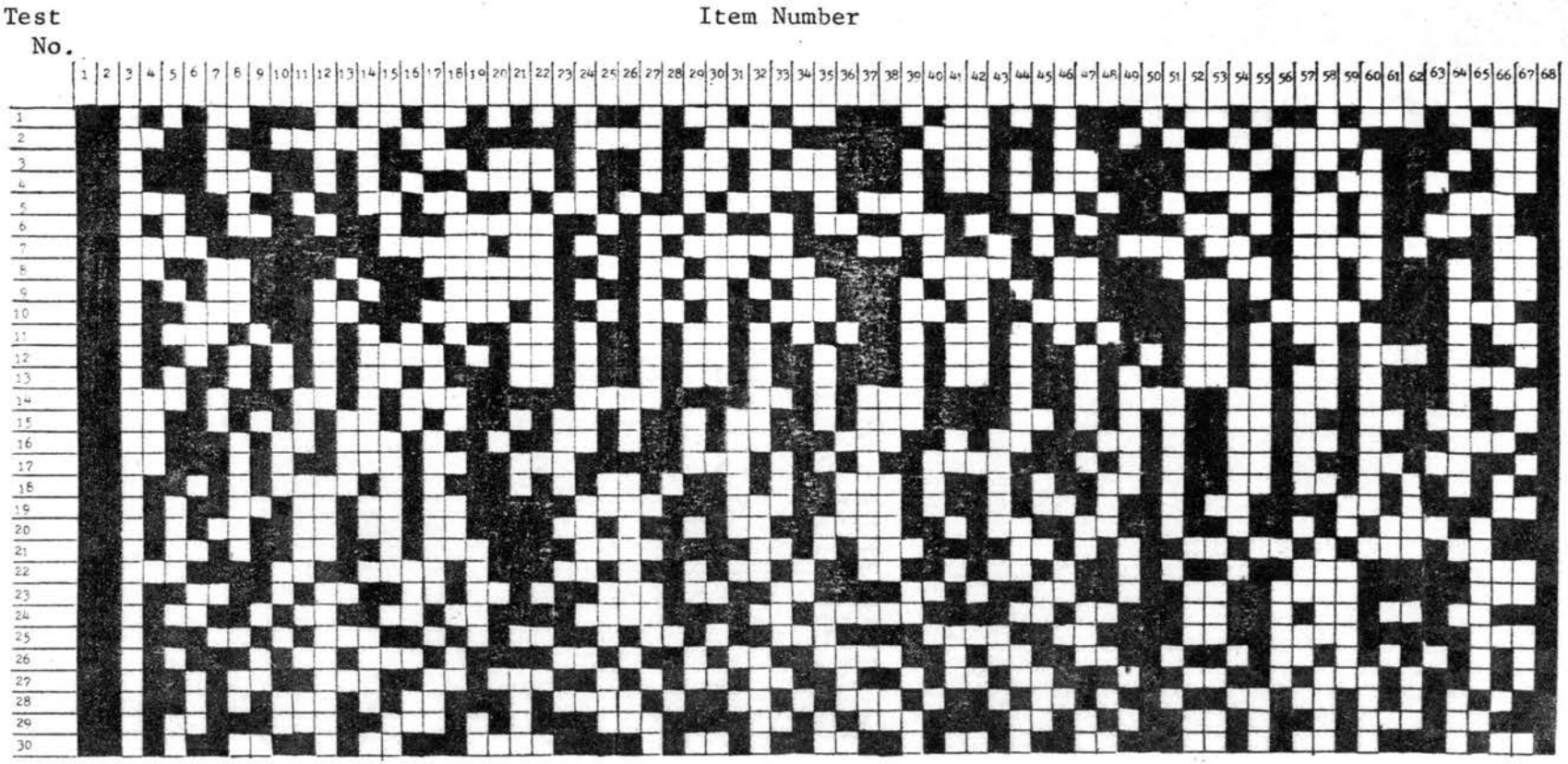
Objective:

1. The student will be able to recognize procedures for proper care of garments.

APPENDIX C

PATTERN OF ITEM SELECTION

PATTERN OF ITEM SELECTION



■ = items appearing on each test

APPENDIX D

SAMPLE QUESTIONNAIRE

SAMPLE QUESTIONNAIRE

You have just taken a test written by a computer. By using computer written tests a student will be able to take a test any time he is ready for it and to repeat a test whenever necessary. The test can be graded immediately so the student will know the correct answers.

1. Which would you prefer? A teacher written test____or a computer written test____

2. Where would you prefer to mark your answers? On the test____or on a separate answer sheet____

3. What did you like the <u>most</u> about this test?

4. What did you like the <u>least</u> about this test?

VITA

Winellen E. Wilkins

Candidate for the Degree of

Master of Science

Thesis: AN INVESTIGATION OF THE FEASIBILITY OF USING COMPUTER GENERATED TESTS IN A CLOTHING SELECTION COURSE AT OKLAHOMA STATE UNIVERSITY

Major Field: Clothing, Textiles, and Merchandising

Biographical:

Personal Data: Born in Oklahoma City, Oklahoma, December 12, 1947, the daughter of Mr. and Mrs. Winston Ward Halbert.

Education: Graduated from Dale High School, Dale, Oklahoma, in 1966; attended Eastern Oklahoma State College in 1967 and 1968; received the Bachelor of Science degree from Oklahoma State University in 1970, with a major in Vocational Home Economics Education; completed requirements for the Master of Science degree at Oklahoma State University in July, 1971.