DEVELOPMENT OF A PAPER AND PENCIL TEST AND A PERFORMANCE TEST TO ACCOMPANY A SELF-INSTRUCTIONAL PROGRAM ON THE SEWING MACHINE

by

Carolyn Elizabeth Ross

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
June, 1965

Approved by

Director

APPROVAL SHEET

This thesis has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro, Greensboro, North Carolina.

Louise Lowe
Thesis Director

Oral Examination Committee Members

Hildegarde Johnson

Irwin D. Sperry

Gail M. Dennis

Helen K. Staley

March 25, 1965
Date of Examination

ROSS, CAROLYN ELIZABETH. Development of a Performance Test and a Paper and Pencil Test to Accompany a Self-Instructional Program on the Sewing Machine. (1965) Directed by: Miss Louise Lowe. pp. 82.

In this study a performance test and a paper and pencil test were developed to accompany a self-instructional program on the sewing machine. The items were analyzed to determine the test reliability and the discriminating power of each test item.

The performance test consisted of student directions, a check list for the teacher, and was accompanied by a list of procedures for administering the test. The ability of the student to fill the bobbin, thread the upper and lower parts of the machine, adjust the tension and stitch-length regulator, sew a 5/8 inch seam, and replace the zipper foot was tested.

The performance test was administered to thirty-five students in the high school in which the researcher was teaching. These students were scored independently by three judges. Correlations were computed between the scores of the judges. Scores of Judge One correlated higher with the scores of the other two judges than their scores correlated with each other's scores. Therefore, they were used to determine the upper and lower 27 per cent of the group in order that the discriminating power of the items could be computed by Flanagan's method.

The paper and pencil test, "A Quiz From Sewing Sue," consisted of three parts. Parts A and B were matching items concerning the description and function of various machine parts. The multiple choice items in Part C measured the student's ability to apply this information.

This test, with a possible score of 63, was administered to 241 students, and their scores ranged from 24 to 60. The mean was 45.7, and the median was 46.

Items were analyzed on test papers of the sixty-five students who scored in the upper and lower 27 per cent of the entire group. Reference to Flanagan's Table indicated the discriminating power of the test items. Items with a value of .20 or above were considered satisfactory.

Sixteen of the eighteen matching items in Part A and Part B of "A Quiz From Sewing Sue" showed sufficient discrimination between the upper and lower groups of students. Forty of the forty-five multiple choice items in Part C showed adequate discrimination. All of the responses in the multiple choice items were rated as functioning responses since each was selected by more than 3 per cent of the total number of persons taking the test.

The coefficient of reliability was determined by dividing the test into two subtests which used the odd items for one subtest and the even items for the other subtest. The scores of the two subtests were correlated and a coefficient of reliability of .699 was determined for a test one-half the length of the test. The Spearman-Brown modified formula was used to estimate the reliability of the entire test which was shown to be .822.

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to all those who made this study possible through their cooperation, participation, and encouragement: Miss Louise Lowe for her time and effort and constant encouragement which enabled the completion of this thesis; Dr. Hildegarde Johnson for help and guidance in planning and securing data for this study; her family for their encouragement during graduate study; Mrs. Helen K. Staley and Dr. Gail Hennis for their helpful suggestions and encouragement as committee members; and the cooperating home economics teachers, the student teachers, and the high school student for their active participation.

LIST OF TABLES

CHAPTER I

INTRODUCTION

Evaluation is considered one of the most important aspects in the field of education. The nature of evaluation is determined by the aims of the educational experience, but whatever the experience, the purpose of evaluation is to improve learning. The evaluation program, therefore, must be broad, diversified, and continuous. The process begins with the planning of the experience, continues during the experience to identify signs of progress or of difficulties, and terminates at the completion of the experience when evidence of progress has been provided.

Techniques of evaluation are numerous and varied, and one of the greatest problems facing teachers is that of selecting the best way of appraising pupil progress toward given goals. While some experiences can best be evaluated by paper and pencil tests, the degree of achievement reached through other experiences may better be measured by performance tests. For more complex experiences or in special situations, appraisal through both written and performance tests may be desirable. The belief that this dual type of testing would be most effective in appraising the progress of pupils in a self-instructional program has led to this study.

Background of the Study

Research in the area of home economics education at the University of North Carolina at Greensboro during the past two years

has evolved around the concept of programmed learning. A self-instructional program on the sewing machine was begun by Moore (15) and revised and field tested by Shoffner (18). This program was written for students who had had no instruction in the use of the sewing machine. As a result of studying the program, the student would be expected to verbalize, identify, and locate designated parts of the sewing machine and to perform certain specified operations on the sewing machine.

Moore (15) recommended that the criterion test used in her study be revised, and a performance test be developed. She further suggested that the revised criterion test be administered to a sufficient number of students to effectively determine the item validity as well as the reliability of the test. A criterion test was used in her study, but revision of that test was needed to make it more effective in measuring the student's knowledge after completing the program. No performance test was developed by Moore to measure the student's ability to perform designated operations on the sewing machine, although the need for one was recognized. Both a criterion test and a performance test were needed in the winter of 1963 in order to evaluate the effectiveness of Moore's program after revision when it was field tested by Shoffner (18).

Purpose of the Study

A larger project is planned at the University to include a series of programs in the area of home economics. Three tests are to be designed to accompany each of the programs--a pretest, a test composed of items almost identical to the frames in the program, and a test composed of application items of the basic principles included in the program.

As self-instructional programs are new resources in the field of learning, it would seem desirable to find effective means of evaluating student progress in these programs.

This study is concerned with two evaluation devices to accompany the self instructional program on the sewing machine. The three-fold purpose is to (1) design a paper and pencil test which will measure the student's ability to identify, locate, and verbalize the functions of the designated parts of the sewing machine, (2) design a performance test which will measure the student's ability to perform designated operations on the sewing machine, and (3) determine the reliability of the tests and the discriminating power of each test item.

Definitions of Terms Used

There is a need for definition of terms commonly used in the area of tests and measurements and in self-instructional programs. For clarity in understanding the development and analysis of the tests the following definitions of terms will be used in this study.

Program: A learning device composed of small segments of information, arranged in sequential order, to which a student responds at his own rate of speed.

Frame: A single segment of material to which a student responds at one time.

Performance test: A test designed to measure the student's performance of some task, or the final product resulting from the performance of the task.

Paper and pencil test: A test planned to measure knowledge and the

student's ability to verbalize information learned through the self-instructional program.

Test item: A numbered statement in one of the tests.

Nonfunctioning distractor: A response selected by less than 3 per cent of the upper and lower 27 per cent of the population.

Organization of the Thesis

The remaining chapters of this thesis will include a review of literature related to performance tests, the methods followed in developing both the performance and paper and pencil tests, and the findings obtained from the tests. A summary with suggestions of ways for further improving the two tests will complete this report.

CHAPTER II

REVIEW OF LITERATURE

The review of literature for this study was limited to references pertaining to performance tests. In the first section investigation of the development, uses, and scoring of performance tests was made; and the writings of specialists in the field of psychometrics were reviewed. Consideration was given to two research studies in which the development of a performance test was a major purpose in the second section. The tests, in these studies, were similar to the performance test developed in this study in that they required little equipment and were relatively easy to administer.

Ryans and Frederiksen (16, pp. 457-8) suggested that performance of an activity under supervision is one of the earliest forms of testing. The Greek games and races before the time of Christ were performance tests in which individual differences in skill were measured. Later, performance tests were used to measure achievement in music, art, and literature. In recent years performance tests have been used in industry to help an employer judge the level of skill that might be expected of an applicant for employment.

Kinds of Performance Tests

Adkins has defined a performance test as a "test in which the subject is directed to carry out some activity." (1, p. 211). A

performance test is described by Ryans and Frederiksen (16, p. 455) as a test that measures a kind of achievement beyond knowledge of facts and principles. According to Hall and Paolucci (8, p. 310), a performance test is a means of evaluating observed behavior, a process which includes the appraisal of the step by step behavior of the person performing a task as well as the appraisal of a final result or product.

Performance Tests of Aptitude

Adkins said aptitude type performance tests are "designed to predict the subject's potential skill." (1, p. 211). Ryans and Frederiksen (16, p. 457) reported that performance tests of the aptitude type are used as a substitute for intelligence tests with students who have language deficiencies, such as deaf students, foreigners in a culture different from their native culture, and uneducated individuals. The Knox Cube Test, the Seguin Form Board Test, and the Portesu Maze Test are examples of performance tests which measure aptitude. Some performance tests measure such special skills as spatial perception, manual dexterity, and neuromuscular coordination. Since these special abilities are used in the performance of many tasks, the tests measure aptitude for certain kinds of work.

Since the performance test developed in this study was intended to measure student achievement in using the sewing machine, the remainder of this review of literature has been limited to a review of performance tests concerned with achievement.

Performance Tests of Achievement

Ryans and Frederiksen (16, p. 458-66) stated that performance tests which measure achievement are concerned primarily with the performance of a particular task. These tests are of three major types: recognition tests, tests involving simulated conditions, and work sample tests. They emphasized that a recognition test "attempts to measure the individual's ability to recognize essential characteristics of a performance or product of performance, or to identify objects such as geological or botanical specimens." (16, p. 458). These tests distinguish between a right and a wrong situation whether it be the use of the equipment, the method used, or the finished product. Examples cited of the uses of this type of performance test are: (1) finding errors in musical selections, (2) selecting correct samples of wire splicing, (3) finding defects in equipment, and (4) identifying different leaves. Ryans and Frederiksen said that this type of performance test is more easily developed than is the work sample test. However, an individual's procedure, technique, or mastery of a skill is not measured by the recognition type of test.

The second type of performance test used to measure achievement is a test involving simulated conditions. Such a test must be a copy of a real life situation. Miniature tests often used in industry are those in which apparatus has been constructed especially for the testing purpose. Examples of these are the miniature punch press and the Wisconsin Miniature Test of Engine-Lathe Operations. These tests have the advantage of being convenient for multiple sets. Their disadvantage lies in the fact that often conditions other than the physical set-up

influence the performance. The emotional factor, for example, has an influence on the output. According to Ryans and Frederiksen:

> Simulated conditions and miniature tasks must be used with caution and with as complete knowledge as possible of the relationship between the results of the miniature tests and the more complete performance with actual equipment and under actual conditions. (16, p. 462).

A work sample test may be considered a "controlled tryout" under the conditions in which the work is usually carried out. This type of performance testing is usually more valid than the other types, and reliable estimates of performance can be secured. There are two principle kinds of work sample tests according to Ryans and Frederiksen:

> ...those in which clear cut distinction between 'rightness' or 'wrongness' of the execution of a skill is possible and which, therefore, are more or less automatic in scoring, and those which must depend upon the judgment of observers for evaluation and assignment of a score or rank. (16, p. 462).

