

The Woman's College of  
The University of North Carolina  
LIBRARY



CQ  
no.508

COLLEGE COLLECTION

Gift of  
Mary McNeely Solomon

SOLOMON- MARY McNEELY; Development of an Evaluation Device to Measure Quality of Workmanship on a Blouse. (1967) Directed by: Dr. Hildegard Johnson. pp. 105.

The objectives of this study were to develop, test, and revise an evaluation device for quantitatively scoring the quality of blouses constructed by students who learned to sew by following a self-instructional program. The program was developed as part of United States Office of Education research project No. 5-1042.

Various kinds of evaluation devices were studied. Since rating scales are most frequently used in areas where measurement is dependent upon observation, a rating scale was selected as the appropriate device for this quantitative scoring of the blouse.

A rating scale in dichotomous form, on which the judge recorded that the blouse was satisfactory or unsatisfactory with respect to each descriptive phrase, was first tried but later discarded in favor of a three-level scale on which judges could appraise each construction process by considering several levels of quality. The three-level scale was then subdivided to form a six point scale.

The rating scale was organized into nine sections: (1) grainlines, (2) staystitching, (3) plain seams, (4) facings, (5) darts, (6) sleeves, (7) sleeve hems, (8) blouse hems, and (9) general appearance. These sections included 128 items. A supplement, containing both diagrams and explanations and titled "Instructions to the Judges", accompanied the rating scale.

Twenty-one blouses secured from Home Economics I students and from research staff members were used for testing the scale. Five judges, four juniors in home economics education and one a graduate student in clothing, participated in three training sessions before

scoring the blouses during a period of one month.

The development of an objective rating scale with high reliability was one of the desired outcomes of this study. Items on the scale were analyzed and suggestions made for revision and elimination of items when the data indicated that judges disagreed with respect to scores on an item or time required for scoring was disproportionate to amount of information gained.

The mean scoring time was 100 minutes. The range of scores for the blouses indicated the device distinguished among blouses of varying quality.

Reliability for the scale was based on intercorrelations among pairs of judges. These correlations ranged from .65 to .89, the average being .82. Although the reliability was considered satisfactory, it was believed the reliability could be improved if some of the descriptive statements for the scale had been worded more briefly and, in a few instances, more objectively. The scale still required occasional subjective ratings that resulted in disagreement among judges' ratings.

Since the rating scale was developed for use in the research project, the scale is too lengthy and detailed to be used in its present form by high school home economics teachers. The researcher hopes the rating scale through its emphasis on clear instructions, accurate visual aids, and definite categories or limits for varying levels of quality of a product can serve as a guide to teachers when they develop their own evaluation devices.

DEVELOPMENT OF AN EVALUATION DEVICE  
TO MEASURE QUALITY OF WORKMANSHIP ON A BLOUSE

by

Mary McNeely Solomon

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 in  
Home Economics

Greensboro  
April, 1967

Approved by

*Hildegard Johnson*

Director



APPROVAL SHEET

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

Thesis  
Director

Hildegarde Johnson

Oral Examination  
Committee Members

Gail M. Dennis

Nancy White

Bonnie Lowe

Hubert K. Staley

April 17, 1967

Date of Examination

## ACKNOWLEDGMENTS

The researcher would like to express her appreciation to the committee members who offered not only their time but also their guidance in order that this thesis could be completed. The researcher especially wants to thank her major professor, Dr. Hildegard Johnson; without her guidance and assistance, this thesis would have been greatly lacking.

Appreciation is also expressed to "Scrooge" for his endurance. The researcher would also like to express appreciation to the faculty with whom she worked while a graduate student at the University of North Carolina at Greensboro in the School of Home Economics.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS . . . . .	iii
LIST OF TABLE . . . . .	vi
LIST OF FIGURES. . . . .	vii
 CHAPTER	
I. INTRODUCTION . . . . .	1
Philosophy Guiding the Development of the Rating Scale. . . . .	1
Purpose of the Study . . . . .	3
Definitions of Terms Used. . . . .	3
II. REVIEW OF LITERATURE . . . . .	5
Differences Among Checklists, Rank Order Scales, and Rating Scales. . . . .	5
Definitions and Examples of Types of Rating Scales . .	6
Graphic Rating Scale . . . . .	7
Product Rating Scale . . . . .	8
Examples of Rating Scales. . . . .	9
Need for Improvement of Rating Scales. . . . .	11
Decisions Which Influence the Quality of Rating Scales . . . . .	12
Number of Scale Intervals. . . . .	12
Spacing of the Descriptions on the Rating Line . . .	13
Phrasing of Descriptions . . . . .	14
Weighting of Items . . . . .	15
Validity . . . . .	16
Reliability. . . . .	17
Selection and Training of Judges . . . . .	18
Training of Judges . . . . .	18
Number of Judges . . . . .	19
Errors of Judges . . . . .	20

CHAPTER	Page
III. METHOD OF PROCEDURE . . . . .	23
Content of the Rating Scale . . . . .	23
Form for the Rating Scale . . . . .	28
Selection and Training of Judges. . . . .	31
Scoring of Blouses. . . . .	33
IV. FINDINGS. . . . .	34
Description of the Scale. . . . .	34
Procedure for Training Judges . . . . .	35
Increasing the Efficiency of the Scale. . . . .	37
Analysis of Judges' Scoring Time. . . . .	45
Blouse Scores . . . . .	50
V. SUMMARY . . . . .	54
BIBLIOGRAPHY. . . . .	59
APPENDIX. . . . .	62

## LIST OF TABLES

Table	Page
1. Sub-Scores on the Rating Scale . . . . .	36
2. Scores, Item Nine, Width of Seam Allowances. . . . .	38
3. Number and Percentage of Blouses for Which Pairs of Judges Agreed on Scoring of Shoulder, Side, and Armseye Seams (Item 9) . . . . .	40
4. Scores on Item 72, Scorching and Shininess from Overpressing . . . . .	41
5. Scores on Item 2 for Right and Left Sides of Garment. . . . .	44
6. Number of Times Judges Scored Item 2A and 2B the Same f for Right and Left Sides . . . . .	43
7. Minutes Used by Judges to Score Sections of Six Blouses. . . . .	47
8. Correlations Between Scores of Pairs of Judges . . . . .	53



LIST OF FIGURES

Figure	Page
1. Time Used By Five Judges to Score Every Other Blouse . . . .	46

Research in the field of education has made rapid advances in recent years, resulting in increased demand for improved instruments for measuring outcomes of experiential research. Great measurements in many fields of evidence has been possible and the availability of instruments for exact measurement contributed to the use of the scientific method. An improvement of measurement techniques in education will stimulate new to improved empirical research in education.

Development During the Development of the Rating Scale

As the number of research projects in education has expanded, the demand for improved measuring instruments has increased. However, in the area of educational research, instruments of measurement necessary for objective evaluation of the experiments do not exist. There is a need for measuring instruments which have been designed with acceptable standards.

Some authors expressed the opinion that the quality of evaluation activities has not improved sufficiently to meet the needs of research. Weightman mentioned that the term evaluation is relatively new and she stated that a very definite purpose of evaluation was to "designate a new comprehensive concept of measurement that is implied in conventional tests and similar items" (25, p. 309).

That said that evaluation "... must be considered as a

## CHAPTER I

### INTRODUCTION

Research in the field of education has made rapid advances in recent years, creating an increased demand for improved instruments for measuring outcomes of experimental research. Exact measurement in many fields of science has been possible and the availability of instruments for exact measurement contributed to the use of the scientific method. An improvement of measurement techniques in education will likewise lead to improved empirical research in education.

#### Philosophy Guiding the Development of the Rating Scale

As the number of research projects in education has expanded, the demand for improved measuring instruments has increased. Moreover, in some areas of educational research, instruments of measurement necessary for effective evaluation of the experiments do not exist. There is a need for measuring instruments which have been completed with acceptable standards.

Some authors expressed the opinion that the quality of evaluation devices has not improved sufficiently to meet the needs of research. Wrightstone mentioned that the term evaluation is relatively new and also stated that a very definite purpose of evaluation was to ". . . designate a more comprehensive concept of measurement than is implied in conventional tests and examinations" (25, p. 929).

Fleck said that evaluation ". . . must be considered an unde-

veloped frontier" (9, p. 300), and she stressed an acute need for evaluation in home economics. Blackwell further emphasized this when she said: "In home economics, at least, few of the professional research workers and full time graduate students have come up with instruments of high quality" (7, p. 87).

The need for evaluation instruments will be met as researchers diligently apply themselves to investigations in the area of evaluation. Fleck stated: "Effective evaluation does not materialize through caprice, intuition, or accident" (10, p. 9), and Leonard and Eurich (3, p. 286) stressed that effective evaluation presents mankind with a challenge for the greatest creative inventiveness and ingenuity he can attain.

The need for evaluation devices with high reliability was apparent at the time a proposal for research in home economics education was submitted by Johnson<sup>1</sup> to the United States Office of Education. Johnson's proposal concerned a field experiment in which two methods of teaching were to be compared. The evidence of superiority of one method over the other would be based on students' learnings as measured by five evaluation devices. One of these devices was to be a rating scale for scoring quality of workmanship on blouses made in the field experiment. The proposal by Johnson was funded by the United States Office of Education and the need for the blouse scoring device intensified. No rating scale that met the needs of the field experiment was available.

---

<sup>1</sup>Dr. Hildegard Johnson, Chairman of the Area of Home Economics Education, School of Home Economics, the University of North Carolina at Greensboro, North Carolina.

### Purpose of the Study

The purpose of this study was to construct an evaluation device for scoring quantitatively the quality of workmanship on the blouses made by students participating in the field experiment. Quantification of quality of workmanship could be accomplished by an evaluation device known as a rating scale. Authorities have agreed that one common objective of rating scales is to quantify data resulting from student performance. Bradfield and Moredock (8, p. 59) stated that rating scales are most frequently used in areas where measurement is dependent upon observations. Since a measurement of quality of workmanship on the blouses must rely on observations made by judges, a rating scale was selected as the method for recording and quantifying these observations.

Further objectives of the study were to test the device, make the necessary revisions, and select and train judges who would use the device to score blouses in the field experiment. One use of the data obtained when scoring the blouses was to revise frames in the self-instructional program--one of the methods of teaching being compared in the field experiment. Although the resulting rating scale would be designed to meet the specific needs of the research project, it was hoped that home economic teachers would use the scale as a reference when developing rating scales for their own needs.

### Definitions of Terms Used

Two terms frequently used in this thesis are rating scale and





## CHAPTER II

### REVIEW OF LITERATURE

The purpose of this study was to develop a rating scale to appraise the quality of workmanship on a blouse. In preparation for developing the rating scale, recommendations made by authorities in the field of evaluation concerning the development of rating scales were reviewed. Rating scales have two uses: (1) the rating of products and (2) the rating of people. Emphasis in this review was given to the types of rating scales appropriate for the scoring of products, the forms of these scales, and the selection and training of judges. Authorities were concerned with such aspects of rating scales as form, selection and training of judges, and errors made by judges.

#### Differences Among Checklists, Rank Order Scales, and Rating Scales

The checklist merely allowed for indication of the presence, absence, or frequency of occurrence of stated characteristics of a product (19, p. 163). Checking a statement indicated that this characteristic was observable but did not involve any judgment of the degree to which the characteristic was present (4, p. 187). Any number of statements might be checked (19, p. 163).

The rank order method allowed for the ranking of products on the basis of comparison of quality. The product exhibiting the highest level of quality could be ranked first, and the one exhibiting the

lowest level of quality ranked last. A serial order resulted from this ranking of products (6, p. 78). This method was used when a product was to be evaluated as a whole--rather than evaluating segments of the product separately (6, p. 13).

The rating scale method was used to quantify judgments from observations (25, p. 163) and these judgments were usually recorded on a scale of units or values (6, p. 74). This method was used when a characteristic could occur in varying degrees (19, p. 163).

The data with which one may work differ when the three types of devices discussed above were used. When checklists were used, the outcomes were lists of descriptive phrases, none of which could be quantified (5, p. 187). The result of using rank order procedure was a group of products which have been placed in proper sequence (6, p. 78). When rating scales were used, the product to be evaluated was scored numerically (8, p. 417). Thus, it is apparent that for research purposes, rating scales are preferable to checklists or rank order scales since a rating scale would result in numerical values which might be treated as measurement data (6, p. 81).

#### Definitions and Examples of Types of Rating Scales

There are several different types of rating scales, the classifications listed most frequently being: descriptive, numerical, graphic, product, and man-to-man or paired-comparison. Some of these scales were used to rate humans, some to rate products, while others were used to rate both humans and products. This review has been limited to the two types of rating scales that could be used to rate products--graphic

scales and product scales.

#### Graphic Rating Scale

Bradfield and Moredock (8, p. 417) stated that the graphic rating scale is made up of descriptive phrases stating varying levels of quality related to specific traits of a product. Sometimes these descriptive phrases apply to one trait, sometimes to more than one. Schwartz and Tiedman (19, p. 165) said that these phrases are printed horizontally at various intervals underneath a straight line across the page. Bradfield and Moredock (8, p. 417) added that numbers corresponding to these phrases are found on this line and that the rating of each trait is indicated by placing a check mark on the line at the appropriate number or between the numbers.