Examples of work samples are themes written in English classes, situations in which music students play selections, and situations in which the student transcribes shorthand on the typewriter.

Performance tests of achievement are used for the following purposes: (1) to predict future success of skills, (2) to find deficiencies in performance, (3) to serve as teaching aids, and (4) to form standards of performance. The prediction of job efficiency in industry is more effective when performance tests are used in the classification and placement of employees. Diagnosis of deficiencies in performance provides a means of analyzing behavior in terms that can be easily observed. A performance test can be used as a pretest in the beginning of a classroom unit, and it may also be used to see if the objectives of the unit were achieved. As a criterion device, a performance

test has been used for research purposes, evaluation of a unit, and for determining promotions.

Performance tests have been described by Hardaway (9, pp. 39-44) in the field of business education as tests in which a student has to apply knowledge and skill to perform certain tasks. These type tests are of importance in business education because the end products, for example, a correctly typed letter, are of importance in the field of vocational business education.

Performance tests are used widely in business courses. Tests for typing classes consist of timed copy tests as well as tests on centering typed lines, letter writing, and tabulations. Problems in keeping records are the type of performance tests used in bookkeeping. Tests used in business arithmetic, machine calculation, and clerical office practice are of a performance nature. Shorthand performance tests require the student to write shorthand for word lists or dictated paragraphs or letters. In areas such as shorthand, it is believed that performance tests should be used often in the last part of the course because the basic knowledge should have been learned earlier. A special type of performance test in the field of business education is a production test in which several skills must be applied for work to be done in an office situation. Hardaway stated:

> Authorities have recommended that terminal examinations in the skill subjects shall be tests of productive ability in which speed of production, as well as excellence of performance, shall be the criterion for recommending students for employment. The better employment tests long have been of the performance type (9, p. 43).

Two individual tests of musical performance were described by Kwalwasser (11, pp. 102-6). These tests, the Hillbrand Sight-Singing

Test and the Mosher Test of Individual Singing, were planned to measure sight-singing skill. The Hillbrand test contained six songs which the student sang without help or accompaniement after studying them for a few minutes. Errors of nine different kinds are recorded on a copy of the songs. Mosher's test consisted of twelve exercises; however, only two points, accuracy of time and of tone, needed to be considered in each measure. A high percentage of agreement was found among the fifteen judges used in the latter test.

Development of Performance Tests

Ryans and Frederiksen (16, pp. 483-93) discussed specific steps in the development of performance tests. The steps were: (1) to make a job analysis, (2) to select tasks representative of the job, (3) to develop a form for rating the performance, (4) to survey the limitations of the test, (5) to develop a tentative plan, (6) to tryout and revise the test, and (7) to prepare directions for administration and use of the test. Each of these steps will be discussed subsequently.

A job analysis should first be made to determine the skills and abilities needed to perform a particular job. Adkins (1, pp. 217-24) also recommended a job analysis in which a series of categories were to be used in analyzing the details of the jobs to be performed. The first category included qualities that are directly measurable in the product. Such factors as accuracy, strength, and instrument measurement are included here. The second category concerned qualities that are subject to judgment such as appearance and feel. The proper work methods were recorded in the third category and speed in the fourth category.

The second step recommended by Ryans and Frederiksen (16, pp. 484-92) was that of selecting tasks representative of the job. They reported that it was desirable to design the test to cover as many of the skills and abilities used in the particular operation as possible. One must determine if the test should measure the product or the behavior involved in completing the task. Routine operations should be omitted in favor of the crucial operations involved in a particular task. If the product is measured, samples must be used to yield measures of proficiency.

As a third step, a suitable rating form should be designed to record the behavior observed in the performance. Time may or may not be recorded. Descriptions may be weighted according to their importance. The form may consist of the list of correct operations, or it may list characteristics of good and poor products.

The practical limitations of the performance test must be surveyed as a fourth step. The amount of time which can be used for performance testing must be considered. In some cases, as in an advanced typing class where students are developing skills for future employment, most of the time for instruction can be spent in testing while in other cases this would not be feasible. Hardaway (9, pp. 39-44) used shorthand as an example of a course in which performance tests are needed, particularly near the end of the period of instruction. The equipment needed must be considered as it influences the practicality of a performance test. Another limitation to be considered is the personnel available. Qualified people must be available to evaluate the results of the performance.

The fifth step recommended for use when developing a performance test is to organize all the materials collected and set up a tentative plan. Numerous trial runs may be necessary to estimate time limits and to determine if directions are clear. A complete description of the procedures must be provided. The plan must also contain directions for administering the test, training the judges, checking the equipment, and conducting the test.

The tryout of the test can point out weak spots in the plans which should be revised before the test is given to a large group. This revision is the sixth step in the development of the test. Observations must be carefully recorded. The subjects used in the trial test should represent the population to be tested. Judges should score the tests independently. Comments of both the judges and students must be carefully considered for improving the test. Results of the test tryout can be used to suggest the internal consistency of the test, the reliability, and the validity of the test. The revision process may mean further training of the judges, revisions of the rating chart, or revision of the directions.

The final step listed by Ryans and Frederiksen in their directions for developing a performance test was to prepare for administration and use of the test. Detailed descriptions of the procedures are to be written. Specific instructions must be supplied prior to administering the test, including equipment needed, room arrangement, and preparation of the judges. Directions for administering the test also include instructions for giving directions, timing the test, and maintaining standard conditions for testees. Directions must be given for grading,

recording, and converting the raw scores to standard scores.

Similar steps in planning and formulating performance tests were considered by Meyers and Blesh (14, pp. 86-98). The steps were: (1) to consider objectives and analyze ability to be measured, (2) to select test items in relation to performance objectives, (3) to set up directions to standardize the testing procedure as items are devised, (4) to select a typical class group, (5) to select a means for determining validity of the test, (6) to determine the validity of each test item, (7) to determine the reliability of each item, (8) to determine the objectivity of each test item, (9) to determine which items correlate poorly with each other in order that they may be eliminated, (10) to select the best combination of tests as a battery, (11) to decide the weighting to be assigned each test in the battery, (12) to arrive at standard scores for the test battery, (13) to devise a manual to accompany the test battery, and (14) to prepare norms for the battery. These will be discussed briefly.

Meyers and Blesh, as a first step, stressed the importance of using test items that were developed in accordance with the desired outcome of the teaching. By making a job analysis, it is possible to know the specific skills required and also to gain an indication of their importance one to another. It is then possible to determine skills at different levels of ability.

In connection with the job analysis it is possible, secondly, to select test items that are representative of the performance desired. Preparation of several test items for each skill makes it possible to select the best item. The time and equipment required for performance

testing should be considered in selecting the test items as well as item difficulty.

The authors considered as a third step the preparation of suitable directions. Specific directions, listed as each item is developed, help to eliminate results of doubtful value or an excessive number of variables. These directions also serve to standardize conditions for all groups taking the test.

As a fourth step, it is important to select a trial group of students who are representative of all the students for whom the test is being planned. Factors affecting the size of the tryout group are the range in variablity of the ability displayed and the degree of confidence desired in the test results. The groups by necessity must be smaller if they are rated rather than given a validated test.

Meyers and Blesh's fifth step in formulating performance tests is to select a criterion measure to be used in determining the validity of the test being developed. Criterion measures that meet with the author's approval are a previously validated test, tournament rankings or records, and rating of expert judges or a jury of expert opinion. The most expedient means of obtaining a correlation coefficient is to use a previously validated test or tournament rankings.

The sixth step is to ascertain the validity of each test item. A validity coefficient of .80 or above indicates a correlation that is very good. If the validity coefficient is from .60 to .79, the test item could be combined with other tests. A validity coefficient of .59 and below is a correlation that is illustrative of a generally unuseful test.

The seventh and eighth steps are to determine the reliability of each test item. Reliability can be determined by retesting under similar conditions or by correlating two halves of the test. The Spearmen-Brown prophecy formula is used in the latter case. Objectivity is determined by correlating the test scores reported by various scorers.

As a ninth step, Meyers and Blesh list that of computing the intercorrelation between desirable items. This step helps to eliminate test duplication and unnecessary test items. To combine test items into a battery, "each test in a battery should possess desirable validity itself but show little, if any, relationship to its companion test in the battery." (14, p. 95)

The tenth, eleventh, and twelfth steps, according to Meyers and Blesh, deal with the test after it has been developed. A battery is selected by the best combination of tests determined by the correlations in previous steps. Ease in administration of various tests should be considered, and a method to be used in combining the scores should be established. Meyers and Blesh suggest a regression equation formula. After establishing the battery raw score, a standard score should be developed to interpret the raw score. Finally, norms can be set up if the test is given to enough students. These students should be a representative sampling of the students for whom the test is designed.

Hall and Paolucci (8, p. 314) offered other suggestions for developing performance tests. The situation should be selected on the basis of the basic skills involved in the situation and on the basis of

the cost of the materials needed. A check sheet should be used so that the observer would know specifically which skills to look for, and he could record the results as the skills were observed. Directions should be specific and easily understood by the student. A test should be only long enough to be completed in one class period.

Characteristics and Limitations of Performance Tests

Arny (3, pp. 73-4) presented five desirable characteristics of performance tests. The first characteristic concerned the number of skills to be checked in the test; that is, the test should cover as many of the fundamental skills as possible. In the second place, performance tests should use materials that are relatively inexpensive. This is particularly true in cases where the materials cannot be used a second time, for example, food could not be used a second time in a test of skill in meal preparation. The third characteristic was that the test should be short enough to be completed in the regular class period. In the fourth place, the performance test should be designed so that it can be scored objectively. A rating device should be developed for objective scoring. Arny's fifth desirable characteristic of performance tests was that the test situation should represent a real life situation. She suggested that in the case of a performance test of skill in meal preparation, the recipes should be given to students the day before the meal is to be prepared. Students could avoid strain by planning in advance, and failure by incorrect memory of ingredients used could be eliminated.

Adkins (1, p. 216) compared a performance test in which students

looked at actual tools and a written test in which students were given a picture of the same tools. She reported advantages and disadvantages of the performance test and the written test. The written test, showing the same tools, was available to more students at the same time. The test conditions of the paper and pencil test could be controlled because such factors as directions, seating, temperature, and administration would be the same. A written test would have the additional advantage of being quicker to score, and there would be an opportunity to measure the student's knowledge of the functions of the tools.

The value of performance tests over paper and pencil tests must be considered. Adkins wrote that performance tests "are very useful in examining for certain types of occupations if the same degree of effort and care is applied as in the development of other measuring devices." (1, p. 216). Hall and Paolucci (8, p. 310) stated the advantages of performance tests were: (1) time can be recorded, (2) work habits can be noted throughout the process, (3) selection of correct equipment can be observed, (4) student's self confidence can be noted, (5) group cooperation can be observed.

Scoring of Performance Tests

Performance tests according to Arny (3, p. 199) can be scored by inspecting the completed product or by checking the performance at successive stages. Ryans and Frederiksen (16, p. 475) reported that measurement of the final product is done by comparison with a standard of quality. This quality can be measured by error scores for typing, patterns and guages in wood and metal work, and graded sample quality

scales in handwriting, and in electrical work.

Arny (3, p. 199) suggested that inspection of the performance at successive stages limits the number of students who can be tested at one time. Ryans and Frederiksen (16, p. 470-4) wrote that quality and speed are the two main factors measured in this method. Sometimes one of these factors may be more important than the other, while in another situation the importance of the factors may reverse. Factors to be considered in judging performance in addition to quality and speed are accident rate, attitude toward work, and learning time. A rough point-scale or a recording machine can be used in some cases in measuring performance processes.

Rating Scales

A rating scale is defined by Wrightstone (22, pp. 163-7) as "a selected list of words, phrases, sentences, or paragraphs following which an observer records a value or rating based upon some objective scale of values." (22, p. 156). The five types of rating scales most often used are classified as the descriptive, the graphic, the forced-choice, the rank method, and the paired-comparison. A descriptive rating scale is one in which the observer checked one of the following phrases in connection with some behavior: all the time, most of the time, occasionally, or never. A graphic rating scale is one in which the observer checks the presence or absence of a given trait. The forced-choice rating scale is an attempt to obtain a truer rating of the individual. The rater chooses between paired alternatives, both of which are characteristics but one characteristic is considered more

desirable. The rank order method is one in which the persons being observed are rated in order from high to low levels on the possession of some quality under consideration. The paired-comparison method is the type in which each individual is judged as better or worse than every other individual in the group. Because of the time consuming statistical analysis, this type of rating method is not effective for teacher use.

Meyers and Blesh (14, pp. 98-101) stated that the device should first describe clearly and distinctly what is to be rated in order that there be a clear conception of the quality or skill being rated. In the second place, the number of catagories in the rating should be established, thus defining the range of rating and the amount of discrimination. In general, the larger the number of catagories, the finer the discrimination would be. These authors further stated that each category or scale point should be described exactly. Sometimes pictures or charts are necessary for a complete description of the category. Next, a specific rating procedure should be prepared and plans made for effective rating. Wrightstone (22, p. 169) stated that complete directions with details about each quality or skill to be measured increases the effectiveness of the rating scale. It was further stated that seven units or categories yield optimal reliability.

Check Sheets

Check sheets and score cards are two of the most widely used scoring devices according to Arny. (3, pp. 204-213) Examples of these are the "Food Score Cards" by Brown and others, the "Minnesota Check List for Food Preparation and Serving" by Arny and others. "The Food Score

Cards" evaluate food according to such factors as appearance, color, taste, and flavor. Rating is possible at three levels. The "Minnesota Check List for Food Preparation" lists eleven topics and three descriptions for each topic. The rating on this device is from one to five. Hatcher developed the device, "Check List for Table Setting." Rules, illustrations, and a check sheet are included with this device.

Other Scoring Devices

Other types of scoring devices include product scales and diagnostic charts. Product scales are designed to measure a more or less complex ability as a whole such as a series of photographs which illustrate different levels of quality of product. Diagnostic charts help students analyze their own differences, but they have no value as scoring devices.

Evaluation of Performance Tests

Reliability

Ahmann (2, p. 340-1) believed two factors largely determine the reliability of a performance test, the consistency of both the observer and the performance which is to be evaluated. Sampling of performance is limited, but ideally the student should be evaluated on several occasions under normal conditions. In reality, there may be only one occasion to evaluate the student's performance, and that may not be under normal conditions. If the product of the performance is completed over a period of time, help from others may be received by the student. This may mean that the student's performance has been altered by an

outside source.

A consistent observer is important for the reliability of a performance test. These tests should be scheduled over a period of time to avoid boredom, fatigue, and indifference on the part of the observer. Since it is not entirely possible to eliminate all of such factors, several observers might be used and their scoring compared. The "halo effect" created by partiality or favoritism of a judge toward a student would increase the degree of consistency of the scoring of the judge, but it would reduce the accuracy of the evaluation. A variety of sample products and careful descriptions of the degree of excellence of performance to be expected would increase the reliability of the observer.

Ryans and Frederiksen (16, p. 483) suggested ways to improve the observer's methods. The more capable observers should be used, and even these should be given a short training session in some cases. It was further suggested that an objective rating scale be used to increase the objectivity and consistency of the observers.

Reliability of performance tests is influenced by the length of the test and by the characteristics of the items which were selected for the test. Ryans and Frederiksen stated that operations involving a long series of tasks were usually of higher reliability than ones involving short operations.

Adkins said:

> Reliability also requires that the test measure a sufficiently constant characteristic of each subject and that it not often permit high or low scores to be obtained by accident, so that subjects do not shift markedly in relative proficiency each time they may take the test. (1, p. 315)

Ryans and Frederiksen (16, p. 483) stated that reliability of the

performance can be measured only when the reliability of judging the performance is satisfactory. Correlations of scores on one set of tasks with scores on a second set of tasks is low when a series of tasks are involved because the student may do some tasks well and others not so well. A test-retest type of reliability is not satisfactory in other cases because the administration of the first test would have a practice effect on the second test. Another factor to consider in determining reliability is the variation in the condition of the equipment used in the test. For example, the same performance using a different oven might cause a variation in the quality of cake baked in each oven.

Validity

Validity, the measurement of what an instrument purports to measure, is important for paper and pencil tests as well as performance tests. Bradfield (4, p. 186) said there were two approaches to validity, the logical or rational and the empirical or statistical. In the rational approach the test is analyzed in terms of the general format, the objectives of the test, and the character of the items. Cureton (7, p. 663) referred to logical relevance in this rational approach to validity. He said logical relevance existed only when all parts of the test could be observed and scored without systematic biases.

As validity is affected by the tasks selected for inclusion in the test, Ryans and Frederiksen (16, pp. 467-8) suggested that a job analysis helps to determine which tasks are important enough to be in the test. The tasks that are selected should, as much as possible,

represent the general abilities needed for the performance. Factors of time, personnel, and equipment affect the selection of the tasks. Adkins said that "a wide variety of work samples must be tried out and statistical analysis used to determine which are the discriminating elements and to ascertain that no important phase of the job has been 'evaporated out' in the process of boiling down the test items." (1, p. 213).

The second approach to validity is the statistical or empirical approach as reported by Bradfield (4, p. 186). First, a criterion must be established upon which the test may be validated. The criterion should be reliable and free of bias. It is not possible to establish such a criterion in some cases of achievement testing, but in others the criterion can be developed in terms of later performance as in tests used for prediction or aptitude tests. Sometimes the test can be validated by referring it to a similar test for which validity has been assumed.

After establishing the criterion, the test should be administered to a typical group. Each person should be rated by a group of experts in actual situations that demand various levels of the particular comprehension being considered. A correlation coefficient is then computed between the test scores and the ratings of the experts. This coefficient of correlation is called the validity coefficient. This method is limited because of the difficulty in developing a suitable criterion.

Research Studies Using Performance Tests

Two studies using performance tests have been selected for this

review of literature because they were similar to the performance test developed in this study. These performance tests require apparatus that would normally be available in the particular situation. The first study was in the area of wood shop performance, and the second was in the area of clarinet playing.

McPherson (13) developed a method of objectively measuring shop performance in which the students copied a model wood block using a saw, drill, and chisel. The model was presented in four stages for the students to see but not to touch. A pattern of the block was drawn on transparent plastic to be superimposed over the student's models. Lines with definite numerical values were drawn on the pattern to show deviations of the student's work from the models.

This method was presented to fifty-nine boys enrolled in a woodshop course. Each student was given a board and asked to make his board look like the model. There was no time limit and only casual adult supervision. Only one student performed the task at a time.

One semester later, fourteen subjects were given the same problem for the purpose of measuring any change in skill that one semester of shop work had made. Each raw score was multiplied by a number from one to six, the number depending on the difficulty of the tasks and the number of times the operation was scored. A point scale was used to evaluate neatness and method of determination of position of operation.

The gains in the retest scores and the distribution of scores indicated that the method was of value. The difference between the mean scores of the two testings was fifteen since 55.92 was the mean of the first scoring and 70.92 the mean of the second scoring. Reliability of

the scoring method was indicated when the boards were rescored eight months later with little variation.

A similar study for metal-shop measurement was conducted by the same staff using similar techniques to those that were used in the wood-shop study. The scoring pattern was of the same type as that used in the first study.

Woelflin (21) was interested in conserving the classroom time of the instrumental music teacher. He developed a teaching machine program to teach beginning clarinet students clarinet fingerings and such factual knowledge as the names of the parts of the clarinet. The students in his study were divided into three groups. The first group, composed of seven members, was the control group. The second group, an experimental group, had six members who used the teaching machine and held a clarinet which could not be played. The third group, an experimental group of five members, used the teaching machine and played the clarinet.

Two tests were administered during the course of the experiment. The first, a written pretest, was given to all groups before the experiment was begun. This test covered factual knowledge about the keys and registers of the clarinet as well as the fingerings. The student's score was the number of errors made in the test. This test was also administered at the end of the experiment. The total score for the student at the completion of this experiment was determined by subtracting the score of the pretest from the final score.

The second test was a performance test in which students were required to play scales and to play a sight reading exercise. The purpose of this test was to determine if the student could use the

information and skills which he had learned about clarinet playing. All fingering errors, incorrect notes, and incorrect articulations were to be marked on the musical passages on the rating sheet where they occurred. Tone quality, intonation, embouchre, and tonguing were rated on a ten point scale extending from poor to excellent.

A committee of three evaluated the students' performance. Each member of the committee was given a rating sheet for each student. Tone quality, intonation, embouchre, and tonguing were rated by each judge. These scores were averaged for the final score. For the rest of the test one committee member graded incorrect notes, another incorrect articulations, and the third graded incorrect fingerings. The final score for the student was composed of the sum of all the scores for the different parts of the test. The score of each pupil was a summation of errors made; therefore, a low score was a good score while a high score was a poor score.