The term "continuum rating scale" was interpreted as meaning essentially the same thing as the term "graphic rating scale". Bradfield and Moredock (8, p. 55) identified the continuum scale as one that had a line representing the variation between two extremes. The rater checked a point on the line to indicate the status of the product. Ahmann and Glock described the continuum style by these statements.

Characteristics of the performance or product are identified and a continuum representing degrees of merit is established for each characteristic. . . . Various positions along the continuum are often identified by numbers and by a brief description of the degree of merit observed (2, p. 225).

It was easy to recognize in these statements the factors that are characteristic of the graphic style of rating.

Gerberich, Greene, and Jorgensen (12, p. 250) stated that the graphic rating scale or variations of that form were widely used rating

devices. An example similar to a graphic rating scale as reported by Ahmann and Glock (2, p. 238) follows.

---

1	2	3	4	5	6	7	8	9	10	Score
A) Straightness: Are nails driven straight, heads square with the wood, no evidence of bending?										_____

---

In this case the descriptions were in the form of questions rather than descriptive phrases. Moreover, the questions described the optimum level of performance. The rater had to form his own low-level descriptions to correspond with the larger numbers. The rater was instructed to fill in the number in the blank at the right that most accurately indicated the quality of the product being evaluated.

#### Product Rating Scale

Army; Barr, Davis, and Johnson; and Wrightstone, Justman, and Robbins presented similar definitions of a product rating scale. Wrightstone, Justman, and Robbins (24, p. 169) defined a product rating scale as one consisting of a series of products arranged in a sequence determined by a jury. Army (4, p. 189) stated that by the application of proper statistical techniques, a limited number of sample products that would vary from one another by approximately equal amounts of quality could be selected from this series. This limited number of sample products would represent a graded series with range from a poor to a good product.

Wrightstone, Justmann and Robbins (24, p. 169) explained that a product being rated was matched with a product on this scale and assigned

the numerical value associated with the product on the scale. Barr, Davis, and Johnson (6, p. 80) said that this matching allowed for quantification since the product being evaluated was matched to a specific product on the scale to which researchers had previously assigned a definite score value.

According to Barr, Davis, and Johnson, product rating scales involve procedures typical of both rank order and rating scales.

Quantification in the case of a product scale involves certain aspects of rank-order scaling. The actual ranking of products is completed before the product scale is used to evaluate the merit of a given specimen. . . . Assignment to a position on the scale of values involves an act of rating (6, p. 80).

Army distinguished between product and graphic rating scales in this way: "Product scales actually illustrate instead of using verbal descriptions to define different levels of quality" (3, pp. 284-5).

Barr, Davis, and Johnson emphasized further:

If this rating is accurately made in the case of a specimen of the product, the process of using a product scale results in achievement of a certain amount of objectivity, since the scale may be standardized upon observable characteristics of the product desired (6, p. 80).

#### Examples of Rating Scales

Most rating scales used in home economics are modified graphic rating scales. An example of such a scale was found in the "Illinois Teacher" (7, pp. 88-90). It represented a modification because making check marks along the continuum was replaced by recording of scores in a column; a section of this scale follows.

The directions were: ". . . score yourself on each item, rating 10, 20, or 30 depending upon whether your behavior corresponds to the



description in the left-hand column, in the right-hand column, or falls between the two" (7, p. 88).

---

SCORE SHEET ON GARMENT MAKING

---

Descriptions	10	20	30	Score
.....				
<u>Applied Sciences</u>				
29. Uses poor techniques in pressing fabric during construction			Understands and follows label directions for care of fabric	_____
30. Abuses machine and other equipment, as leaving water in iron			Applies physics principles to care of equipment used in classroom	_____
31. Fails to see relationship between finish of fabric and method of handling			Is aware of implication for handling of any fabric with chemically treated finish	_____

(7, p. 90)

Another example of a modification of a graphic scale was developed by Huntzicker. A portion of this device follows.

---

 APPEARANCE, BECOMINGNESS, FIT
 

---

Item:	5*	4	3	2	1	Score
Fabric	Satisfactory choice for individual and for the blouse design		Poor choice for individual or for blouse design			_____
	Attractive, well designed fabric		Unattractive, poorly designed fabric			_____
Color	Becoming, suitable color choice for individual		Poor color choice for individual			_____

---

 \*Excellent

(14, p. 111)

Need for Improvement of Rating Scales

The development of high quality rating scales has been slow. Leonard and Eurich (3, p. 2) stated that standardized rating scales have not been improved to the point that they can be used for evaluating all types of behavior. Various types of standardized rating scales can be found on the market; however, these scales are not popular. The lack of normative data accompanying the scales appears not to support their purchase. Schwartz and Tiedman (19, p. 166) emphasized that standardized scales have not as yet been constructed that meet the local needs of the school or teacher.

Improvement is required because rating scales are important instruments of measurement. There is a need to increase the objectivity of rating. Army (5, p. 168) reported that first attempts at developing rating scales were unsuccessful, as evidenced by the fact that ratings of different judges did not agree any more closely than did teachers' scores on essay tests. The improvement of rating scales has been slow because

crude techniques have been used, and "highly objective rating devices are very difficult to construct. . . " (4, p. 165).

#### Decisions Which Influence the Quality of Rating Scales

A number of suggestions for improving rating scales were given in the literature. In constructing a rating scale, there are many decisions that must be made which will affect the quality of the device. These decisions are concerned not only with the form of the rating scale but also with the selection and training of the judges using the scale.

#### Number of Scale Intervals

The number of scale intervals into which to divide the scale is one of the major decisions to be made when developing rating scales. Schwartz and Tiedman (19, p. 190) and Bradfield and Moredock (8, p. 58) agreed that there should be more than two and less than ten. Army (5, p. 192) expressed strong preference for three intervals. Remmers and Gage (17, p. 369) and Latchaw and Brown (15, p. 149) recommended three or five intervals. Seven intervals was suggested as optimal by Symonds (21, p. 83); Wrightstone, Justman, and Robbins (24, p. 169); and Barr, Davis, and Johnson (6, p. 109).

Principles for making this decision were suggested in two references. These principles could be applied each time the test constructor must decide the number of intervals necessary. The first rule as stated by Barr, Davis, and Johnson was: "The ability of the rater to discriminate imposes a limit upon the number of degrees of discrimination which are effective" (6, p. 109). They defended this rule when they stated:

In a rating scale, the more steps that involve fine distinctions, the more frequently are errors in judgment likely to result. . . . because of the rater's inability to discriminate among fine shades of gradation (6, p. 109).

The second rule for deciding the number of intervals was given by Bradfield and Moredock:

The principle to be followed in designing a rating scale is that the number of scale intervals should approximate the number of clearly discernible differences in the dimension being appraised. For measurement tasks that require great precision, a greater number of scale units may be necessary ( 8, p. 58).

The test constructor when applying Bradfield's and Moredock's rule, might first ask a person who is qualified to be a judge to rank a series of products with respect to the item being scored. If it were impossible for the person to distinguish between the samples of progressive quality, the number of samples should perhaps be reduced until a point was reached at which the person could clearly specify at which interval the sample belonged. According to Bradfield's and Moredock's rule, the number of intervals that could be used is, therefore, a function of the difficulty of placing samples at the various intervals.

#### Spacing of the Descriptions on the Rating Line

Another major decision to be made when developing a rating scale concerned the spacing of descriptions on the rating line. Weinland (22, p. 130) cautioned against divisions such as spacing the descriptions evenly, believing that equal spacing inclines the judge to check the average category without giving sufficient thought to the process. Weinland stressed that since overworking the average description was so

frequently a problem, care must be taken to prevent raters from doing this. Schwartz and Tiedman (19, p. 168) and Remmers and Gage (17, p. 369) suggested frequent interchanging of the top and bottom ends of the scale to keep the judges alert and to prevent checking from becoming automatic.

#### Phrasing of Descriptions

Two decisions must be made in relation to developing the descriptions for the rating line. They are concerned with: (1) the time to establish the wording for the phrases and (2) the construction form that will result in these phrases being useful and comprehensive to all raters.

In regard to the problem of time to word the phrases, authorities seemed to be in general agreement that the phrases describing the product should be stated prior to setting up the rating scale. Latchaw and Brown (15, p. 191) and Weinland (22, p. 130) recommended working with the wording of the descriptions and setting up a description of what is considered good before actually constructing the rating scale. Weinland further suggested a method of procedure to be followed that would enable the developer of a rating scale to find the best descriptive adjectives for use in the phrases on the rating line. The developer should first make a list of similar adjectives that describe a trait and then ask others to indicate which of the adjectives are the most meaningful to them. A tallying of the opinions would indicate the descriptive adjectives that are best for describing a specific trait. Sells (20, p. 429) approved Weinland's method of procedure and said that the reliability of



scales could be improved by using this method.

When discussing the characteristics that make a descriptive phrase comprehensive and useful to all raters, most authorities emphasized specificity. These authorities stressed that judges must be in agreement with respect to what is being rated and that this was possible only when the descriptive phrases were specific. Both Latchaw and Brown (15, p. 194) and Mahler (16, p. 72) urged specificity when they warned against the use of such ambiguous words as good, average, or excellent.

Wrightstone (25, p. 929) pointed out the need for specific definitions for descriptive phrases. He suggested that specificity could best be accomplished by: (1) using descriptive phrases that contain adequate definitions and (2) avoiding overlapping of traits. Mahler (16, p. 69) agreed with Wrightstone that one of the major faults of descriptive phrases is that frequently each phrase does not refer to a single type of activity. Weinland (22, p. 132) suggested avoidance of exaggerating the descriptive phrases at either end of the rating line since the rater tends to select the more conservative middle range of phrases.

#### Weighting of Items

The selection of items to be weighted is another decision that must be made. Authorities presented varying opinions concerning the necessity of weighting. There was no consensus among the authorities; rather, there seemed to be two contradictory ideas presented in the literature.

Mahler (16, p. 74) stated one of these when he indicated that weighting of individual items was not always required and might be detrimental. The variability among products could still be shown by totaling all the points on a scale with no weighting of individual items. Furthermore, weights were frequently assigned to specific items without the use of any qualifications or standards. This incorrect procedure of assigning weights defeated any purpose that weighting might serve in relation to increasing the reliability of a rating scale.

The second idea presented in the literature, as expressed by Wrightstone (25, p. 931), was that weighting is necessary for successful rating. He suggested that the importance of each item should establish its relative weight.

### Validity

The most meaningful suggestions found in the literature in regard to methods for increasing validity were made by Wrightstone and Barr, Davis, and Johnson. According to Wrightstone (25, p. 931), validity could be increased if the traits being rated were defined objectively.

Barr, Davis, and Johnson stated: ". . . It is difficult to formulate precise rules indicating the relative validity of any scales" (6, p. 109). Depending on the circumstances, a particular rating could be characterized by either high or low validity values. Validity is not even a constant factor for a rating scale when the same procedures are followed. Validity could also be expected to vary with the type of scale (6, p. 109).

### Reliability

Wrightstone, Remmers, and Gage made significant suggestions with respect to reliability. Wrightstone (25, p. 931) stated that reliability may often be increased by subdividing a trait into a number of sub-traits; however, he also stated that there is no one formula that will apply to all situations. Wrightstone (25, p. 929) further expressed belief that the reliability of descriptive phrases pertaining to specific traits could be increased in cases where those phrases were included which a judge is likely to have had opportunity to observe.

Remmers and Gage (17, p. 363) stated that reliability of final scores could be increased by averaging the ratings of several raters. Wrightstone (25, p. 931) also stressed this.

A further problem for the person attempting to develop a rating scale of high reliability is the number of ratings required of each judge on a single product. Furfey (11, p. 43) believed that increasing the number of judgments increases reliability; whereas, Hahn and MacLean said, "Two ratings by the same judge are no more valid than one" (13, p. 163).

### Selection and Training of Judges

Most authorities discussed the selection of judges because of their influence upon rating scales. Schwartz and Tiedman stressed the relationship between judges' skills and rating scales when they stated that rating scales are ". . . no better or worse than the skill of the

rater using them" (19, p. 166), and that ". . . most of the errors in rating are due not to the scale itself but to the raters using the scale" (19, p. 163).

Barr, Davis, and Johnson explained the way in which judges influence ratings when they said: "The rater's opinions, attitudes, and fund of general experience are all involved in the activity of rating," and ". . . his final conclusion may remain in a state of suspension not known even to himself, until the moment that he is required to make the rating" (6, p. 109).

Barr, Davis, and Johnson also emphasized the importance of the selection of judges. Since ". . . no amount of statistical manipulation can compensate for errors and inaccuracies which are especially likely to occur during the subjective phases of rating and rank order methods. . . ." (6, p. 109), successful rating depends upon selection of skilled judges.

#### Training of Judges

Authorities agreed that training of judges was necessary. Remmers and Gage stressed judges' training as one of the most crucial factors in using rating scales and stated that "training and instruction are essential for raters. . . if the various kinds of errors are to be minimized" (17, p. 363).

Schwartz and Tiedman made recommendations that would necessitate some type of training for judges when they stated that judges should be allowed to develop a belief in and thorough understanding of the rating scale and that they ". . . must become more objective in their observations and must guard against prejudice and agree on the meaning of

traits to be rated. . . " (19, p. 167). According to Wrightstone (25, p. 931), lack of training of judges is a cause for much ineffective rating.