CHAPTER III

METHOD OF PROCEDURE

The three-fold purpose of this study was to develop a performance test and a paper and pencil test to accompany a self-instructional program on the sewing machine, and to determine the reliability and the discriminating power of the items in the tests. The performance test was designed to measure the student's ability to perform designated operations on the sewing machine. The objective of the paper and pencil test was to measure the student's ability to identify, locate, and verbalize the functions of designated parts of the sewing machine.

These tests were planned so that a classroom teacher could administer them to a class in one class period. The typical class could be divided into two groups; half of the students taking the performance test at the sewing machine while the remainder of the class responded to the items on the paper and pencil test at their tables. Students would exchange places with someone in the other group when they had completed either of the tests.

The first step in the development of the two tests was to study the specific objectives of the sewing machine program. The first major objective stated that the learner should be able to identify, locate and verbalize the function of certain machine parts. The other major objective stated that the learner should be able to perform certain operations at the machine.

Development of the Performance Test

A review of several theses in physical education (6, 10, 20) gave suggestions for ways of directing students to perform a desired task, and ways of observing and scoring the performance of some motor skill. The rating scales used in these studies consisted of five point scales ranging from poor to excellent.

Chadderdon (5) designed a device to be administered to girls with little experience in sewing to determine the kind of help they would need in using the sewing machine. A number of difficulties or errors that might be made while threading the machine and winding the bobbin were listed on a rating chart. On this chart a place was available for the name of each student and a place to check a number corresponding to the difficulties experienced by each student. The test was to be scored by a judge who observed the performance. The lower the score, the better the student's performance as the score was the sum of the errors made.

The form of the performance test to be used in this study was determined after considering a number of tests which included those described above. The test itself would include the instructions and the specific steps for the student to follow. A set of directions for administering the test and a check sheet for scoring would be needed to accompany the test.

The test for the performance of tasks at the sewing machine was developed in four parts, each complete with directions. At the end of each part, the student was instructed to stop, raise his hand, and have his work checked before preceding with the remainder of the test. Part I

involved filling the bobbin. The student was asked to leave the bobbin on the bobbin winder until the precedure was checked. Threading the machine was the procedure tested in Part II. Both the upper and lower parts of the machine were to be threaded. In Part III, the student was asked to sew a 5/8 inch seam with ten stitches per inch and correct tension. After adjusting the tension and stitch-length regulator, the student was instructed to sew the seam while the teacher observed. The presser foot was removed and the zipper foot placed in its place in Part IV of the test.

Two methods of scoring, a rating scale and a check sheet, were considered. The first was a three to five point scale used for appraising quality of each separate step in the performance, similar to the scales used in scoring the performance tests in physical education. The second method of scoring was a chart on which were listed the most common difficulties encountered in the use of a sewing machine. This chart was similar to the one used with the test developed by Chadderdon (5). Errors made by the student were checked on the chart by the teacher who observed the performance.

After working with both methods of scoring, the second method was selected as the basis for developing an evaluation device for this study. This method is objective in that it does not require the observer to make value judgments, instead he responds positively or negatively to a statement. The procedures followed in threading a machine can be objectively scored because each step is either right or wrong. The task of preparing the test for scoring involved detailed listing of points in the process where errors could be made. A list of possible

errors was made for the following tasks: filling the bobbin, threading the upper and lower parts of the machine, adjusting the tension, adjusting the stitch-length regulator, attaching the zipper foot, and sewing a 5/8 inch seam. Specialists in the School of Home Economics at the University of North Carolina at Greensboro were asked to examine the statements for clarity, consistency of form, and logical order of the steps of performance. After the specialists had examined the statements, they suggested minor changes, primarily a rewording of statements.

Errors which a student might make were listed on a check sheet (see Appendix A) which could be used by the teacher. The first four parts of the check sheet corresponded with the four parts of the test. The fifth part was entitled overall procedure, and various work habits were listed in this section.

Instructions to the teacher for administering the test included directions for setting up the machines for the testing program. Materials needed for the test were listed for the teacher. (See Appendix A).

The test was now in a form to be administered to a group of six students chosen from two of the Home Economics I classes at the school in which the researcher was teaching. These students represented varying abilities as determined by the teacher's observation. These students had completed their sewing unit earlier in the year. They had not been taught by the sewing machine program, but each of them had read the sections in the sewing machine program that dealt with the upper tension and the method for testing for correct tension.

Three judges, chosen from the group of home economics teachers in the county, were selected on the basis of their interest in this

test, their closeness to the school, and their willingness to help. These judges had not read the sewing machine program, and they did not read the performance test until immediately before its administration.

The six students used the same sewing machine that they had used during their sewing unit. The machines were placed in a semicircle to facilitate walking from one machine to another. Teachers were given a check sheet for each student, and they were asked to score independently each of the six students. Approximately an hour was needed to administer the tests since the students often had to wait for all the judges to check their work before preceeding to the next step.

The next day each of the six students was asked to state her opinion of the test. Each reported that she had experienced no difficulty in reading and following the directions; no suggestions were made for improving the test instructions.

The cooperating home economics teachers offered several constructive comments. The teachers reported that negative statements were difficult to follow, and that they would prefer positive statements. They suggested that the judge's chart include a statement to be checked when no errors were made by the student. They had some doubt about the practicality of the performance test in a typical home economics classroom.

The investigator's observation of the teachers and the six students revealed some weaknesses in the test and in the administrative procedures. Although the judges knew how to operate each model of the sewing machines, they had a tendency to differ in what they expected the student to do. A joint review of the sewing machine by the investigator

and the participating teachers would have increased the agreement among judges. For instance, one teacher was more concerned about adjusting the tension as a first step in procedure than about threading the machine. The teachers would also have been more alert as to variations in the different models of sewing machines after a short review.

Revisions of the test followed the field trials and were based on judges' suggestions, the students' comments, and the writer's observations of the testing. The teacher's check sheet, the student's directions, and the procedures for the teacher were all revised in light of the comments received.

The revised test was administered to Home Economics I students at the University laboratory school. The judges were two graduate students and a staff member of the School of Home Economics. Each student was scored by only one judge; therefore, it was not possible to measure agreement among the judges in this second testing situation.

This same revised test was given to students in a school in which the self-instructional program on the sewing machine was being used. In this case, students proceeded through the program at their own pace, and took the test as soon as they finished the program. Some students were being tested while others were still working on the program. Graduate students and staff members of the School of Home Economics who administered the self-instructional program scored the students, each student being scored by only one judge.

A second revision of the test was necessary to incorporate the suggestions and comments made by the judges in both situations described above. Items on the scoring sheet were revised to make them more

specific. Additional statements about tension adjustment and the bobbin case were considered for a more thorough evaluation of this part of the test. The section on general work procedures was deleted because no students made errors of the type described in this section.

The second revision of the test was administered to thirty-five Home Economics I students at the school in which the researcher was teaching. These students were in their clothing construction unit, but they had not seen the sewing machine program. In the test situation each student used the sewing machine with which she was familiar.

The three judges for this administration of the test were the researcher and two student teachers working under her guidance. They were deemed qualified by their knowledge of the different models of machines in the home economics department and their familiarity with the sewing machine program itself. Each judge was familiar with the program and all parts of the test. Also, these judges had instructed the students in the use of the machines in the department.

Development of the Paper and Pencil Test

A paper and pencil test was developed to measure the student's ability to identify, locate, and verbalize the functions of designated parts of the sewing machine. This type of test was designed to measure the student's ability to apply knowledge about the functions of the designated parts of the machine.

Matching items were selected for the first two parts of the test. (See Appendix B). These matching items were developed to measure recognition of the part of the machine which is associated with each

description of a machine part and with each function. Column I in Part A consisted of a list of descriptions, and column I in Part B consisted of a list of functions. Column II in both matching sections was composed of a list of sewing machine parts, some of which were used more than once and some of which were not used at all.

Part C of the paper and pencil test (see Appendix B) was made up of multiple choice items stated in language typical of the students. To secure responses in the language of the students, six questions in essay form were given to forty-five students. The questions were:

1. How would you recognize a thread take-up on a machine with which you are not familiar?
2. Why is it important to loosen the stop-motion screw before winding the bobbin?
3. What do thread guides have in common?
4. Where should your hands be when you start to sew?
5. What two machine parts should be observed when you stop stitching?
6. Why is it important to keep your hand on the hand wheel when you stop and start your machine?

Student responses to the questions were categorized, then phrases were selected which expressed correct answers and wrong, but plausible, answers. The latter were student misconceptions of principles and how they apply when a machine is used.

The form of the multiple choice items was considered carefully, with attention directed toward stating both correct and incorrect responses in approximately the same number of words. Most of the items were stated in the form of questions. Two columns of items were used rather than one column to keep the total number of test pages to a minimum.

The test items were submitted to three specialists in clothing at the University of North Carolina at Greensboro. Each was asked to

examine the items for clarity and to react to each so that a scoring key could be determined. If the response was marked by the specialists, it was considered correct and included on the scoring key. Items 8, 9, 37, 41, and 42 were discarded because of lack of agreement among the specialists.

A check sheet was developed for ease in scoring. The responses by the specialists were tallied on an answer sheet which had been glued to a piece of cardboard. Each correct response was then punched. After checking to see that only one response was marked for each item, the stencil could be placed over a student answer sheet and the correct responses would show in the punched openings. Matching items would be checked by a strip key before the multiple choice items were checked.

A data sheet (see Appendix B) was prepared to be sent with the test materials to the teachers administering the paper and pencil test to help them in clarifying their observations. Specific information was requested as to the length of time it took students of various abilities to complete the test. Also, the teachers were asked to note the comments of the students and to evaluate the test themselves. The information from the data sheet was helpful in analyzing the paper and pencil test.

Numerous home economics teachers volunteered to help secure data for this thesis. Four teachers were chosen from the group, and letters (see Appendix B) were sent requesting their cooperation at this time. They were asked to return a post card stating the time they could give the test and the number of students available to take the test.

Test booklets and answer sheets were mailed to these teachers. There were enough answer sheets for each participating student and

enough test booklets for the largest class in each school. The data sheet and specific instructions for administering the test were sent in a letter (see Appendix B) which accompanied the test materials.

Treatment of Data

Performance Test

Three judges scored the performance of each of the thirty-five students. Correlations were computed between the scores of pairs of judges. Judge One correlated higher with each of the other two judges than they correlated with each other. Tests were then arranged according to the number of items missed as determined by the scores of Judge One. The upper and lower 27 per cent of the group were selected, and Flanagan's method (17, pp. 286-92) of estimating ρ in a bivariate normal population was used to determine how the groups differed in their response to the items.

Paper and Pencil Test

Answer sheets were returned from the schools administering the test and were scored using a scoring stencil. Tests were then arranged according to the number of items missed. The upper and lower 27 per cent of the papers were selected for an item analysis. Flanagan's Table (17, pp. 290-1) was used to determine the discriminating power of each item. To determine the functioning power of items in this test, it was decided to drop any response in an item which was not selected by at least 3 per cent of the total number of persons taking the test as suggested by Scott and French (17, p. 282).

The coefficient of reliability was determined by dividing the test into two subtests, using the odd items for one subtest and the even items for the other subtest, and correlating scores on the subtests. The coefficient of correlation was an estimate of the reliability of a test half the length of the present test. The estimate of reliability of the full length test was computed using the Spearman-Brown formula (17, pp. 328-9).

CHAPTER IV

FINDINGS

Two types of tests and two analyses of each are included in the findings. A performance test and a paper and pencil test were developed to accompany a self-instructional program on the sewing machine, and the data collected by administering both tests were analyzed to determine reliability of the tests and the discriminating power of each test item. It was not a purpose of this study to improve the items that did not discriminate adequately according to the standards which were used.

Performance Test

The performance test was divided into four parts for convenience in performing the specified operations as well as for ease in observing the performance. The four parts were: filling the bobbin, threading the upper and lower parts of the machine, adjusting the tension and stitch-length regulator and stitching a 5/8 inch seam, and attaching the zipper foot. The participating students were expected to perform each operation according to directions on an instruction sheet and to summon the teacher when each task was completed. Errors were to be corrected by the teacher before the student proceeded with the test.

This test was administered to thirty-five students who were scored independently by three judges. Students took the test in various open periods during the day and during the regular home economics class

period. Some finished two parts on one day and completed the other parts on the following day. It was found that the three judges could supervise only six to eight students in a regular class period.

The three judges scored each student independently with a few exceptions. The following decisions were made jointly: if the stop motion screw were loose, if the zipper foot screw were tight, if the screw on the stitch-length regulator were tight, and if the cut threads were three inches long. The judges further checked together on the threading of the 66 model machines for winding the bobbin since this model provided a thread guide which the other models did not use. Threads were considered pulled back if they were more back than forward of the needle position. The metal hook near the upper tension regulator on some models was accepted as being the thread guide.

Further study of the performance test suggested changes in the construction of the test and in the check sheet. The judges recommended that Part II, Item 11 and Part III, Item 1, "the thread take-up is at its highest point," should be considered correct if the thread take-up was more than mid-way to the top of the machine. Part II, Item 15, "both threads are pulled under the presser foot and toward the back," was found inadequate.

The judges suggested further that in Part III, a statement should be inserted between Items 2 and 3 about how far the threads should be pulled behind the presser foot. Item 10 in Part III, "the thread take-up and the needle are left at their highest points," needed to be expanded into two items. Threads were found by the judges to be too long in some instances so Part III, Item 12, "the threads are cut so that

three inches are left beyond the needle," should also state that the threads should not be longer than four inches.

The judges discovered that it was difficult to determine in Part III, Item 8, "the hand wheel is used to make the last few stitches," whether the students were actually using the hand wheel to take stitches at the end of the seam or if they were only stopping the machine.

In a number of cases in Part III, the judges found that the student used a diagonal fold in testing the tension, but when asked to make the row of stitching for the judges, they used a single layer on the straight grain. To eliminate this occurrence, the student instructions for Part III should be restated to emphasize repetition of the original performance for the judges. The three judges concluded that use of the same color of thread would make it easier to determine whether the stitching was the same on both sides of the diagonal fold.

Students had no particular questions or comments about the student instructions. They seemed to be able to go about their work independently and with reasonable accuracy.

Correlations between scores assigned by pairs of judges were computed. All correlations among judges were above .90. Scores of Judge One correlated .95 with scores of Judge Two; those of Judge One correlated .92 with scores of Judge Three; and the scores of Judge Two correlated .91 with scores of Judge Three.

The scores of Judge One were used in finding the discriminating power of the test items. There were nine students in the upper and lower 27 per cent of the thirty-five students who had taken the test. Tables 4-7 in Appendix A show the results of this item analysis.

Since this was a mastery test, a different standard was necessary to interpret the results of the performance test than for the paper and pencil test. Such basic steps, as filling the bobbin, threading the upper and lower parts of the machine, and adjusting the stitch-length regulator and tension, must be completed satisfactorily before the total correct performance can be expected. As far as the observation is concerned, the judge must check the steps of performance of each basic step in sequence. Sometimes repetition of a basic step as "The needle is at its highest point," is necessary for checking further performance.

As the study progressed, it became evident that reliability could not be determined for the performance test because the participants would naturally learn by retesting since the errors were corrected before the could proceed to another part of the test. Also, the scores of the thirty-five students who had taken the test were not a sufficient number to give dependable results if reliability were determined by dividing the test into two parts, using the odd items for one part and the even items for the other.

Paper and Pencil Test

The paper and pencil test, "A Quiz From Sewing Sue," consisted of eighteen matching items and forty-five multiple choice items. Part A was concerned with descriptions of machine parts, and Part B was concerned with functions of the machine parts. Part C was composed of multiple choice items related to the use of the machine. While most of the multiple choice items had four responses from which choices were to be made, some items had only two responses.

The possible score on this test was 63. The scores of the 241 students from four schools ranged from 24 to 60. The mean was 45.7, and the median was 46.

The coefficient of reliability was determined by dividing the test into two subtests, that is, by using the odd items for one subtest and the even items for the other subtest. The scores on the two subtests were correlated and a coefficient of reliability of .699 was determined for a test one-half the length of the test. The Spearman-Brown modified formula was used to estimate the reliability of the entire test. The reliability of the test was shown to be .822.

Item discrimination was computed by Flanagan's method of estimating ρ in a bivariate normal population as reported by Scott and French (17, pp. 286-92) and Walker and Lev (19, p. 472). This method determined the ability of each item to differentiate between the upper 27 per cent and the lower 27 per cent of the group. First, it was necessary to compute the percentage of both the upper and lower groups answering each item correctly. These percentages were translated into the values of r by using the table developed by Flanagan. The values ranged from -1 to +1. Positive values indicated that more students in the upper than in the lower group responded correctly. In the same manner, a negative value indicated that more students in the lower than in the upper group responded correctly. If an item had an index of .20 or more, it was considered a discriminating item. Items with an index lower than .20 were considered to be of poor discriminating power.

Eighteen matching items comprised Part A and Part B of "A Quiz

From Sewing Sue." The results of the item analysis and item discrimination are shown in Tables 1 and 2. All of the items except two had an index of .20 or more and were considered to be discriminating. The two items that had an index of 0 were Items 1 and 2. Item 1 was the description,"spindle on which the spool rests," and Item 2 was the "wheel used for starting and stopping the machine."

Table 1. Analysis of responses of upper and of lower 27 per cent of the students: "A Quiz From Sewing Sue," Part A.

	Number responding correctly				
	High Group		Low Group		
Item	Number	Per Cent	Number	Per Cent	r
1	65	100	65	100	.00
2	65	100	65	100	.00
3	65	100	62	95	.30
4	65	100	63	97	.23
5	65	100	48	74	.59
6	64	98	59	91	.30
7	62	95	31	48	.57
8	65	100	57	88	.43
9	11	17	3	5	.30

Table 2. Analysis of responses of upper and of lower 27 per cent of the students: "A Quiz From Sewing Sue," Part B.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	65	100	59	91	.40
2	65	100	54	83	.51
3	65	100	48	74	.59
4	55	85	34	52	.37
5	63	97	47	72	.44
6	59	91	36	55	.45
7	64	98	50	77	.50
8	64	98	46	71	.55
9	64	98	46	71	.55

Part C of "A Quiz From Sewing Sue," consisted originally of fifty items. Item 41 was discarded because of lack of agreement among the specialists at the time the scoring stencil was set up. The item could be revised for future use. Items 8, 9, 37, and 42 were also discarded because of lack of complete agreement among the specialists who responded to the test items.

Forty of the forty-five items in Part C showed items discrimination of .20 or above. These items will not be discussed, but data concerning them is presented in Table 3.

Items 7, 32, 33, 43, and 45 did not show adequate discrimination between the upper and lower groups. The two items concerned with the needle, Items 33 and 43, were similar also in the fact that they were reversals. More students in the low than in the high group reacted correctly to these items. Item 32 tested knowledge of the proper location of the needle when working with the underneath parts of the sewing machine. Location of thread guides was checked in Item 7 while Item 45 concerned the method of tightening the upper tension. Since all possibilities were listed in several of these items, revision of the entire item is necessary. In other cases, the choices may be revised.

The functioning of the discriminators in multiple choice items was determined by using the method described by Scott and French (17, p. 282). When judged by this standard, any response in an item which was not selected by at least 3 per cent of the total number of persons taking the test was considered non-functioning. In this test five students or 3 per cent of the group selected each response indicating that all test items functioned satisfactorily.

In summary, sixteen of the eighteen matching items in Part A and Part B of "A Quiz From Sewing Sue," showed adequate discrimination between the high and low groups. Part C was composed of forty-five multiple choice items with two, three, or four responses. Forty items were found to have a discriminating power of .20 or above between the upper and lower 27 per cent of the students. Also, the responses in the multiple choice items proved to be functioning responses since each was selected by more than 3 per cent of the total number of persons taking the test.

Table 3. Analysis of responses of upper and of lower 27 per cent of the students: "A Quiz From Sewing Sue," Part C.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	61	95	26	40	.62
2	58	89	42	65	.32
3	58	89	31	48	.46
4	54	83	29	45	.41
5	36	55	22	34	.21
6	59	91	42	65	.36
7	36	55	31	48	.06
10	62	95	42	65	.45
11	64	98	56	86	.37
12	61	94	28	43	.61
13	61	94	32	52	.55
14	49	75	33	51	.26
15	54	83	34	53	.34
16	57	88	43	66	.30
17	62	95	49	75	.36
18	41	63	09	14	.51
19	61	94	37	58	.50
20	58	89	44	70	.26
21	63	97	33	52	.60
22	57	89	10	16	.70

Table 3 —Continued.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
23	46	71	25	39	.33
24	65	100	57	88	.43
25	52	80	10	16	.63
26	37	57	15	23	.36
27	62	95	41	63	.47
28	56	86	40	62	.31
29	65	100	58	89	.43
30	61	94	36	55	.53
31	54	83	11	17	.65
32	61	94	58	89	.15
33	18	28	25	40	-.13
34	56	86	29	45	.47
35	59	91	46	71	.30
36	51	78	29	47	.34
38	65	100	31	50	.72
39	54	83	36	58	.28
40	52	80	31	49	.35
43	30	46	30	48	-.02
44	59	91	39	61	.40
45	46	71	36	58	.13
46	53	82	39	61	.27

Table 3 --Continued.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
47	57	88	36	56	.39
48	54	83	33	52	.34
49	60	92	35	55	.49
50	44	68	26	41	.29

Two important facts were gleaned from the comment sheets of the teachers. First, the teachers had not taught their students in as much detail as did the self-instructional program. Second, it was stated by one teacher that she found through experience that ninth graders need not adjust the upper tension, and consequently she checked and adjusted the tension for her students.