#### Number of Judges

Most authorities agreed that more than one judge should appraise each product and that an average of these ratings should be computed in order to increase the reliability and validity of the ratings. Wrightstone, Justman, and Robbins said that using only one judge resulted in a weak evaluation and that "reliability of ratings may be increased by pooling the judgments of a number of persons. . ." (24, p. 168). Recommendations of other authors recorded by Wrightstone (25, p. 931) ranged from the pooling of no fewer than three judges' independent ratings to the pooling of the ratings of twenty-two judges. Schwartz and Tiedman (19, p. 169) advocated two, three, or more as the number of judges needed and an average of their ratings should be computed. Ahmann and Glock (2, p. 244) also suggested that several judges should evaluate independently and then compare scores.

Remmers and Gage said: "Since no single rating is ever perfectly reliable." . . . "steps should be taken to increase the reliability by averaging the ratings of several raters" (17, p. 363). They also presented the relationship between reliability and validity and number of judges and the limit to set on the number of judges.

. . . The reliability and validity of the ratings obtained from equally well-trained and instructed raters should increase as the number of raters increases. But usually the increase



not worth the trouble after about ten ratings (17, p. 363).

#### Errors of Judges

Schwartz and Tiedman (19, p. 166) stated that personal bias or prejudice of the judge affected the rating he made. This bias might cause him to overrate some parts of the product and minimize others as a result of his preconceived notions or beliefs. Evidence of personal bias affecting ratings could be illustrated by comparing the scores of several judges on the same item. These scores would usually differ by a marked degree. Three types of personal bias with suggestions for compensation or prevention were found in the literature reviewed.

One type of personal bias of which the judge must be reminded is the "halo effect". According to Arny (5, p. 224), the "halo effect" is a marked tendency for judges while scoring various items to be influenced by their self-impression of the person rather than the quality of the product. Arny (5, p. 224); Rothney (18, p. 16); and Barr, Davis, and Johnson (6, p. 82) all urged judges to resist this influence of the "halo effect".

Remmers and Gage (17, p. 365) discussed another type of personal bias termed the "logical error". They stated that raters' preconceptions of the relationships between certain parts of products causes this error. When the rater believed that two items were related, he would tend to give the same score to both items.

Remmers and Gage (17, p. 365) suggested that the "halo effect" and that the "logical error" could be avoided by scoring all products on the

same item before the next item was scored. Although Schwartz and Tiedman (19, p. 168) also advocated the above procedure, Mahler (16, p. 74) indicated that it was seldom practiced.

Mahler (16, p. 83) discussed another type of personal bias which is the judges' preference for favorable as opposed to unfavorable descriptive statements. Mahler suggested that compensation for this tendency to preferable phrases could be achieved by eliminating all descriptive statements that make a rating seem to appear unattractive or derogatory.

Another error of judges, termed the "generosity error" by Schwartz and Tiedman (19, p. 167) was rating too high. They said that ratings which were too high often resulted from the natural tendency of the judges to give the subject the benefit of the doubt. Symonds stated:

A rater may be helped to make more valid and reliable ratings if he is given some indication about the per cent of cases that should fall into various step intervals of the scale (21, p. 98).

Schwartz and Tiedman (19, p. 168) suggested that since this "generosity error" usually occurred when the judge was not certain about the definition, degree, or even existence of a specific quality, the error could perhaps be avoided by permitting or even encouraging the rater to indicate his competence or incompetence to appraise the specific item in question.

On the other hand, Schwartz and Tiedman (19, p. 168) stressed that some judges rated parts of the product lower than the parts deserved to be rated. Schwartz and Tiedman gave two explanations for this tendency to underrate parts of a product: (1) the judges, because of

feelings of inadequacy or insecurity, might rate part of a product as average rather than rate a deserving part as superior. (2) the judges might not be aware that extreme variations in range could and often do exist in the items that must be rated and that items can be rated very high as well as very low.

Several authors made suggestions that could compensate for, or perhaps even prevent, these errors made by judges. Wrightstone, Justman, and Robbins (24, p. 169) suggested that a set of instructions explaining specific details about the meaning of each quality or characteristic included on the scale should be issued to each judge. Both Ahmann and Glock (2, p. 225) and Arny (4, p. 165) emphasized the importance of practice for the judges. Ahmann and Glock (2, p. 225) stated that repetition was necessary for successful rating, and Arny expressed agreement when she said: "In reality, no matter how objective the rating device, people must have some directed experience in using it before their ratings will be very accurate" (4, p. 165). Wrightstone, Justman, and Robbins (24, p. 169) stressed that clear headings could eliminate much of the judges' confusion. It was suggested by Mahler (16, p. 74) that providing some additional space for comments on the rating scale would enable the judges to indicate their opinions more truthfully.

## CHAPTER III

### METHOD OF PROCEDURE

The purpose of this study was to develop, test, and revise a rating scale to be used for evaluating blouses constructed by first year students in high school home economics classes. These students were participating in the United States Office of Education project No. 5-1042. The rating scale was designed for use by juniors majoring in home economics education at the University of North Carolina at Greensboro. These juniors were to be experienced in the area of clothing construction. Students with high achievement in clothing were to be selected and given special training in the use of the rating scale.

#### Content of the Rating Scale

The first step in the development of the rating scale was to formulate a description of each construction detail of a blouse. The following were used in phrasing these descriptions: evaluation devices of instructors in the School of Home Economics at the University of North Carolina at Greensboro and of home economics teachers in North Carolina, authorities in the field of evaluation in home economics, and theses concerned with evaluation and rating scales. Additional descriptions were formulated by the researcher since some techniques and construction details specified in the self-instructional program were not included in the above devices. For example, a review of

available evaluation devices presented no descriptions of superior quality for tubular hems. These descriptions were developed by constructing a series of tubular hems according to directions in the self-instructional program and then describing the hem of superior quality.

Descriptions were placed on cards and when there was duplication, the cards were coded to refer to one another. For instance, in one rating scale, trimming a plain seam was considered a separate technique. In another, trimming was considered a step in the construction of a plain seam. These statements concerning trimming were recorded on both the cards referring to trimming and to plain seams. These cards were then coded to show this reference.

The next step was the development of a lengthy description of a top-level blouse<sup>2</sup> constructed on the basis of knowledge gained from the self-instructional program. Three or four phrases or statements, taken from the detailed descriptions on the coded cards, were used to describe the characteristics that were judged superior by the researcher. They were not yet worded for use in evaluating various degrees of workmanship. Any of the phrases and statements which needed further clarification or interpretation were foot-noted and explained in a supplement called "Instructions to the Judges".<sup>3</sup>

---

<sup>2</sup>See Appendix, pp.63-67.

<sup>3</sup>See Appendix, pp.68-70.



This supplement to the description of a top-level blouse contained both written statements and diagrams related to those phrases and statements which needed clarification. Staff members associated with the research project reacted to these statements and diagrams, appraising their clarity and usefulness. Revisions were made, using the suggestions of staff members.

The next step was to organize the description of the blouse and to divide it into sections such as staystitching and grainline. The selection of the heading for each section was influenced by the following reasoning: headings should indicate what is to be evaluated in that section, be as brief as possible, separate various major construction details, and separate certain techniques which could be evaluated more easily when listed independently.

The order or sequence of the separate sections in the top-level description was influenced by the following considerations: the sequence for construction details and techniques should be natural and not forced; it should coincide as nearly as possible with the sequence that judges would use if they were evaluating blouses without a guiding reference; in addition, the sequence should follow the same orderly pattern and not cause the judges to move at random from one section of the blouse to another unless there was a specific reason for doing this. If there were one technique that could more easily be rated in all areas of a blouse at one time rather than rating it throughout the device at each point where it occurred, the sequence should make provisions for this. When results from tryouts of the device indicated a need for change in sequence,

revisions were made at various stages in the study.

Objectives used in the writing of the self-instructional program were reviewed as an aid to statement of pertinent descriptions of the top-level blouse. Descriptions were then altered and items deleted so that only descriptive phrases and statements were used which pertained to specifications of the blouse used in the experiment. Altering and deleting descriptive phrases and statements was a continuous process.

Three members of the research staff working on the self-instructional program appraised the top-level descriptions. Their suggestions prompted another revision.

The evaluation consultant for the research project (23), a psychologist from the University of North Carolina at Chapel Hill, suggested that twenty to twenty-five blouses would provide a sufficient range of quality for initial testing of the rating scale. Since testing of the rating scale would necessitate a supply of blouses to be rated, provision for obtaining blouses was made during the initial stages of the study. Home Economics I students in Forsyth and Stokes counties were selected to construct the blouses as home projects with no help from their teachers.

It was believed that a blouse completed as a home project with no teacher assistance and that followed a clothing unit in the classroom would be similar to blouses made the following year in the field study. Blouse specifications<sup>4</sup>, including construction details and requirements for design and fabric, and a list of suitable pattern numbers were

---

<sup>4</sup>See Appendix, p.71.

mailed to the students. The students were asked to lend their blouses for a period of three months with the promise that as compensation for their co-operation, buttonholes would be machine-worked in their blouses. In addition, the students were told that the identity of the blouses would not be revealed. The judges did not know which student had made each blouse or from which school the blouse had come.

When some students failed to send blouses, it was decided that members of the research staff would construct the additional blouses needed. In constructing these blouses, the staff made particular effort to make blouses of quality varying from low to top-level and to alter techniques somewhat. This was done to determine whether or not rating scale scores would vary in relation to the actual range in qualities of the blouses.

The top-level description was checked again when blouses began arriving. Even if it were difficult to see certain construction details or results of techniques as described on the scale, these phrases and statements remained in the top-level description at this time. Final decisions concerning deletions, additions, and revisions were made after the final test-run of the rating scale.

After the first trial run, the description of a top-level blouse was sent to the clothing consultant for the research project for the purpose of checking the phrases and statements. A conference was held with the clothing consultant and project director to discuss the descriptions and final revision of the description incorporated their suggestions. The description was set aside to be used when the rating scale

was constructed. During the time in which the top-level description was being developed, evaluated, and revised, a corresponding low-level description was being developed following the same procedure.

The phrases and statements from both the top-level and low-level descriptions were phrased so that objective rather than subjective judgments could be made. A subjective statement such as "Staystitching is neat" would be made objective by rewording it as follows: "Staystitching is within one-eighth of an inch from the marked seamline toward the cut edge." All statements and phrases were analyzed to insure that they were worded in as objective a manner as possible. Instructions for judging were developed with directions for measuring certain construction details. Analysis and revision of each statement continued throughout the study.

#### Form for the Rating Scale

Since content for the rating scale had now been selected, the next problem was the choice of form. The evaluation consultant suggested a dichotomous form. The phrases and statements from the top and low-level descriptions were organized into a variation of a dichotomous form.<sup>5</sup> Instead of only "yes" and "no" response columns, a column entitled "I do not see this" was added. Statements and phrases were reworded so that affirmative and negative responses could be given.

Illustrations were developed to accompany many of the phrases and statements. These illustrations were checked by staff members for

---

<sup>5</sup>See Appendix, p.72.

clarity and revisions were made. It was also necessary to make provisions in the directions for suggesting the deletion of items pertaining to construction details that either were not present or could not be seen in a particular blouse.

When trying out the first dichotomous scale, the research staff members were given a set of instructions on which they were asked to indicate directions that were not clear, items that seemed repetitious, phrases that were difficult to interpret, as well as their opinions in regard to the feasibility of the amount of time required to rate items. Their written comments along with individual conferences prompted further revisions in the content and form of this first dichotomous scale.

Research staff members were given two forms for recording appraisals of each item on the scale: (1) a separate sheet comparable to an answer sheet, and (2) an attached numbered strip. The staff mutually indicated preference for the latter.

A second dichotomous scale was then developed and either affirmative or negative answers were the only responses required of a judge using the scale. This second dichotomous scale was tried out by the staff members; individual conferences allowed them to express their opinions concerning the form of the scale. Generally, all three staff members believed that the dichotomous form failed to meet the needs of the evaluation device required by the experiment. Such a scale did not allow for adequate scoring of many parts of the blouse because a "yes" or "no" response was not sufficient to indicate the range of quality of



workmanship.

The decision was made to use a deviation of the graphic type of rating scale in order to give judges the opportunity to appraise each construction process by considering several levels of quality rather than the two extremes. The following criteria were used to select statements describing each level of quality: (1) The top-level category, as far as possible, was to be set so high that the category would apply to only 20 or 25 per cent of the blouses scored; at least half of the blouses were to fall at the middle-level; 20 to 25 per cent of the blouses in the low-level. (2) In preparation for setting the limits for each of the three levels, various parts of the blouses were measured or observed closely. For example, seam widths on each of twenty-three blouses were measured. The difference between each measurement and a corresponding measurement on a top-level blouse was used to express a range of errors. This range of errors influenced the specificity of the three levels on the rating scale. Measurements descriptive of the seams of highest quality were used in the specification describing the top-level seams. Measurements of seams of poorest quality were used in the specification describing the lowest level. Between these two, estimates of average quality were used in statements describing the middle-level.

Additional changes were made in the three-level scale as it was being developed. Provision was made for judges to indicate when particular construction techniques were omitted. Changes they recommended in directions and diagrams were also incorporated. Some items were eliminated that were very difficult to score or that proved too time

consuming for the results obtained.

Descriptions of the three levels were recorded horizontally. A range of one to six points was to be used for each item. If the quality of a blouse part was described by the limits within the top-level, then the blouse part could receive a score of six or five, depending upon its quality. The score would then be recorded on the detachable score sheet located at the far right of the scale.