The teachers further indicated that the test required from twenty-five to forty minutes. The amount of time required to complete the test did not seem to be related to the ability of the student. For example, the better student might hurry through the test or cautiously recheck her answers.

Comments made by students, but included in the teachers' reports, seemed to be typical of students. Some indicated the test was easy while others thought it was difficult. A helpful suggestion offered by

students was to have a sewing machine open in the room while the test was being administered.

CHAPTER V

SUMMARY AND RECOMMENDATIONS

This study was planned to develop two tests to accompany the self-instructional program on the use of a sewing machine by ninth grade students in homemaking classes. Test items were developed in light of the objectives of the program. The reliability and the discriminating power of each test item were determined by field testing and standard methods of evaluation.

The forms for the performance test and the matching and multiple choice items in the paper and pencil test were submitted to specialists in clothing and in home economics education to insure uniform approval of the techniques and methods involved. Their suggestions were given consideration in the revision of the tests. Only those items on which all specialists agreed were analyzed.

Development of the performance test was begun by listing errors which might be made in filling the bobbin, threading the upper and lower parts of the machine, adjusting the tension and stitch-length regulator, sewing a 5/8 inch seam, and replacing the zipper foot. These errors were compiled on a check sheet prepared for use by the teacher in administering the test.

At the same time directions for the student and procedures for the teacher were developed to complete the forms for the performance test. The test and the teacher's check sheet were divided into four

similar parts for ease in administering and scoring. A list of testing equipment and directions for setting up the machines for the testing program were included in the procedure for the teacher to follow in administering the test.

After revisions were made, the performance test was administered to six first year home economics students who were representative of students for whom the program and test were planned. Three home economics teachers in the county independently rated each of the students on a check sheet. On the basis of information gleaned from these teachers and students, the test was further revised. A major change in the check sheet was the substitution of positive statements for those which had been stated in a negative form.

This test was given also to students in the school in which the self-instructional program was being used and in the University laboratory school. Although judges did not score each student, constructive comments were received and used in further revision of the test.

The final form of the performance test was administered to thirty-five students in the high school in which the researcher was teaching. These students were scored independently by three judges. Correlations were computed between the scores of the judges. Scores of Judge One correlated .95 with scores of Judge Two and .92 with scores of Judge Three, and scores of Judge Two correlated .91 with scores of Judge Three. Scores of Judge One were used to determine the upper and lower 27 per cent of the group in order that the discriminating power of the items could be computed by Flanagan's method (17, pp. 286-92 and 19, p. 472).

The paper and pencil test, "A Quiz From Sewing Sue," was developed

in a different manner. To get responses for multiple choice items in student terminology, homemaking students were given essay type questions concerning the sewing machine. Fifty multiple choice items with two, three, or four responses were then formulated. The description and function of various machine parts were tested with eighteen matching items.

Test forms and answer sheets were mailed to four schools in the area. Accompanying the tests were instructions for administering the test. Two-hundred and forty-one students took the test, and their scores were used for the analysis.

The four teachers participating in the test were asked to return a comment sheet about the testing situation. Answer sheets for each of the students taking the test were returned and scored with a scoring stencil.

The coefficient of reliability was determined by dividing the test into two subtests which used the odd items for one subtest and the even items for the other subtest. The scores on the two subtests were correlated and a coefficient of reliability of .699 was determined for a test one-half the length of the test. The Spearman-Brown modified formula was used to estimate the reliability of the entire test. The reliability of the test was shown to be .822.

The maximum possible score on the paper and pencil test was .63. Each correct item was given a score of 1. Scores of the 241 students ranged from 24 to 60. The mean was 45.7, and the median was 46.

Items were analyzed on test papers of the sixty-five students who scored in both the upper and lower 27 per cent of the group. Reference to Flanagan's Table (17, pp. 290-1 and 19, p. 472) indicated the

discriminating power of the responses in the test items. Responses with a value of .20 or above were considered good items, while responses with a value less than .20 were considered poor and in need of revision.

Sixteen of the eighteen matching items in Part A and Part B of "A Quiz From Sewing Sue" showed sufficient discrimination between the upper and lower groups of students. Forty of the multiple choice items in Part C showed adequate discrimination with discriminating powers of .20 or above between the upper and lower 27 per cent of the students. All of the responses in multiple choice items were rated as functioning responses since all were selected by more than 3 per cent of the total number of persons taking the test.

The analysis of the data from both the performance and paper and pencil tests suggested the need for a few improvements. The following steps are recommended as a follow-up of this study:

1. Revise selected items in the performance test using the suggestions in Chapter IV.

2. Administer the performance test to additional students for a more dependable item analysis. Thirty-five students were too few for representative item analysis.

3. Use a minimum of three qualified judges with the performance test to compute the coefficients of correlation which showed the extent of agreement among the judges.

4. Select a school in which the self-instructional program is being used and administer the performance test at the conclusion of the program. In this study, the performance test was given to students who had not studied the self-instructional program.

5. Use both the performance test and paper and pencil test in order to determine their effectiveness together. This study did not indicate this factor.

6. Revise the five multiple choice items in Part C of the paper and pencil test which had a discriminating value less than .20.

7. Develop a minimum of three responses for all items when feasible.

8. Administer the revised paper and pencil test to approximately the same number of students as used in this study, and analyze the scores in the same manner.

The researcher suggests that the revised tests could serve the research staff in the School of Home Economics and high school home economics teachers. In the first place, the tests developed in this study contain test items which may be used in the larger study now in progress at the University. Home economics teachers could use both of these tests to determine their students' knowledge of the sewing machine. The paper and pencil test might serve as a pretest before the study of the sewing machine.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. ADKINS, DOROTHY C., et al. Construction and Analysis of Achievement Tests. Washington: Government Printing Office, 1948. 292 pp.

2. AHMANN, J. STANLEY. Evaluating Pupil Growth. Boston: Allyn and Bacon Inc., 1959. 605 pp.

3. ARNY, CLARA BROWN. Evaluation in Home Economics. New York: Appleton-Century-Crofts, Inc., 1953. 378 pp.

4. BRADFIELD, JAMES M., and MURDOCK, H. STEWARD. Measurement and Evaluation in Education. New York: The Macmillian Co., 1957. 509 pp.

5. CHADDERDON, HESTER. Evaluation in Clothing, Ninth Grade. Ames, Iowa: Home Economics Research, Iowa State University, April 1963. (Mimeographed).

6. CHESSMAN, MARY JANE. The Relationship of Height to the Performance of College Women in the Repeated Volley Test. Master's thesis. Greensboro: Woman's College of the University of North Carolina, 1956. 45 pp.

7. CURETON, EDWARD E. "Validity," Educational Measurement. (Edited by E. F. Lindquist.) Washington: American Council on Education, 1951. 819 pp.

8. HALL, OLIVE A., and PAOLUCCI, BEATRICE. Teaching Home Economics. New York: John Wiley and Sons, Inc., 1962. 397 pp.

9. HARDAWAY, MATHIDLE, and MAIER, THOMAS B. Tests and Measurements in Business Education. Second Edition. Cincinnati: South-Western Publishing Co., 1952. 434 pp.

10. HULAC, GEORGIA MAY. The Construction of an Objective Indoor Test for Measuring Effective Tennis Serves. Master's thesis. Greensboro: Woman's College of the University of North Carolina, 1958. 42 pp.

11. KWALWASSER, JACOB. Tests and Measurements in Music. Boston: C. C. Birchard and Company, 1927. 146 pp.

12. McCLOY, C. H. Tests and Measurements in Health and Physical Education. New York: Crofts and Co., 1942. 412 pp.

13. McPHERSON, M. W. "A Method of Objectively Measuring Shop Performance." Journal of Applied Psychology 29: 22-26; 1945.

14. MEYERS, CARLTON R., and BLESH, T. ERWIN. Measurement in Physical Education. New York: The Ronald Press Company, 1962. 473 pp.

15. MOORE, CATHERINE PORTER. Development of a Self-Instructional Program on the Sewing Machine. Master's thesis. Greensboro: University of North Carolina at Greensboro, 1963. 98 pp.

16. RYANS, DAVID G., and FREDERIKSEN, NORMAN. "Performance Tests of Educational Achievement." Educational Measurement. (Edited by E. F. Lindquist.) Washington: American Council on Education, 1951. 819 pp.

17. SCOTT, M. GLADYS, and FRENCH, ESTHER. Evaluation in Physical Education. St. Louis: The C. V. Mosby Company, 1950. 348 pp.

18. SHOFFNER, SARAH MOORE. Revision and Field Test of a Self-Instructional Program on the Sewing Machine. Master's thesis. Greensboro: University of North Carolina at Greensboro, 1964. 140 pp.

19. WALKER, HELEN M., and LEV, JOSEPH. Statistical Inference. New York: Henry Holt and Company, 1953. 510 pp.

20. WEST, CHARLOTTE. A Comparative Study Between Height and Wall Volley Test Scores as Related to Volleyball Playing Ability of Girls and Women. Master's thesis. Greensboro: Woman's College of the University of North Carolina. 1957. 57 pp.

21. WOELFLIN, LESLIE E. An Experimental Study on the Teaching of Clarinet Fingerings with Teaching Machines. Paper Presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, February 18, 1963. (Offset).

22. WRIGHTSTONE, J. WAYNE; JUSTMAN, JOSEPH; ROBBINS, IRVING. Evaluation in Modern Education. New York: American Book Company, 1956. 481 pp.

APPENDIX A

Performance Test

Teacher's Check Sheet

Procedure for Teachers

Tables 4, 5, 6, 7

This is a sample of the Student Instructions.

SEWING SUE'S PERFORMANCE TEST

STUDENT INSTRUCTIONS

Read the directions carefully. Your score will be influenced by how you follow the directions. Do each step as accurately and as rapidly as possible. If you have to wait for the teacher, study the next steps. At each "STOP" sign, raise your hand to call the teacher. She will score what you have done thus far.

PART I BOBBIN

1. Fill an empty bobbin about 1/3 full. Do not cut the threads. Do not remove the bobbin from the bobbin winder.

STOP -- Raise your hand.

PART II THREADING THE MACHINE

1. Thread the machine, both upper and lower parts. Use a different color of thread on the upper part than you use in the bobbin.

2. Bring the bobbin thread up through the hole in the needle plate.

3. Do anything that is necessary so that the machine is ready for you to put the fabric in place and stitch.

STOP -- Raise your hand.

PART III TENSION, STITCH-LENGTH REGULATOR AND STITCHING

1. Set the stitch-length regulator to sew 10 stitches per inch.

2. On the fabric provided, check the machine to see if the tension is correct. Continue to adjust the tension regulator and to check the machine until the tension is correct.

STOP -- Raise your hand. The teacher will ask you to make a row of stitching 5/8 inch from the edge on another piece of fabric so she can check the way you stitch and the way you test for tension.

PART IV ZIPPER FOOT

1. Remove the presser foot.

2. Attach the zipper foot.

STOP -- Raise your hand.

1. Replace the presser foot and turn in all materials.

This is a sample of the Teacher's Check Sheet.

School_____ Name_____

Date_____

SEWING SUE'S PERFORMANCE TEST

TEACHER'S CHECK SHEET

Check each student's work at each "STOP" sign. (Note: The statements indicate the correct performance.) Place an (x) in the blank preceding each statement if the student performs incorrectly. Place a check ( ) in the blank preceding each statement if the student performs correctly.

PART I BOBBIN

Performance

_____ 1. The spool of thread is on the bobbin spool pin.
_____ 2. The thread guide between the bobbin spool pin and the bobbin winder is used.
_____ 3. The bobbin is pushed far enough to latch onto the spindle.
_____ 4. The stop-motion screw is loose.

PART II THREADING THE MACHINE

_____ 1. The stop-motion screw is tightened.
_____ 2. The spool of thread is placed on the spool pin so that it rotates in a counter clockwise direction.
_____ 3. The first thread guide is threaded.
_____ 4. The tension regulator is threaded.
_____ 5. The thread is between the discs of the tension regulator.
_____ 6. The thread pushes against the wire spring.
_____ 7. The thread guide (s) on or near the tension regulator is (are) threaded.
_____ 8. The thread is put through the thread take-up after the tension regulator.
_____ 9. The thread is put through the hole in the thread take-up.
_____10. The thread is put through all the thread guides in the area of the needle.
_____11. The thread take-up is at its highest point.
_____12. The needle is at its highest point.
_____13. The needle is threaded from the same side on which the last thread guide is located.
_____14. The bobbin thread is brought to the surface through the needle hole.

Incorrect performance (x)
Correct performance ( )

_____15. Both threads are pulled under the presser foot and toward the back.
_____16. The slide plate is closed.
_____17. The bobbin thread is placed in the bobbin case so that the bobbin turns in the correct direction for this machine.
_____18. The thread is placed in the slot and under the spring on the bobbin case. The thread lies across the bobbin.

PART III TENSION, STITCH-LENGTH REGULATOR, AND STITCHING

Note: After the student has adjusted the machine, ask her to make another row of stitching as before. Watch her procedure in stitching.

_____ 1. The thread take-up is at its highest point.
_____ 2. The needle is at its highest point.
_____ 3. A diagonal fold is used.
_____ 4. The bulk of the fabric is to the left of the needle.
_____ 5. The needle is placed in the fabric before the presser foot is lowered.
_____ 6. The hand wheel is used to begin the first few stitches.
_____ 7. The line of stitching is 5/8 inch from the edge.
_____ 8. The hand wheel is used to make the last few stitches.
_____ 9. The student stopped before stitching off the edge of the fabric.
_____10. The thread take-up and the needle are left at their highest points.
_____11. The fabric is removed by pulling it toward the back so the threads are still under the presser foot.
_____12. The threads are cut so that three inches are left beyond the needle.
_____13. The stitch-length regulator is set for 10 stitches per inch.
_____14. The screw on the stitch-length regulator is tightened.
_____15. The line of stitching is pulled to see if the threads break.
_____16. The stitching looks the same on both sides of the fabric.

PART IV ZIPPER PRESSER FOOT

_____ 1. The zipper foot is attached to the presser bar properly.
_____ 2. The screw is tight enough so that there is no danger of its becoming loose during stitching.

PROCEDURE FOR TEACHERS

Materials Needed:

1. Two pieces of fabric 6" square for each student. Have some extra squares on hand for use if needed.
2. Two spools of thread, each in a different color, for each student.
3. A pair of scissors or shears for each student.
4. An empty bobbin for each student.
5. A zipper foot for each machine.
6. Cards indicating how to thread the upper and lower parts of each model of machine with which the teacher is not familiar.

Instructions:

1. Each machine should be in perfect adjustment except that the upper tension is loose and the stitch-length regulator is set for the longest stitch that can be made on each machine.
2. The stop-motion screws should be working. If one does not operate properly, do not count it as a student error.
3. Students should use the model of machine with which they are familiar.
4. Explain the general procedure to the students:
    a. They are to work as rapidly, accurately, and as independently as possible.
    b. They will be checked at the conclusion of each part of the test.
    c. The thread on the bobbin will be a different color from that used on the spool.
    d. Their mistakes will be corrected after being checked in order that they may proceed with the test.

Table 4. Analysis of responses of upper and of lower 27 per cent of the students: Performance Test, Part I.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	9	100	9	100	.00
2	9	100	9	100	.00
3	9	100	9	100	.00
4	9	100	6	67	.65

Table 5. Analysis of responses of upper and of lower 27 per cent of the students: Performance Test, Part II.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	9	100	9	100	.00
2	9	100	6	67	.65
3	9	100	9	100	.00
4	9	100	9	100	.00
5	9	100	9	100	.00
6	9	100	8	89	.43
7	8	89	7	78	.16
8	9	100	9	100	.00
9	9	100	9	100	.00
10	9	100	9	100	.00
11	8	89	9	100	-.43
12	8	89	7	78	.16
13	9	100	9	100	.00
14	9	100	8	89	.43
15	9	100	8	89	.43
16	9	100	9	100	.00
17	9	100	8	89	.43
18	9	100	5	56	.70

Table 6. Analysis of responses of upper and of lower 27 per cent of the students: Performance Test, Part III.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	9	100	7	78	.55
2	9	100	8	89	.43
3	7	78	2	22	.56
4	9	100	9	100	.00
5	8	89	4	44	.49
6	8	89	5	56	.39
7	9	100	4	44	.75
8	6	67	1	11	.60
9	7	78	2	22	.56
10	6	67	0	0	.79
11	6	67	3	33	.35
12	7	78	4	44	.36
13	9	100	8	89	.43
14	4	44	4	44	.00
15	7	78	1	11	.68
16	9	100	6	67	.65

Table 7. Analysis of responses of upper and of lower 27 per cent of the students: Performance Test, Part IV.

Item	Number responding correctly: High Group Number	High Group Per Cent	Low Group Number	Low Group Per Cent	r
1	9	100	8	89	.43
2	9	100	8	89	.43

APPENDIX B

A Quiz From Sewing Sue

Answer Sheet

Letters to Teachers

Comment Sheet

This is a sample of the paper and pencil test.

A QUIZ FROM SEWING SUE

PART A. Read each description in column 1. Select from column 2 the part of the sewing machine that is described. Place the number which identifies the sewing machine part in the blank provided on the answer sheet. The parts in column 2 may be used more than once or they may not be used at all.

Descriptions	Parts
____1. Spindle on which the spool rests.	1. bobbin
____2. Wheel used for starting and stopping the machine.	2. bobbin case
____3. Container in which the bobbin is placed.	3. bobbin winder
____4. Cover which can be pushed aside so you can remove the bobbin.	4. feed dog
____5. Metal foot with two toes near the needle.	5. hand wheel
____6. Rubber wheel and metal spindle used to wind the bobbin.	6. presser foot
____7. Metal spool on which the thread is wound to furnish the lower thread on the machine.	7. slide plate
____8. Metal "teeth" under the presser foot which carry the cloth forward.	8. spool pin
____9. Metal holder which contains a spring and a screw which regulate the bobbin tension.	9. tension regulator
	10. thread guide
	11. thread take-up

A QUIZ FROM SEWING SUE (Continued)

PART B. Read each purpose listed in column 1. Select from the list of sewing machine parts the <u>one</u> which performs this work. Write its number in the blank provided on the answer sheet. The parts in column 2 may be used more than once or they may not be used at all.

Purposes	Parts
____1. Adjusts the length of stitches.	1. bobbin
____2. Moves the material along.	2. bobbin case
____3. Regulates the tightness or looseness of the stitches.	3. feed dog
____4. Pulls up the slack in the thread and locks the stitch each time the needle comes up.	4. hand wheel
____5. Holds the cloth in place.	5. presser foot
____6. Moves up and down as a stitch is formed.	6. slide plate
____7. Is turned to move the needle and the thread take-up to correct position for beginning and ending a seam.	7. spool pin
____8. Keeps the thread in correct position as it goes from the spool to the upper tension.	8. stitch-length regulator
____9. Holds the bobbin and keeps the lower thread under tension.	9. tension regulator
	10. thread guide
	11. thread take-up

A QUIZ FROM SEWING SUE (Continued)

PART C. Each of the questions or incomplete statements in this part of the quiz is followed by a number of suggested answers. Decide which <u>one</u> of these answers is correct. On the answer sheet cross out the letter of the answer you choose.

Example: Which of the following is an animal?
A. toy
B. car
C. cat
D. foot

(On answer sheet) A B C D

1. If you are asked to thread a sewing machine with which you are not familiar, in which direction will you thread the needle?

A. away from you.
B. from the flat side of the needle.
C. from the side of the last thread guide.
D. from the side of the needle eye.

2. How can the thread take-up be identified?

A. by watching to see what part, in addition to the needle, goes up and down as a stitch is formed.
B. by finding the part of the top of the machine which has a hole through the center.
C. by looking at the top of the machine.

3. How is the bobbin case threaded?

A. attach the bobbin to the bobbin case.
B. put the thread on the lower spool pin.
C. lead the thread into the slot in the bobbin case and under the spring.
D. push the bobbin until you hear it click.

4. Where should the ends of the thread lie when you have completed a seam and cut the threads?

A. under the presser foot.
B. under and behind the presser foot.
C. behind the presser foot.
D. toward you.

5. Why is it important to put your hand on the hand wheel when stopping at the end of the seam?

   A. to prevent breaking the needle.
   B. to raise the needle and the thread take-up.
   C. to put the needle on the cloth.
   D. to be sure you do not run off the fabric.