This three-level scale was now completed and ready for use<sup>6</sup>. Scoring of the blouses by each of the five judges was recommended by the measurement consultant as basis for statistical appraisal of the device.

#### Selection and Training of Judges

Judges were carefully screened and interviewed before they were allowed to participate in the project. Four juniors in home economics education and one graduate student in clothing were selected. Since four judges would be used in the field experiment the following year, juniors, rather than seniors in home economics education were chosen because it was hoped that these same juniors would evaluate the blouses from the field experiment. Criteria necessary for selection were that the students have an over-all average of "B" or better and that they have some previous experience in evaluating clothing. It was not considered necessary for selection that the judges be technical experts in the fields of evaluation or clothing since one of the secondary objectives of this study was to construct a rating scale such that its

---

<sup>6</sup>See Appendix, p.73.

use would not be limited to highly qualified personnel. Any person capable of following directions and of measuring accurately when given specific instructions should be able to use the rating scale.

It was believed, however, that any judge, no matter how well qualified, needed an introductory training session to allow her to become familiar with the scale. The judge should have an opportunity to voice her misunderstandings and prejudices or biases. The person responsible for the development and use of the device could then learn where the device needed improvement or where the judge needed assistance in interpreting it.

Three training sessions of two hours each provided opportunity for the discussion of any difficulties encountered by the judges. Each judge had two copies of the device, one of which was a personal study copy. The judges were requested to review this device to aid in understanding of and familiarity with its parts. These training sessions were taped in order that the tape could be reviewed for future training sessions. The judges scored blouses in these training sessions and thus had first hand experience with the rating scale. In addition, the judges were also warned about bias and urged to rate as objectively as possible. The judges were timed on each of the nine sections of the device and if any section were extremely time consuming, the importance of the information gained from this section was evaluated in terms of the time required to score this section.

While the blouses were being scored, changes were made within some sections as a result of the time required to complete them, the

quality and reliability of information gained, and the judges' difficulties with these sections. The device was then set up for use by the five judges. Twenty-one blouses were to be scored by each judge. The rating scale consisted of nine sections and a total of 128 items. Answer strips for recording of the scores were attached to a master copy of the scale for each judge.

#### Scoring of Blouses

The judges were required to score the twenty-one blouses over a period of one month. This scoring was done between eight and five o'clock Monday through Saturday in the educational research laboratory in the School of Home Economics at the University of North Carolina at Greensboro. Each judge indicated completion of the scoring of a blouse by signing her number to a master sheet listing the blouses. The judges did not have to follow a sequence in the scoring of the blouses. The researcher was available at almost all times when the judges were scoring to answer any questions that might arise. The judges timed themselves when scoring each section; these time scores were attached to the tally sheets and later analyzed.

Compensation was made to the judges for their time by a payment of \$.85 an hour. Funds from the project made this payment possible. All judges completed scoring the twenty-one blouses within the one month period. Since it was hoped that these same judges could be used the following year, it was beneficial that they learn to budget their time and be able to score blouses within a given time limit.

## CHAPTER IV

### FINDINGS

The rating scale itself is the major finding of this study. Descriptive data about the scale are reported in this chapter. Two forms of the scale were developed--one was dichotomous and one provided for responses on a six point scale numbered from six to one. Procedure for developing the dichotomous scale appeared in the Method of Procedure. Since it became evident early in the study that the six point scale was superior to the dichotomous, only the six point scale is discussed in this chapter.

#### Description of the Scale

The scale covered each detail of blouse construction as these were listed in the description of a top-level blouse. The scale was organized into nine sections: (1) grainlines, (2) staystitchings, (3) plain seams, (4) facings, (5) darts, (6) sleeves, (7) sleeve hems, (8) blouse hems, and (9) general appearance.

Descriptive statements at three levels composed the items within each section. The first statement described a blouse of superior quality, the second a blouse of average quality, the third an inferior blouse with respect to the particular characteristic rated in this statement. Each of the three levels was subdivided on the six point scale with a result that a maximum of six points rather than three could be given to each item rated. The rater was to record four points



for a blouse slightly above average, three points for a blouse slightly below average and to divide the superior and inferior levels likewise into two categories each. A score of zero was recorded only if the student omitted one of the required steps.

The total number of items in the nine sections was seventy-six. More than seventy-six scores were recorded however. Many items had more than one part to be scored, and a particular item could be worth up to thirty-six points. The total number of scores recorded was 128.

Since there were 128 items, each of which could receive six points if the items being scored were top-level, it might be expected that the maximum score would be 768. Six items were, however, weighted times two because they were believed to be most important. These items made a greater influence on the appearance of the finished blouse. For example, items three A and three B,<sup>7</sup> found in the appendix on page seventy-four, were weighted times two. The cross-grain of the sleeve was considered most important because direction of grain would greatly affect the overall appearance of the blouse. The maximum score on the device was 804.

The 804 points on the rating scale were distributed among sections as shown in Table 1.

#### Procedure for Training Judges

The use of the scale is discussed in the report of the training session for the judges. The researcher decided that a training session in which the judges would have an opportunity to use, question, and dis-

---

<sup>7</sup>See p. 69, where cross-grain is shown.

cuss the scale was necessary since authorities in the area of rating scales had emphasized the importance of proper training of judges.

TABLE 1

## SUB-SCORES ON THE RATING SCALE

Sections	Scoring Points
I . . . . .	36
II . . . . .	24
III . . . . .	234
IV . . . . .	114
V . . . . .	102
VI . . . . .	102
VII . . . . .	42
VIII . . . . .	90
IX . . . . .	60
TOTAL	804

At the start of the training session, the researcher presented a short introduction on rating scales and discussed the qualities characteristic of a reliable judge. The judges were informed that the main purpose of the training session was to give each of them an opportunity to use the rating scale and the "Instructions to the Judges" as they rated a blouse during the training session and used the required equipment for measuring certain details on the blouse. Each item was rated separately. If a diagram or instructions for measuring accompanied the item statement, the specified procedure was followed.

The judge recorded a score for the item in the box on the far right on the tally sheet for this blouse. After completing the rating of a blouse, the judge detached all tally sheets for this blouse and stapled them together. After each judge had rated one blouse, it was suggested that several items be removed from the scale because the judges found these items difficult or impossible to rate.

After scoring the blouse during a trial session, each judge then scored the twenty-one blouses. Data recorded by the judges in this trial session were not used in the analysis.

#### Increasing the Efficiency of the Scale

Since scoring of the blouses required about one month, the researcher did not want to wait until all the scores were recorded before evaluating the reliability of the items on the scale. Each blouse was scored and individual records were kept for each of the seventy-six items on which scores by the five judges were recorded. The decision was made to improve the rating scale while the judges were rating the blouses, rather than after they had judged all the blouses.

One indication that an item needed revision was evidenced by disagreement among judges. Some disagreement was anticipated, but when points assigned to an item by the five judges varied by two or more out of the possible six points, there was need for revision. An example of such an item is shown in Table 2 . Item number nine<sup>8</sup> is concerned with the width of the seam allowances for the shoulder, side, and armseye

---

<sup>8</sup>See Appendix, p:77.

TABLE 2  
 SCORES, ITEM NINE, WIDTH OF SEAM ALLOWANCES

Blouse Number	SHOULDER					SIDE					ARMSEYE				
	Judges					Judges					Judges				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
1	4	4	4	4	2	3	4	2	3	2	4	3	4	4	4
2	0	0	0	0	0	0	0	0	0	0	3	3	3	3	2
3	0	0	0	4	0	0	0	0	2	0	3	4	2	2	2
4	5	4	5	4	5	4	5	4	4	5	4	4	5	5	4
5	4	5	4	4	2	4	4	4	4	4	3	4	3	2	3
6	4	4	4	4	3	4	5	5	4	5	2	4	2	4	2
7	5	5	4	4	3	4	4	2	4	2	4	4	4	4	3
8	1	1	1	2	1	1	3	3	2	1	2	4	4	4	1
9	4	4	4	4	4	2	4	2	2	1	4	4	4	4	2
10	4	5	4	4	4	0	5	5	5	5	4	4	3	2	3
11	4	6	5	2	5	3	3	3	2	3	4	3	2	2	2
12	5	5	4	5	4	3	3	3	4	3	4	5	4	4	3
13	3	4	4	4	3	3	3	3	4	4	4	4	3	4	3
14	4	4	5	4	4	5	5	5	4	4	4	4	4	5	4
15	5	4	4	4	5	5	4	5	5	5	4	5	4	3	5
16	5	6	4	5	6	2	4	2	4	1	3	4	2	4	1
17	5	4	5	4	3	4	4	5	3	4	4	4	5	5	5
18	4	4	3	2	2	2	3	2	2	1	2	3	3	3	3
19	0	4	4	5	0	0	4	4	4	0	0	0	0	5	0
20	5	5	4	4	1	4	4	1	2	1	4	4	4	3	4
21	3	4	4	2	4	3	3	4	2	2	0	3	0	1	1

seams.

Shoulder seams on blouse nineteen were scored zero by two judges, four by two judges, and five by one judge. Similarly, on blouse eleven, shoulder seams were scored six by one judge, five by two judges, four by one judge, and two by another judge. There were too many instances of disparity among judges on item nine. The procedure described above was used to analyze all of the items in the rating scale.

All items on the scale were informally analyzed by pairing the judges and counting the number of times they were in complete agreement about the score assigned. An example of these data for item nine is recorded in Table 3. The percentage of agreement between judges was below fifty in most instances, further evidence that item nine needed revision.

Item nine was reviewed and discussed with the judges. The general agreement was: (1) that all judges were using the same method for measuring the seam allowances, (2) they understood the method, (3) they used the measuring instruments in the same ways, yet, (4) human error entered the measuring process. The failure of the judges to agree on the scores and the time required for this particular scoring prompted the researcher to suggest that the item be removed from the scale.

All items on the rating scale were analyzed in the manner described above. Examples of items on which judges had difficulty in agreeing are numbered one, two A, two B, eleven A, eleven B, eleven C, thirteen A, thirteen B, twenty-two A, forty-eight, fifty-three, seventy-five, and these are found on pages seventy-three through 105 of the appendix. Items that showed disparity among the judges were either



revised or suggestions for elimination of these items from the rating scale were made.

TABLE 3

NUMBER AND PERCENTAGE OF BLOUSES FOR WHICH PAIRS OF JUDGES AGREED ON SCORING OF SHOULDER, SIDE, AND ARMSEYE SEAMS (ITEM NINE)

PAIRS OF JUDGES	SHOULDER		SIDE		ARMSEYE	
	N*	%	N	%	N	%
I & II	11	52.4	11	52.4	10	47.6
I & III	10	47.6	12	57.1	12	57.1
I & IV	9	42.9	9	42.9	7	33.3
I & V	10	47.6	8	38.1	7	33.3
II & III	10	47.6	11	52.4	8	38.1
II & IV	10	47.6	6	28.6	8	38.1
II & V	7	33.3	9	42.9	6	28.6
III & IV	10	47.6	8	38.1	11	52.4
III & V	9	42.9	10	47.6	12	57.1
IV & V	5	23.8	6	28.6	6	28.6

\*There were twenty-one blouses.

Another example of an item for which deletion from the rating scale was suggested is shown in Table 4. Item seventy-two<sup>9</sup> is con-

<sup>9</sup>See Appendix, p.105.



cerned with the scorching and shininess that results from over-pressing a blouse. The numbers indicate the scores given this particular item by the five judges. It became evident early in the testing of the scale that almost every blouse was receiving a score of six on this item. The only blouses rated less than six were numbered one and three. Blouse one was deliberately scorched and made to be shiny so that the effectiveness of the rating scale in differentiating among blouses with respect to this characteristic could be tested. The only blouse made by a student that scored less than six on this item was numbered three; furthermore, only one judge, number IV, rated this item less than six. This item had no effect on the scale other than increasing the time required for rating; therefore, suggestion was made for its deletion. Other items where deletion was suggested because the items received the same score regardless of the blouse being rated were numbered sixty-eight, sixty-nine, seventy, seventy-one and seventy-three.<sup>10</sup>

Scoring of some processes on only one side of the garment, rather than on both right and left sides would be desirable from the standpoint of saving time in cases where quality of construction could be consistent from one side to the other. Item two<sup>11</sup>, in which the grain-lines for the right and left lengthwise front edges of the blouse are scored, is an example of such an item. Scores given this item by the

---

<sup>10</sup>See Appendix, pp.104-05.

<sup>11</sup>See Appendix, p.73

TABLE 5  
 SCORES ON ITEM 2 FOR RIGHT AND LEFT SIDES OF GARMENT

BLOUSE NUMBER	JUDGE NUMBER									
	I		II		III		IV		V	
	RIGHT	LEFT	RIGHT	LEFT	RIGHT	LEFT	RIGHT	LEFT	RIGHT	LEFT
1	5	4	4	4	4	4	4	4	4	4
2	4	3	4	4	4	4	4	4	3	3
3	2	2	2	3	2	2	2	2	1	1
4	5	5	4	4	5	4	4	4	4	5
5	4	4	6	5	4	4	6	4	4	5
6	6	5	6	5	6	5	6	5	4	5
7	5	4	4	5	5	5	6	4	5	5
8	4	4	5	4	5	4	5	5	3	3
9	5	4	4	4	4	4	4	4	4	4
10	4	5	6	5	4	5	5	5	3	3
11	2	2	4	4	4	4	1	2	3	1
12	4	5	5	5	5	5	5	5	1	1
13	4	4	4	5	5	4	4	4	4	5
14	4	4	5	5	4	5	5	4	5	5
15	5	5	5	5	5	5	5	5	4	4
16	4	5	5	4	4	4	4	4	4	4
17	5	5	5	5	5	5	5	5	4	4
18	5	4	6	4	5	5	5	4	5	4
19	5	5	6	6	5	4	3	4	4	4
20	4	4	4	5	4	4	4	4	4	3
21	4	4	5	5	6	4	4	6	4	4

five judges are indicated in Table 5. It may be observed that the scores on the right and left sides are usually the same. Perfect consistency would result in each judge scoring the right and left sides the same on each of the twenty-one blouses.

The number of times a judge scored items two A and two B the same for both right and left sides is recorded in Table 6. Since over half of the pairs of scores are identical and since all except one of the remaining pairs differ by no more than one point, the researcher suggested that only one side of the garment needed to be evaluated. This suggestion was also made for items numbered forty-seven A, forty-seven B, forty-seven C, forty-nine A, and forty-nine B found on pages ninety-seven and ninety-nine of the appendix.

TABLE 6

NUMBER OF TIMES JUDGES SCORED ITEM TWO A AND TWO B  
THE SAME FOR RIGHT AND LEFT SIDES

JUDGE NUMBER	NUMBER OF IDENTICAL FOR RIGHT AND LEFT
I . . . . .	12*
II . . . . .	12
III . . . . .	13
IV . . . . .	13
V . . . . .	14

\*Highest possible number is 21.



### Analysis of Judges' Scoring Time

The judges were instructed to record the time required for rating each blouse. The average time used by the five judges to rate one blouse in the training session was 381 minutes. Each judge used progressively less time for successive blouses until she reached a rather uniform rate. The mean time for the five judges rating the twenty-one blouses was 100 minutes.

There was considerable variation among blouses in scoring time. Since the blouses were not scored in the order in which they were numbered, the variation in time among judges for any particular blouse resulted from previous scoring experience, as well as from the varying pace of scoring among judges after each had reached her own plateau.

Figure 1 shows the time used by the five judges to score every other blouse. The first blouse scored required a great deal of time. For each judge, the third blouse scored required about one-third as much time as the first. Time for scoring successive blouses demonstrates that each judge reached her own plateau by the time she had scored five blouses. After the fifth blouse was scored, there were small deviations in scoring time.

The judges were also instructed to record the time required for each of the nine sections of the scale. These time records were then studied and if a particular section required too much time, it was reviewed to see if the time necessary for rating was worth the value gained from the rating.

Table 7 shows the time used by the five judges to score separate

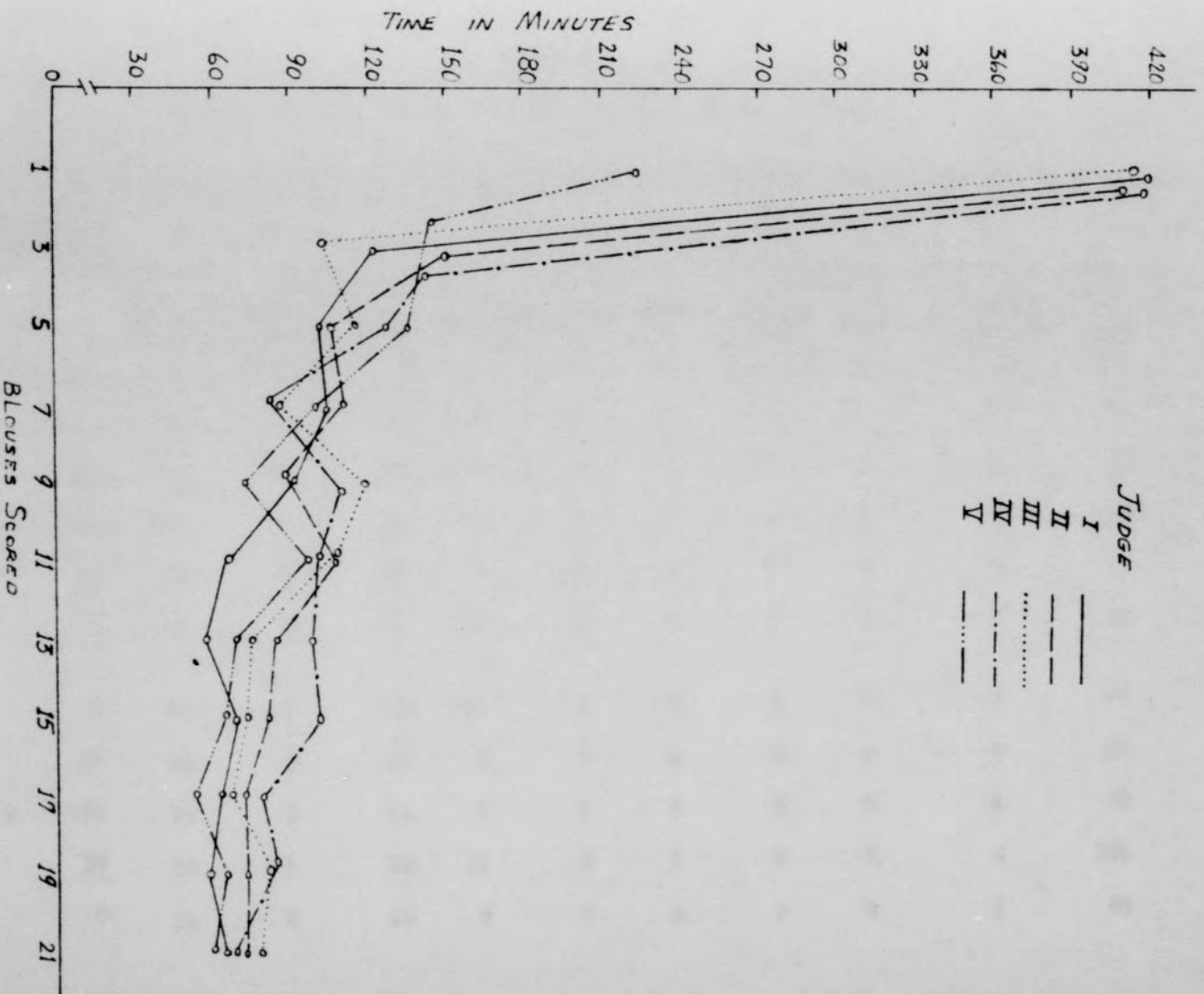


FIGURE 1 Time Used By Five Judges To Score Every Other Blouse

TABLE 7

MINUTES USED BY JUDGES TO SCORE SECTIONS OF SIX BLOUSES

BLOUSE NUMBER	JUDGE NUMBER	I	II	III	IV	V	VI	VII	VIII	IX	TOTAL TIME
		GRAIN- LINE	STAY- STITCHING	PLAIN SEAMS	NECKLINE FACING	DARTS	SLEEVE HEM	SLEEVE HEM	BLOUSE HEM	GENERAL APPEARANCE	
1	I	30	2	25	5	4	4	2	4	4	80
	II	15	3	30	5	4	7	3	5	5	78
	III	18	1	19	5	4	5	4	5	3	64
	IV	35	6	28	6	8	6	5	8	6	92
	V	12	3	21	10	5	3	3	5	4	66
2	I	28	2	19	15	5	5	4	3	2	83
	II	15	5	29	9	7	4	4	4	4	80
	III	15	3	21	7	5	4	5	5	4	69
	IV	30	3	28	11	7	5	8	8	4	106
	V	15	6	26	8	8	4	7	8	3	85

TABLE 7--Continued

BLOUSE JUDGE NUMBER NUMBER	I	II	III	IV	V	VI	VII	VIII	IX	TOTAL	
	GRAIN- LINE	STAY- STITCHING	PLAIN SEAMS	NECKLINE FACING	DARTS	SLEEVE	SLEEVE HEM	BLOUSE HEM	GENERAL APPEARANCE	TIME	
	I	12	2	18	5	1	8	2	6	2	56
	II	10	2	20	5	1	5	5	3	5	56
3	III	12	2	20	6	3	2	3	5	4	57
	IV	35	5	35	12	3	7	5	11	5	118
	V	10	3	18	8	3	2	4	3	3	54
	I	21	3	25	6	4	7	3	4	2	75
	II	19	8	35	10	10	8	5	10	6	111
7	III	17	3	30	5	10	6	4	5	3	83
	IV	14	3	28	5	10	10	8	5	10	93
	V	10	3	28	6	5	5	5	5	5	70

TABLE 7--Continued

BLOUSE JUDGE NUMBER NUMBER	I	II	III	IV	V	VI	VII	VIII	IX	TOTAL TIME	
	GRAIN- LINE	STAY- STITCHING	PLAIN SEAMS	NECKLINE FACING	DARTS	SLEEVE	SLEEVE HEM	BLOUSE HEM	GENERAL APPEARANCE		
	I	31	1	48	6	10	7	5	8	4	120
	II	30	5	27	11	9	10	8	5	3	108
11	III	17	1	36	7	10	7	4	8	4	94
	IV	25	5	37	10	13	7	4	7	3	111
	V	24	12	25	8	10	7	7	6	3	102
	I	31	1	40	8	10	6	5	5	3	109
	II	20	10	45	10	10	10	5	15	10	135
21	III	17	2	35	12	8	8	5	10	7	104
	IV	35	5	35	10	5	5	7	20	10	132
	V	21	4	44	19	12	7	7	9	5	128



sections of eight of the twenty-one blouses. Similar time records were kept for the scoring of all of the blouses.

Section III, plain seams, required the greatest amount of scoring time, an average of twenty-eight minutes. This was approximately one-fourth of the total time. Two hundred and thirty-four or 30 per cent of the total 804 points on the scale came from Section III. Approximately one-fourth of the scoring time was spent determining 30 per cent of the total score. Therefore, the scoring of Section III required a proportionate amount of time for the information gained. It was recommended that the major part of this Section III be included in the rating scale.

Section I, grainline, was second in amount of time required for scoring. Average scoring time for Section I was twenty-two and three-tenths minutes. This was approximately one-fourth of the total time. Thirty-six or 4 per cent of the total 804 points on the scale came from Section I. Approximately one-fourth of the scoring time was spent determining 4 per cent of the total score. Thus, the scoring of Section I required a disproportionate amount of time for the information gained. For this reason, it was recommended that the method for rating grainline be altered.

#### Blouse Scores

A mean score for each blouse was computed. These means ranged from 538.0 to 716.0, the average being 641.3. The wider the ranges of scores in a distribution, the better a scoring device differentiates

between blouses with respect to quality of workmanship. This range of 177.8 points indicates that the device clearly distinguishes among blouses of varying quality.

Careful study of the data indicated that the range would be equally good if those items were deleted which were scored similarly by the five judges and which, therefore, failed to differentiate among blouses of varying quality. It was also suggested that those items showing disagreement by two or more points between judges and those items requiring a disproportionate amount of time for scoring as compared with the information gained be deleted.

Scores assigned to the twenty-one blouses by pairs of judges were correlated and these correlations are recorded in Table 9. The correlations range from .65 to .89. Scores of Judge II correlated lower with scores of the other judges than did scores of each of the other judges with each other.

The correlations were transformed to z scores so that an average correlation could be computed. The z scores were averaged and the resulting score was converted back to a correlation coefficient. This above method was used because the consultant on measurement aspects of project No. 5-1042 recommended it. According to Adkins<sup>12</sup>,

Translation of r's to z's is also recommended if an average of some r's is sought, because the units in which degree of correlation are expressed are not rewarded as equal in different parts of the range from 0 to  $\pm 1$ . The procedure is to trans-

---

<sup>12</sup>Dorothy C. Adkins, Statistics (Columbus, Ohio: Charles E. Merrill Books, Inc., 1964), p. 275.

late each  $r$  to the corresponding  $z_r$ , get the arithmetic mean of the  $z_r$ 's and then find the  $r$  that corresponds to the mean  $Z_r$ .

The average correlation among judges in this study was .818.

This was considered satisfactory. The consultant on measurement problems had indicated in a conference with the researcher that a correlation coefficient of .80 or higher would give reliable ratings for purposes of this study.

TABLE 8  
CORRELATIONS BETWEEN SCORES OF PAIRS OF JUDGES

JUDGE NUMBER	I	II	III	IV	V
I		.65	.89	.89	.85
II			.67	.80	.78
III				.84	.81
IV					.87
V					

## CHAPTER V

### SUMMARY

Students were required to construct a blouse as participants in the field experiment of United States Office of Education research project No. 5-1042. The students followed a self-instructional program that had been written as a part of this research project and that guided them through the process of constructing this simple blouse.

The purpose of this study was to construct, test, and revise an evaluation device for scoring quantitatively the quality of workmanship on these blouses made by the students. Various kinds of evaluation devices were studied. Since rating scales are most frequently used in areas where measurement is dependent upon observation, a rating scale was selected as the appropriate device for this quantitative scoring of the blouses.

The first step in the development of the evaluation device was to describe a hypothetical blouse of high quality made by following the self-instructional program. A description of each construction detail of this blouse was formulated by the author and reviewed by specialists in the area of clothing. These detailed descriptions were organized into a top-level description which was to become a part of the evaluation device. This top-level description was then divided into sections and these in turn were sequenced.

Alterations of and deletions from this description preceded



appraisal of the description by staff members of the research project. Revisions continued until it was considered acceptable to the researcher and director of the study.

Blouses secured from Home Economics I students in Forsyth and Stokes counties and from staff members were used for initial testing of the rating scale. Five judges obtained to score each of these twenty-one blouses were carefully selected.

At the same time that this description of the top-level blouse was being developed, evaluated, and revised, a corresponding description of a blouse of inferior quality was being developed by the same procedure. The phrases and statements from both the top-level and low-level descriptions were then placed in a form that could only be used by an observer in the process of objectively rating a blouse. The trial form of the rating scale was accompanied by a supplement titled "Instructions to the Judges". This supplement contained both diagrams of and explanations for any phrases or statements that needed clarification.

A dichotomous form, on which the judge recorded that the blouses were satisfactory or unsatisfactory with respect to each descriptive phrase, was chosen for the rating scale. Later, the dichotomous form was discarded in favor of a three-level scale on which judges could appraise each construction process by considering several levels of quality. Judges believed they could score more accurately when using such a scale. The three-level scale was a deviation of a graphic form of rating scale.

Statements from the description of the top-level blouse formed

the basis for the upper level of the three-level scale and statements from the description of the low-level blouse formed the basis for statements at the lowest level. Statements describing a blouse of average quality were added for the middle-level. Statements were reworded to include measurements characteristic of blouses at each level; for example, the width of seam allowances pressed open at the top-level was to be one and one-fourth inch or miss this by no more than two threads. The limits for measurements in these statements were set by measuring and observing blouses made under project specification.

Additional changes were made in the three-level scale as it was being developed. Each of the three categories or levels was subdivided to form a six point scale. If the quality of a blouse item were described by the limits within the top-level, then the blouse item could receive a score of six or five, depending upon its quality. Scores were recorded on a detachable score sheet located at the far right of the scale.

Scoring of the blouses by each of the five judges was recommended by the measurement consultant as the basis for statistical appraisal of the device. These five judges were carefully screened and interviewed before participating in the project. Four of them were juniors in home economics education and one was a graduate student in clothing. The average time used by the five judges to rate one blouse in this training session was 381 minutes or six hours and twenty-one minutes.

The rating scale as it was then set up for use, was organized into nine sections: (1) grainlines, (2) staystitching, (3) plain seams,

(4) facings, (5) darts, (6) sleeves, (7) sleeve hems, (8) blouse hems, and (9) general appearance. These sections included seventy-six items with a maximum score of 804. Answer strips for recording of the scores were attached to a master copy.

The time analysis of the scale resulted in the following findings: (1) The range in mean scoring time for the twenty-one blouses was from 67.2 to 147.3 minutes, the over-all mean being 100.0 minutes. (2) One section, grainlines, required a disproportionate amount of time for the information gained; this section was reviewed and revised.

A mean score was computed for each blouse. These means ranged from 538.0 to 716.0, the average being 641.3. This range of 178.0 points indicated that the device distinguished among blouses of varying quality.

The measure of reliability for the rating scale was based on inter-correlations among pairs of judges--each of whom scored twenty-one blouses in a period of one month. These correlations ranged from .65 to .89, the average being .82.

Since the development of an objective rating scale with high reliability was one of the desired outcomes of this study, all of the items on the scale were analyzed for disagreement among judges. Suggestions were made that the following items be revised or eliminated from the scale since the data indicated that the judges varied by two or more points in their assignment of score values for these particular items on the same blouse: one, two A, two B, nine, ten, eleven A, eleven B, eleven C, thirteen A, thirteen B, twenty-two A, forty-eight, fifty-three, and seventy-five.

Analysis of data also indicated that some items would have no

effect on the scale other than increasing the time required for rating since all blouses were scored the same or very nearly the same on these items. Usually these items were descriptions of steps performed correctly by all students. Items described above and numbered sixty-eight, sixty-nine, seventy, seventy-one, seventy-two, and seventy-three were deleted from the rating scale.

It was suggested for items two A, two B, forty-seven A, forty-seven B, forty-nine A, and forty-nine B where the quality of construction was consistent from one side of the blouse to the other that only one side of the blouse be scored.

The following suggestions were made by committee members to improve the rating scale: (1) Diagrams of the blouse should include shoulder darts. (2) Item two could be stated more accurately if worded cross-grain of sleeve is perpendicular to foldline of sleeve--rather than is perpendicular to the underarm seam.

Although the reliability for the rating scale was considered satisfactory, it was believed the reliability could be improved if some of the descriptive statements for the scale had been worded more briefly and, in a few instances, more objectively. The scale still requires occasional subjective ratings that result in disagreement among judges' ratings.

Since the rating scale was developed for use in the research project, the scale is too lengthy and detailed to be used in its present form by high school home economics teachers. The researcher hopes the rating scale through its emphasis on clear instructions, accurate visual aids, and definite categories or limits for varying levels of quality of a product can serve as a guide to these teachers when they develop their own evaluation devices.

BIBLIOGRAPHY



## BIBLIOGRAPHY

1. Adkins, Dorothy C. Statistics. Columbus, Ohio: Charles E. Merrill Books, Inc., 1964.
2. Ahmann, J. Stanley, and Glock, Marvin D. Evaluating Pupil Growth. Boston: Allyn and Bacon, Inc., 1963.
3. An Evaluation of Modern Education. Edited by John Paul Leonard and Alvin C. Eurich, with cooperation of Reginald C. Bell. A report sponsored by the Society for Curriculum Study, New York: D. Appleton-Century Co., Inc., 1942.
4. Arny, Clara Maude (Brown). Evaluation and Investigation in Home Economics. New York: F. S. Crofts and Co., 1941.
5. \_\_\_\_\_. Evaluation in Home Economics. New York: Appleton-Century Crofts, 1953.
6. Barr, Arvil Sylvester, Davis, Robert A., and Johnson, Palmer O. Educational Research and Appraisal. New York: J. B. Lippincott Co., 1955.
7. Blackwell, Sara. "Utilizing Different Kinds of Evaluating Instruments," Illinois Teacher, Vol. V, No. 2 (1955), 87-93.
8. Bradfield, James M., and Moredock, H. Stewart. Measurement and Evaluation in Education. New York: The MacMillan Co., 1957.
9. Fleck, Henrietta. "Evaluation in Practice," Journal of Home Economics, Vol. 48 (May, 1954), 300-02.
10. \_\_\_\_\_. "How to Evaluate Students," (Bulletin), Bloomington, Illinois: McKnight and McKnight Publishing Co., 1953.
11. Furfey, P. H. "An Improved Rating Scale Technique," Journal of Educational Psychology, Vol. 17, No. 1 (January, 1926), 45-48.
12. Gerberich, Joseph Raymond, Greene, Harry Andrew, and Jorgensen, Albert N. Measurement and Evaluation in the Secondary Schools. New York: Longmans-Green Co., 1943.

13. Hahn, Milton E., and MacLean, Malcom S. General Clinical Counseling in Educational Institutions. 1st. ed. New York: McGraw-Hill, 1950.
14. Huntzicker, Jane. "The Effectiveness of Selected Methods of Clothing Construction Used in Teaching First Year Students at Yakima Senior High School." Published Master's thesis, Oregon State College, June, 1957.
15. Latchaw, Marjorie, and Brown, Camille. The Evaluation Process in Education, Physical Education, and Recreation. New York: Prentice-Hall, 1962.
16. Mahler, Walter P. "Some Common Errors in Employee Rating Practices," Personnel Journal, Vol. 26 (June, 1947), 68-74.
17. Remmers, H. H., and Gage, N. L. Educational Measurement and Evaluation. New York: Harper and Brothers, 1943.
18. Rothney, John M. "Evaluating and Reporting Pupil Progress," A report to the Department of Classroom Teachers, American Education Association. Washington, 1955.
19. Schwartz, Alfred, and Tiedman, Stuart C. Evaluating Student Progress in the Secondary School. 1st. ed. New York: Longmans-Green, 1957.
20. Sells, S. B. "Observational Methods of Research: rating scales," Review of Educational Research, Vol. 18 (December, 1948), 424-47.
21. Symonds, Percival M. Diagnosing Personality and Conduct. New York: The Century Co., 1931.
22. Weinland, James D. "Better Words on Rating Scales," Personnel Journal, Vol. 25 (October, 1946), 131-34.
23. Wood, Dorothy A. Personal Interview. University of North Carolina at Chapel Hill, Department of Psychology, April, 1965.
24. Wrightstone, Jacob Wayne, Justman, J., and Robbins, Irving. Evaluation in Modern Education. New York: American Book Co., 1956.
25. Wrightstone, Jacob Wayne. "Observational Techniques," Encyclopedia of Educational Research, (1960), 929-31.

I. GRAINLINE:

1. Grains run in all directions.
2. Lengthwise grain runs in direction of length.
3. Crosswise grain runs in direction of width.

II. SEAMING:

1. Straightening the grain of the fabric.
2. Straightening the grain of the fabric.
3. Straightening the grain of the fabric.
4. Straightening the grain of the fabric.
5. Straightening the grain of the fabric.

APPENDIX

III. SEAM FINISH:

1. Seams are even in width.
2. Shoulder, side, and sleeve seams are finished with 1/4" in width.
3. Seam allowances are finished on the outside of the garment.
4. Seams are made with neat finish.
5. Seams are pressed open.
6. Seams are pressed open before creased by another seam or fold.
7. Seams are pressed lightly enough so the seam lap does not show on the right side.
8. Seams are made at underarm.
9. Stitching looks the same on both sides (balanced finishing) throughout the garment.

## Description of a Top-Level Blouse

(\*: See "Instructions to the Judges")

## I. GRAINLINE:

1. Center back is on straight grain.\*
2. Lengthwise front edges are on straight grain.\*
3. Crosswise grain of sleeve is perpendicular to underarm sleeve seam.\*

## II. STAYSTITCHING:

1. Staystitching is done at the following places:
  - A. Neckline....shoulder to center on front and back. (4 units)
  - B. Extended facing....shoulder to center front. (2 units)
2. Staystitching stops at center back.
3. Stitch length is 12 to 15 stitches per inch.

## III. PLAIN SEAMS:

1. Seams are even in width.
2. Shoulder, side, armseye, and underarm seam allowances are  $\frac{5}{8}$ " in width.
3. Seam allowances have straight or true curved edges, except for notches.
4. Notches on side seams match.
5. Notches on armseye seams match.
6. Seams are pressed open.
7. Seams are pressed open before crossed by another seam or hem.
8. Seams are pressed lightly enough so the seam imprint does not show on the right side.
9. Seams coincide at underarm.
10. Stitching looks the same on both sides (balanced tension) throughout the garment.

## Description: (continued)

11. Stitch length is 12 to 15 stitches per inch.
12. There is slight ease in each back shoulder seam.
13. This slight ease in each shoulder seam is evenly distributed.

## IV. FACING:

## 1. Neckline:

- A. Grain of the facing matches grain of garment piece.
- B. Seam joining back neck and extended front facing is:
  - a. True.
  - b. Trimmed to  $1/4$ ".
  - c. Pressed open.
- C. Free edges of the facing are:
  - a. Turned and cleanfinished\* to form a smooth curve or straight line.
  - b. Topstitched an even distance from the turned back edge.
  - c. Pressed and lie flat.
  - d. Attached to blouse at shoulder seams only.
- D. Seam attaching facing to garment is:
  - a. On the marked curve or forms an even and symmetrical neckline.
  - b. Trimmed to  $1/4$ ".
  - c. Clipped to the seam line before understitching every  $1/2$ ".
- E. Understitching of facing:
  - a. Is done at the neckline within  $1/8$ " from the seam.
  - b. Steps an equal distance from each corner ( $1/4$  to  $1/2$ ").



Description: (continued)

F. Neckline corners are:

- a. Trimmed to eliminate bulk.
- b. Turned so they approximate right angles.

G. Neckline is pressed so facing does not show from the right side.

2. Lower edge of extended front:

- A. Is trimmed at hemline to 1/4".
- B. Has corners at hemline that approximate right angles.
- C. Is pressed so it lies flat.

V. BLOUSE HEM:

1. First turning of hem is 1/4" from the raw edge.
2. Hem is stitched:
  - A. An even distance from the turned edge.
  - B. Within 1/8" from the turned edge.
3. Hem is secured by backstitching or a knot on the wrong side.
4. Hem has well-pressed, sharp edges.

VI. DARTS:

1. A single straight line is traced for each stitching and fold line.
2. Small crosswise lines show where stitching should stop.
3. Tracing color is a shade darker or lighter than fabric to be marked.
4. Stitching tapers evenly to a point for each dart.
5. Corresponding darts are even in length.
6. Threads are hand-tied at point of each dart.
7. Threads are trimmed to approximately 1/2" for each dart.

Description: (continued)

8. Darts are pressed so they appear smooth on the right side.
9. Darts are pressed before they are crossed with a seam.
10. Horizontal darts are pressed toward the hem.
11. Vertical darts are pressed toward the center.

VII. SLEEVE:

1. Tracing on each sleeve cap is  $5/8$ " from the edge of each sleeve.
2. Each sleeve is cut so the edges are smooth or evenly curved.
3. Ease threads that would show on the right side are removed.
4. Stitching is on the marked seamline or forms a true armseye curve.
5. Center of each sleeve cap matches shoulder seam.
6. Ease is evenly distributed between notches on each sleeve cap.
7. There are no tucks or puckers in each sleeve cap area.
8. Each sleeve hem:
  - A. Is even in width.
  - B. Has stitching  $1/8$ " from the edge and equidistant.
  - C. Is secured by stitching approximately  $1/2$ " beyond the starting point.
  - D. Has no puckers or wrinkles.

VIII. GENERAL APPEARANCE:

1. Markings from the tracing wheel do not show on the right side.
2. No ends of threads show on the right side.
3. Threads are secured and pulled to the wrong side.

## Description: (continued)

4. Blouse is:
  - A. Free from soil.
  - B. Not scorched.
  - C. Not shiny on the right side.
5. Thread matches fabric.
6. Staystitching does not show on the right side.
7. Right and left front edges are the same length.

Instructions to the Judges

INSTRUCTIONS FOR PREPARING BLOUSE FOR SCORING GRAINLINE:

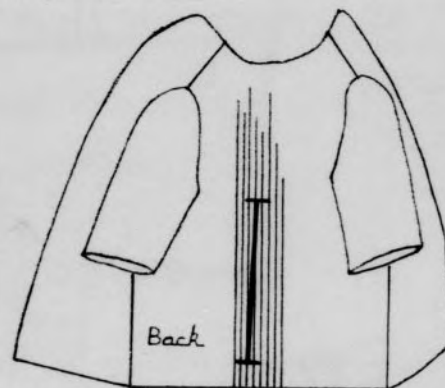
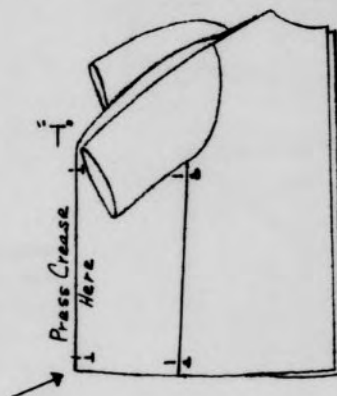
A demonstration will accompany this.

I. Center back: (center front if blouse opens at center front)

Place side seams so they coincide. Pin at top, then at hemline. Hand press a crease down center back. Pin. Open blouse and see if this pressed crease is on the straight grain.

Pinning at top, make pin level with underarm intersection as shown at "T".

Pinning at bottom, do this as close as possible.



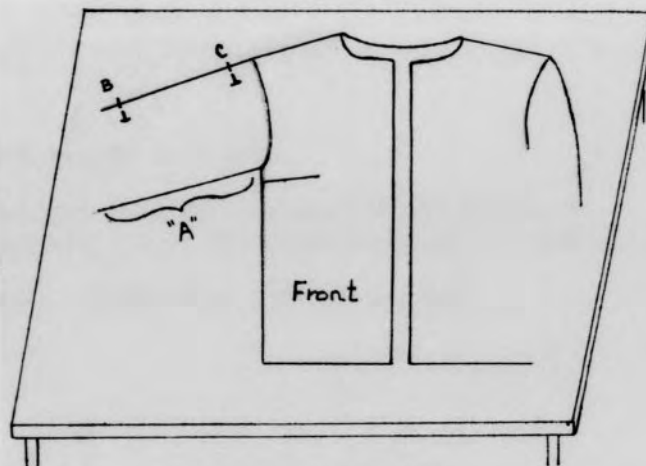
Crease off-grain; therefore, center back is off-grain.

II. Lengthwise front edges: (back edges if the blouse opens at center back)  
To be demonstrated.

Instructions to the Judges: (continued)

### III. Sleeves:

1. Lay right sleeve on the table as illustrated so blouse front faces upward. Hand crease along the stitching of seam at "A" and along the opposite side. Pin at "B" and "C".



2. As shown in diagram 1, place the triangle so that one side of the right angle coincides with the cross-wise grain of the sleeve at "D" and so the end of the other side of the right angle is in line with the edge of the sleeve as at "E". Measure the amount at "F" that the sleeve is off-grain.

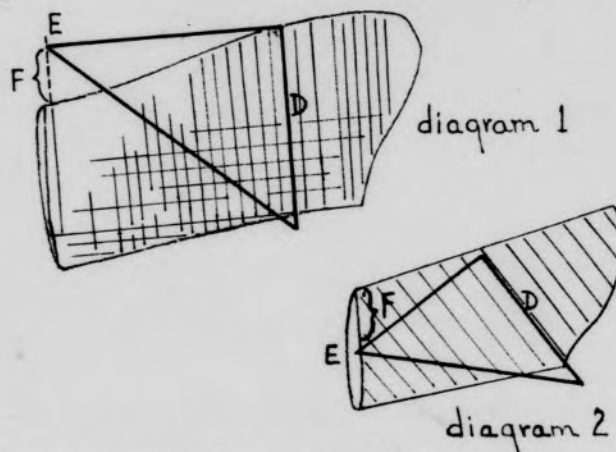


Diagram 2 shows how the grain could be off-grain in another direction.



Instructions to the Judges: (continued)

Definitions:

1. Free edge: the outside edge that is not caught in a seam.
2. Bridgestitch line: a line of stitching 1/4" from the cut edge of the fabric on which the free edges are turned for cleanfinishing or hemming.
3. Cleanfinished edge: the free edge that is turned once and topstitched.

71 SPECIFICATIONS FOR BLOUSE TO BE MADE DURING FIELD EXPERIMENT:

I. Pattern:

- A. Overblouse with machine-stitched hem.
- B. Collarless with extended front and fitted back neck facing.
- C. Set-in sleeves with machine-stitched hems.
- D. Shoulder and underarm darts.

II. TECHNIQUES:

- A. Understitching of facing: a line of stitching that keeps the facing from rolling to the right side. This stitching is placed 1/8" from the neckline seam and is machine-stitched through the facing and seam allowances.
- B. Bridgestitching: a line of stitching 1/4" from the cut edge of the fabric on which the free edges are to be turned for hemming or clean finishing.
  - 1. Free edges of the facing
  - 2. Lower edge of the blouse
  - 3. Lower edge of the sleeve
- C. Reinforcing of armseye seam: a row of stitching between the notches on the lower portion of the armseye seam. This stitching is placed 1/4" from the armseye seam.
- D. Clipping of armseye seam allowance: a row of clipping to the second row of stitching in the lower portion of the armseye. This clipping is done every 1/2".
- E. Finishing of lower front corners: facing of these corners.

DIRECTIONS:

1. Check the appropriate column:  
 yes--if the criterion has been met.  
 no--if the criterion has not been met completely.  
 I can not see this--if you can not evaluate the step.
2. Write a number in the last column to indicate how sure you are of the accuracy of scoring:  
 1--I could see this clearly.  
 2--I think I knew what I was looking for.  
 3--I took a wild guess.

3. Write a letter in the column to the left of the items to indicate:  
 a--the item repeats another item unnecessarily.  
 b--the item repeats another item but this is necessary.  
 c--the item does not repeat another item.

4. If an item seems to be especially time consuming, circle the item number and explain directly on the sheet.
5. Any additional comments are welcome.
6. The symbol \* means that you are to see "Instructions to the Judges".

Repeat	CRITERION	Yes	No	I see this.	How sure I am.
--------	-----------	-----	----	-------------	----------------

I. GRAINLINE:

1. Center back foldline is on straight grain.\*

2. Lengthwise front edges are on straight grain.\*

3. Cross-grain of each sleeve is perpendicular to its underarm seam.\*

II. STAYSTITCHING:

(If there is no staystitching, skip items numbered 2-8.)

### Rating Scale for Scoring a Blouse

1. Depending upon the description to which an item's quality corresponds or descriptions between which it falls, rate each item from 6-1. Record this score in the box on the right. If the construction process described was omitted, leave the item blank.
2. Symbols: a) \*: See "Instructions to the Judges".  
b) \*\*: To be used in scoring later.
3. Use a ruler and a tape measure for measuring where these are indicated.
4. Tear the tally sheets from each page and staple them together for each blouse scored.

6	5	4	3	2	1
---	---	---	---	---	---

#### I. GRAINLINE:

Blouse No. \_\_\_\_\_  
Time \_\_\_\_\_  
Name \_\_\_\_\_

- |   |   |   |
|---|---|---|
| <p>1. Center back foldline is</p> <p>on straight grain* (or misses by no more than two threads).</p>  | <p>off-grain by more than two threads, but no more than 1/4".*</p>    | <p>off-grain by more than 1/4".*</p>    |
| <p>2. Lengthwise front edge for</p> <p>a) left side (as wearing)</p> <p>b) right side</p> <p>is on straight grain* (or misses by no more than two threads).</p> | <p>is off-grain by more than two threads, but no more than 1/4".*</p> | <p>is off-grain by more than 1/4".*</p> |

GRAINLINE (continued)

6	5	4	3	2	1
---	---	---	---	---	---

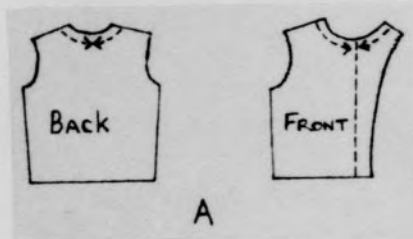
3. Cross grain of  
 a) left sleeve (as wearing)  
 b) right sleeve

is on grain\* (or misses by no more than two threads).

misses being on grain by more than 1/4".\*

is off-grain by more than 1/4".\*

II. STAYSTITCHING:



4. At the following places, (See diagram "A".)  
 a) neckline...shoulder to center on front and back (4 areas)  
 b) extended front facing...shoulder to center front (2 areas)

staystitching is done on

all areas (6).

all but two areas.

three to one areas.

(If only part of staystitching is done, score that part that is done for items 5-7.)

5. The number of stitches per inch is

12 or more.

(11)

(10)

(9)

(8 or less.)



## GRAINLINE (continued)

6	5	4	3	2	1
---	---	---	---	---	---

6. (If there is no marked seamline, measure from the stitched seamline.)

Staystitching is

within 1/8" from the marked seamline toward the cut edge.

between 1/8" and 1/4" from the marked seamline toward the cut edge in any unit.

more than 1/4" from the marked seamline toward the cut edge in any unit.

7. Staystitching shows

ONLY on the wrong side.

on the right side in one or two places

on the right side in more than two places.

### III. PLAIN SEAMS:

8. For the following seam types

a) shoulder

b) side

c) armseye (Do not count reinforcing stitching such as a second armseye stitching right outside the other.)

there is only one line of stitching for each seam.

there is part of a second line of stitching for either or both of these seams.

there are two or more lines of stitching for both of these seams.

6	5	4	3	2	1
---	---	---	---	---	---

(If seam allowances are trimmed, skip items 9-13.)

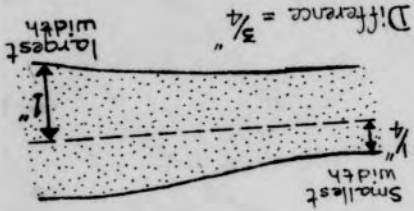
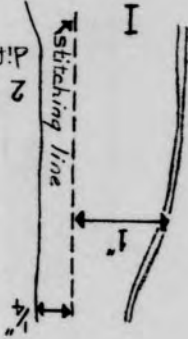
In item 9, you are measuring the amount in fractions of an inch by which the widths of the two seam allowances of a seam differ. See diagram "I".

In item 10, you are measuring the number of times the seam is uneven in width.

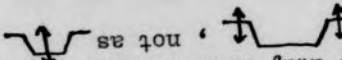
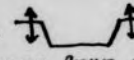
In item 11, you are comparing the combined width of the two seam allowances of a

seam with the width  $1\frac{1}{4}$ ".

2 Seam allowances differ in width by  $\frac{3}{4}$ "

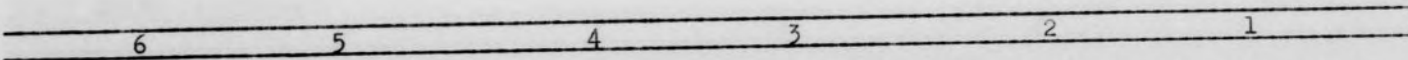


DIRECTIONS FOR MEASURING IN ITEMS 9 AND 11:

1. Use clear see-through ruler.
2. Measure to within  $1/16$ ".
3. Measure ravel of seam allowance. (This is considered a part of the seam allowance.)
4. If there are two lines of stitching, measure from the one farthest away from the raw edge.
5. Do not measure where there are notches; measure to either side as  , not as  .

DIRECTIONS FOR MEASURING IN ITEM 9:

For one seam type, locate the seam allowance with the largest width. Measure this width. On this same seam type, measure the smallest width on the other seam allowance. Compute the difference. Do not forget to do the same for the other seam of that same type.



9. Seams are stitched so that the widths of the two seam allowances

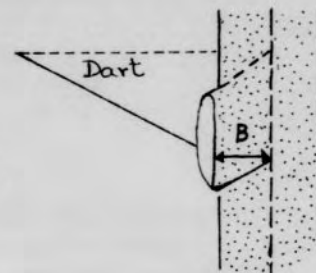
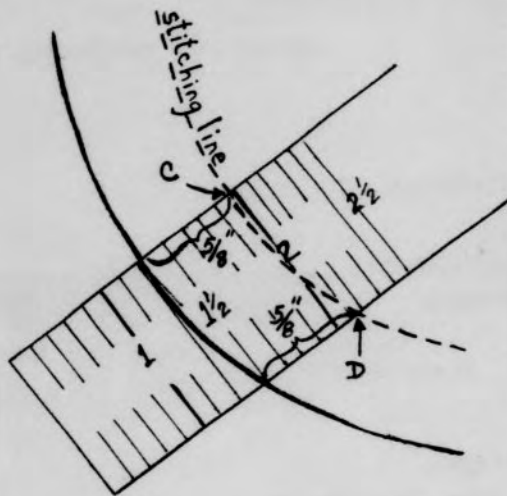
are the same for both seam types.

miss being the same for either or both seams of a type by no more than 1/8" (by no more than 1/4" for the armseye seam).

miss being the same for either or both seams of a type by more than 1/8" (by more than 1/4" for the armseye).

for the following seam types:

- shoulder (Measure only that portion that can be seen.)
- side (At the dart, measure the actual seam allowance as shown at "B".)
- armseye (Remember the different measure for this seam.) Lay the ruler on the seamline at any point so the ruler's one-inch or two-inch line coincides with the seamline at each end as shown at "C" and "D".



PLAIN SEAMS (continued)

6 5 4 3 2 1

10. Seams are stitched so that stitch is straight for following seam types.

a) shoulder

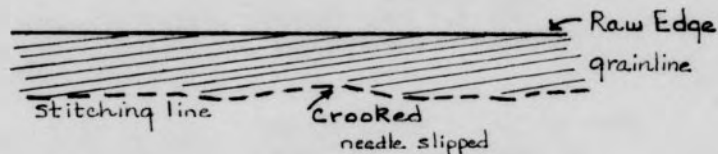
b) side

in all places for both seams.

in all places for one seam of a type and all except three places for the other seam of the same type.

for both seams with more than three places uneven.

Measure how many times the needle slips.

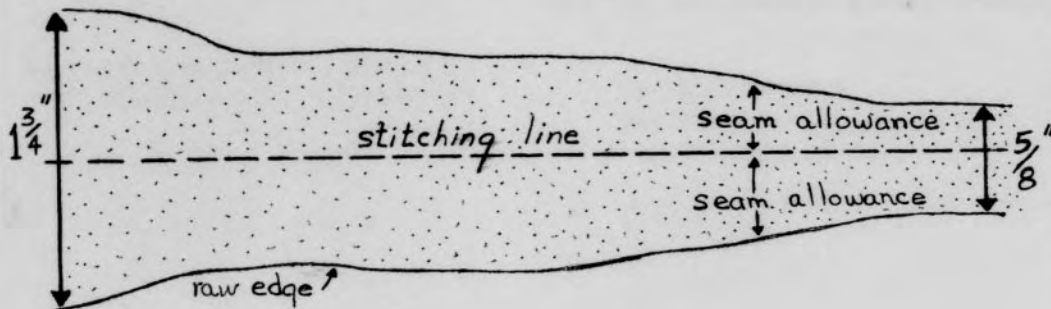


DIRECTIONS FOR MEASURING IN ITEM 11:

- On each seam, measure the largest width of the two seam allowances. Compare this measure with 1 1/4" and compute the difference.
- On each seam, measure the smallest width of the two seam allowances. Compare this measure with 1 1/4" and compute the difference.
- Use the larger of these two computed differences for evaluating each seam.

PLAIN SEAMS (continued)

6	5	4	3	2	1
---	---	---	---	---	---



(b)  $1\frac{1}{4}$ "  
 $-\frac{5}{8}$ "  
 $\frac{5}{8}$ " Difference

(c)  $\frac{5}{8}$ " = larger than  $\frac{1}{2}$ "

11. The width of seam allowances pressed open (if seam allowance is not pressed open, measure each allowance and add for total width.)

is  $1\frac{1}{4}$ " or misses this by no more than two threads for both seams of a type.

misses being  $1\frac{1}{4}$ " by no more than  $\frac{1}{4}$ " for either or both seams of a type (or by no more than  $\frac{1}{8}$ " for either or both shoulder seams).

misses being  $1\frac{1}{4}$ " by more than  $\frac{1}{4}$ " for either or both seams of a type (or by more than  $\frac{1}{8}$ " for the shoulder seam).

for the following seam types:

- shoulder (measure only that portion that shows and remember the different measure here.)
- side (remember to measure seam allowance as shown in 9 B's diagram.)
- armseye

PLAIN SEAMS (continued)

6	5	4	3	2	1
---	---	---	---	---	---

12. On both armseye seams, notches are

in right combination...single  
with single, double with double.

(no middle range  
description)

not in right combination or  
are not in right combina-  
tion for one armseye seam.

13. Points or edges of notches on either or both of following  
seam types

a) side

b) armseye (4 measurements here)

match or miss matching by no more  
than 1 or 2 threads.

miss matching by more than  
2 threads but no more than  
1/8" on either or both of  
side notches; for armseye,  
1 of 4 paired notches could  
miss matching by more than  
1/8", the other 3 by no  
more than 1/8".

miss matching by more than  
1/8".

14. Where crossed by another seam, of the following seam  
types,

a) shoulder (look only at armseye intersection...2  
possible)

b) side (look at armseye and hemline intersection...4  
possible)



PLAIN SEAMS (continued)

6 5 4 3 2 1

all intersections are open.  
(Ignore pleat or pucker; this  
will be evaluated later.)

c) sleeve underarm (look at armseye and hemline inter-  
section...4 possible)

no more than 2 intersections for side and sleeve underarm or no more than 1 intersection for shoulder are not open. neither intersection is open for shoulder; more than 2 intersections for side and sleeve underarm are not open.

15. Excepting armseye seam, both seams for the following seam types,

a) shoulder (feel here if necessary)

b) side

c) sleeve underarm (feel here if necessary)

are pressed open.

look or feel like they were pressed open, but I'm not definitely sure or one seam was and one was not pressed open. were not pressed open or if they have been, it's impossible to tell it.

16. Both seams are pressed open for the following types

a) shoulder (evaluate only that portion you can see)

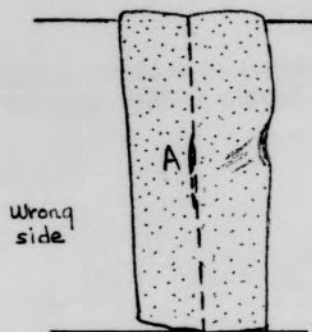
b) side

c) underarm sleeve (evaluate only visible portion)

PLAIN SEAMS (continued)

6 5 4 3 2 1

there are no puckers or pleats as shown at "A".



Wrong side

one of these seams has a small pucker or pleats.

both seams have a small pucker or pleat; or, either or both seams have small puckers or pleats.

17. The tension of stitching for
- a) the right hand side seam
  - b) understitching of facing
  - c) right armseye
  - d) blouse hem

is balanced (stitching looks the same on both sides.).

is not balanced; either the upper or lower tension appears somewhat loose or too tight.

appears unbalanced; either the upper or lower tension is so loose or so tight that the threads loop in the stitch.



PLAIN SEAMS (continued)

6 5 4 3 2 1

21. In the two back shoulder seams, this slight ease (or there may be more ease)

is evenly distributed.

is evenly distributed in one, but not in the other. is not evenly distributed.

DIRECTIONS FOR MEASURING IN ITEM 22:

Measure on right side of blouse (and only that portion that shows on the right side). Use ruler or compare seams and then measure the difference.



22. The length of seams on right and left sides of the blouse for the following seams

a) shoulder

b) side

c) underarm sleeve (unfold hem if turned up and measure as shown at "E").

d) armseye (Use tape measure to measure circular seam-line. Pin tape measure at starting point. Start at underarm sleeve seam intersection.)

are the same or miss it by no more than 1 or 2 threads.

differ by no more than 1/8" (1/4" for the armseye). differ by more than 1/8" (1/4" for the armseye).

IV. FACING: NECKLINE:

23. The neckline facing is cut so lengthwise grain

matches grain of garment.

misses matching grain of garment by no more than 1/8".

misses matching grain of garment by more than 1/8" or is cut on crosswise grain.

FACING: NECKLINE (continued)

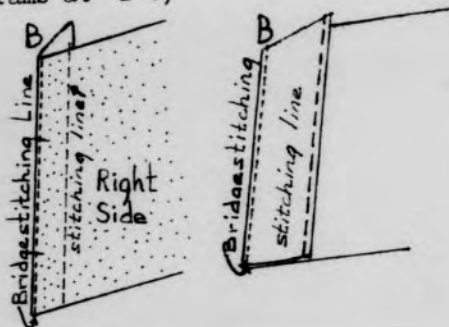
6	5	4	3	2	1
---	---	---	---	---	---

24. Seam joining back neck and extended front facing is

- |   |  |  |
|---|--|--|
| <p>a) straight. (look on right side.)</p> <p>b) trimmed to within 1/4" (not trimmed if only 1/4" seam allowance) for both seams.</p> <p>c) coinciding with shoulder seam at neckline intersection or missing it by 1 or 2 threads for both seams.</p> | <p>a) straight except for one place on one seam.</p> <p>b) trimmed to within 1/4" for one but not for the other.</p> <p>c) coinciding with shoulder seam on one side, but not the other by more than 1 or 2 threads.</p> | <p>a) crooked on one or both seams; or, straight except for one place on both seams.</p> <p>b) not trimmed at all; or, trimmed, but not to within 1/4" for both seams.</p> <p>c) not coinciding with either shoulder seam by more than 1 or 2 threads.</p> |
|---|--|--|

25. Free edges of the facing are

- |   |  |   |
|---|--|---|
| <p>a) turned either on the Bridge-stitch line or so this bridge-stitch line is turned under for the entire facing. (See diagrams at "B".)</p> | <p>a) turned, but the bridge-stitch line was not used because it is still on the topside of the facing; or, the bridge-stitch line was used only part of the time.</p> | <p>a) turned, but there is no bridge-stitch line.</p> |
|---|--|---|



FACING: NECKLINE (continued)

6	5	4	3	2	1
---	---	---	---	---	---

(continued) 25. Free edges of the facing are

- |  |   |  |
|--|---|--|
| b) stitched to form a smooth curve or straight edge for the entire facing. | b) stitched so a smooth curve and straight edge are in all but two places.  | b) stitched so there is not a smooth curve and straight edge in more than two places.  |
| c) stitched within 1/8" from the turned back edge for the entire facing.   | c) stitched within 1/8" from the turned back edge for about two-thirds of the facing.   | c) stitched within 1/8" from the turned back edge in less than two-thirds of the facing.   |
| d) pressed so there is a sharp edge.                                       | d) pressed, but the edges are not as sharp as they should be in all places.   | d) not even pressed.   |
| e) attached to the blouse at each shoulder.                                | e) attached to the blouse at one, but not at the other shoulder seam; or attached so this tacking does show from the right side in one place. | e) not attached at all; or attached at both shoulder seams so this tacking does show from the right side in more than one place. |



FACING: NECKLINE (continued)

6	5	4	3	2	1
---	---	---	---	---	---

DIRECTIONS FOR MEASURING IN ITEM 26:

Fold neckline in half and compare sides as shown at "A" for judging symmetry.



26. The seam attaching facing to garment is

a) on or very near the marked curve or forming an even and symmetrical neckline.

b) trimmed to within 1/4" (or trimmed so no more than 1/4").

a) on or very near the marked curve in all but three places, or forming an even and symmetrical neckline except for two places.

b) trimmed, but not to within 1/4" in all places.

a) not on or very near the marked curve in more than three places; or, not forming an even and symmetrical neckline in more than two places.

b) not trimmed at all; or, if trimmed, it's difficult to tell.

NECKLINE: FACING (continued)

6	5	4	3	2	1
---	---	---	---	---	---

(continued) 26. The seam attaching facing to garment is

- |  |   |   |
|--|---|---|
| c) clipped to within one or two threads of the seamline in all places where clipped or for at least $\frac{3}{4}$ of the seam attaching facing to garment. | c) clipped, but not to within one or two threads of the seamline. | c) not clipped at all, or, if clipped, it's very difficult to tell. |
|--|---|---|

(Skip "d" if no clipping)

- |                                   |   |  |
|-----------------------------------|---|--|
| d) clipped every $\frac{1}{2}$ ". | d) clipped about every $\frac{3}{4}$ ". | d) clipped, but the space between clipping is more than $\frac{3}{4}$ ". |
|-----------------------------------|---|--|

(If no understitching, skip items 27-28.)

27. Understitching of facing

- |  |   |   |
|--|---|---|
| a) is done within $\frac{1}{8}$ " of the neckline.                                     | a) is done within $\frac{1}{8}$ " of the neckline seam in all but two places. | a) is not done within $\frac{1}{8}$ " of the neckline seam in more than two places. |
| b) stops within 1" of each corner.   | b) stops within 1" of one corner but not the other.                           | b) does not stop within 1" of both corners.   |
| c) catches the seam allowances all the way. (Be sure to check neckline intersections.) | c) does not catch the seam allowances in at least 2 places.                   | c) does not catch the seam allowances in more than 2 places.                        |

FACING: NECKLINE: (continued)

6 5 4 3 2 1

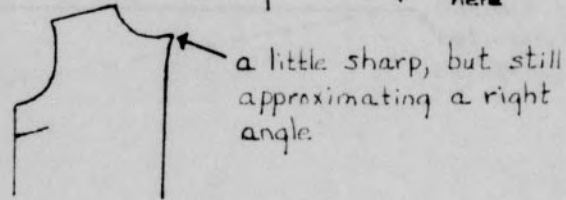