6. Where should your hands be placed before you begin sewing?

   A. one hand on the fabric and the other behind the presser foot.
   B. left hand on the fabric and the right hand on the hand wheel.
   C. left hand on the fabric and the right hand behind the needle.
   D. both hands on the fabric.

7. Mary's machine has more thread guides than your machine has. How can you find these extra thread guides?

   A. they are located at the front of the machine near the tension regulator.
   B. they have an opening for the thread and are in a place where the thread needs guiding.
   C. they can be located by studying the picture of the sewing machine in the instruction book.

8. When you are using Mary's machine, what should you do about the extra thread guides?

   A. run the thread through each one in order until the needle is threaded.
   B. try till the machine works correctly.
   C. use only the ones that correspond to your machine.
   D. study the instruction book.

9. What is the reason for loosening the stop-motion screw when filling the bobbin?

   A. so that the threads will not tangle in the bobbin case.
   B. so the bobbin will thread evenly.
   C. so the needle will not go up and down.
   D. so the bobbin will not rotate.

10. What is the procedure to follow when you reach the end of a seam?

   A. stop, using the hand wheel, with the thread take-up and the needle at their highest points.
   B. tie the threads to see if the seam is straight.
   C. leave the needle in the fabric and lift the presser foot with the presser foot lifter.
   D. see if the needle is threaded and the threads are pulled back.

11. Where does the threading of the sewing machine begin?

A. at the first thread guide.
B. from the top to the first thread guide.
C. from the spool pin to the first thread guide.
D. from the thread guide to the thread take-up.

12. Suppose that you are threading a machine which you have not seen before. In what order would you thread the important parts of the machine?

A. spool pin, thread take-up, tension regulator, thread guide, needle.
B. spool pin, stop-motion screw, bobbin winder, needle.
C. spool pin, thread guide, tension regulator, thread take-up, needle.
D. spool pin, thread guide, thread take-up, feed dog, needle.

13. Which terms describe the kinds of power by which sewing machines are operated?

A. electric and threaddle.
B. treacle and electric.
C. power and hand.
D. treadle and electric.

14. When the knob on the tension regulator is turned to the right, what happens to the numbers?

A. they become larger.
B. they become smaller.
C. they do not change.

15. Which of the following is raised or lowered by a lifter?

A. presser bar.
B. pressure bar.
C. needle bar.

16. When you are ready to stitch, which of the following describes what happens to both threads?

A. both threads are tied together to keep them together.
B. both threads are pulled to the back, the upper one between the toes of the presser foot.
C. both threads are pulled toward the person who is sewing.
D. any of the above as it makes no difference in sewing.

17. What is the correct procedure for turning a square corner in the fabric?

A. stop at the corner, raise the needle and the presser foot, line up the seam, and begin stitching again.
B. stop with the needle in the fabric, raise the presser foot, turn the fabric, lower the presser foot, and begin stitching.
C. make a small curve around the corner sewing slowly.
D. stop at the raw edge, backstitch to the corner, lift the presser foot, turn the fabric, begin stitching again.

18. One way to determine whether the tension is correctly adjusted is to stitch on a bias fold, then give a quick jerk to the stitching. If both threads break, what is the cause?

A. the stitches are too small.
B. the stitches are too large.
C. the tension is the same on the upper and lower threads.
D. the tension is incorrect.

19. When you are testing the tension and the upper thread breaks, what is the cause?

A. the upper tension is too tight.
B. the lower tension is too tight.
C. both tensions are the same.

20. When you are testing the tension and the upper thread breaks, what should be done?

A. the upper tension should be tightened.
B. the upper tension should be loosened.
C. the lower tension should be tightened.
D. the lower tension should be loosened.

21. How do you know from which side to thread the needle?

A. always thread it from the left side.
B. always thread it from the right side.
C. note the location of the last thread guide and thread it from that side.
D. always thread it from front to back.

22. Which machine part uses two closely fitted discs to control the looseness or tightness of the thread?

A. thread take-up.
B. upper tension.
C. lower tension.
D. stitch-length regulator.

23. Which kind of fabric requires the longest stitch?

A. thick fabrics.
B. sheer fabrics.
C. medium weight fabrics.
D. it makes no difference.

24. How can stitches be made longer or shorter?

A. by adjusting the seam guide.
B. by adjusting the stitch-length regulator.
C. by adjusting the tension regulator.

25. What machine part (or parts) controls the tension on the needle thread?

A. the upper tension.
B. the lower tension.
C. both the upper and lower tension.

26. Which of the following parts are located near each other in the order of threading?

A. the tension regulator and the spool pin.
B. the last thread guide and the short groove on the needle.
C. the last thread guide and the long groove on the needle.
D. the tension regulator and the last thread guide.

27. The machine will not stitch unless the thread passes through the hole in one of the following parts. Which part is it?

A. thread take-up.
B. tension regulator.
C. presser bar.

28. Which hand is used to guide your fabric?

A. the left hand.
B. the right hand.
C. either hand.
D. depends upon whether you are right or left handed.

29. Which of the following parts is found on the hand wheel?

A. close fitting discs.
B. stop-motion screw.
C. wire spring.

30. What are the three parts which move up and down as stitches are made?

A. thread take-up, needle, feed dog.
B. thread take-up, needle, upper tension.
C. upper tension, needle, last thread guide.

31. A wire spring is an essential part to which of the following?

A. upper tension.
B. lower tension.
C. stitch-length regulator.

32. At what level should the needle be when you are working with the underneath part of the sewing machine?

A. at its highest point.
B. at its lowest point.
C. it makes no difference.

33. If the needle is put in backwards what will be the result?

A. the needle will break.
B. the needle will come unthreaded.
C. it is not possible to put it in backwards.
D. the tension will be too loose on the upper thread.

34. What kind of groove (s) does (do) a machine needle have?

A. a long and a short groove.
B. two grooves the same size.
C. one long groove.
D. the tension will be too loose on the upper thread.

35. What holds the needle in place?

A. a screw on the needle bar.
B. a screw on the presser bar.
C. the tension regulator.
D. the presser foot.

36. Which number indicates the greatest tension on the upper thread?

A. 0
B. 2
C. 5
D. 7

37. What are the two main parts of the tension regulator?

A. a wire spring and discs.
B. a wire spring and a screw.
C. a wheel and a thread guide.
D. a screw and the backstitch lever.

38. When a machine stitches so that the upper and lower threads lock in the center of the fabric, what does one conclude?

A. the tension is evenly adjusted.
B. the correct size stitch is used.
C. the proper needle is used.

39. What are the two types of controls found in electric machines?

A. hand and foot controls.
B. knee and foot controls.
C. belt and foot controls.
D. hand and knee controls.

40. Which of the following parts is stationary in some machines and removable in others?

A. needle.
B. bobbin case.
C. presser foot.
D. slide plate.

41. How do you lift the presser-foot lifter?

A. reach under the head of the machine and to the right of the needle with your right hand.
B. reach under the head of the machine and to the left of the needle with your left hand.
C. reach under the head of the machine with the most convenient hand.

42. What must be done after the upper and lower parts have been threaded in order that the machine be completely ready to sew?

A. adjust the stitch-length regulator so that the stitch is neither too long or too short.
B. bring the bobbin thread up through the hole in the needle plate and pull the threads to the front.
C. pull the two threads back between the toes of the presser foot and raise the thread take-up to its highest point.

43. On which side of the head of the needle is the short groove?

A. the flat side.
B. the round side.
C. neither.

44. What direction is the stop-motion screw turned to loosen it?

A. toward you as you hold the hand wheel.
B. away from you as you hold the hand wheel.

45. How do you tighten the upper tension?

A. turn the knob to the left.
B. turn the knob to the right.

46. In what direction should you turn the screw on the stitch-length regulator to tighten it after the length of stitch has been correctly set?

A. to the left.
B. to the right.

47. In what direction do you turn the hand wheel of a Singer machine to raise the needle and the thread take-up to their highest points?

A. toward you.
B. away from you.

48. In which direction should the spool of thread turn on the spool pin?

A. clockwise
B. counter clockwise.

49. In what order are the thread take-up and tension regulator threaded?

A. first the thread take-up, then the tension regulator.
B. first the tension regulator, then the thread take-up.

50. Where is the screw which holds the needle in the needle bar?

A. on the right side of the needle bar.
B. on the left side of the needle bar.

This is a sample of the answer sheet.

A QUIZ FROM SEWING SUE

Name_____

School_____

Date_____

PART A.

1._____
2._____
3._____
4._____
5._____
6._____
7._____
8._____
9._____

PART B.

1._____
2._____
3._____
4._____
5._____
6._____
7._____
8._____
9._____

PART C.

1. A B C D
2. A B C
3. A B C D
4. A B C D
5. A B C D
6. A B C D
7. A B C
8. A B C D
9. A B C D
10. A B C D
11. A B C D
12. A B C D
13. A B C D
14. A B C
15. A B C
16. A B C D
17. A B C D
18. A B C D
19. A B C
20. A B C D
21. A B C D
22. A B C D
23. A B C D
24. A B C
25. A B C
26. A B C D
27. A B C
28. A B C D
29. A B C
30. A B C
31. A B C
32. A B C
33. A B C D
34. A B C D
35. A B C D
36. A B C D
37. A B C D
38. A B C

PART C. (continued)

39. A B C D

40. A B C D

41. A B C

42. A B C

43. A B C

44. A B

45. A B

46. A B

47. A B

48. A B

49. A B

50. A B

Sewing Sue wishes you--

Happy Machine Sewing!

310 West Fifth Avenue
Lexington, North Carolina
April 6, 1964

Dear

The Quiz From Sewing Sue is now ready to be tested. I am sending in a separate package copies of the test booklet and copies of the answer sheet. If you find that it is not practical to use the same test booklets with your different classes, I will send additional copies of the tests.

These directions will help in administering the test.

1. See that each student has an answer sheet and a test booklet.

2. Impress upon them the importance of putting answers on the answer sheet and not in the test booklet. You will have to check to see that they do not write in the booklets before using them again.

3. Go over the directions for the test with the students and be sure that they understand the marking procedure.

4. Record the time it takes to administer the test.

5. Fill out the enclosed form and return it with the answer sheets.

Please give the tests as soon as possible. Return only the answer sheets and your comments. Stamps have been placed on the return envelope. You may keep the test booklets for future use or destroy them as you wish.

Thank you again for your help.

Sincerely,

(Miss) Carolyn Ross

This is a sample of the comment sheet.

COMMENTS

School_____

Number of students taking the test_____

Approximate date students received instruction in use of sewing machine_____

Approximate time required to complete the test:

For the best students_____

For the average students_____

For the slow students_____

Comments of students concerning the test:

Suggestions: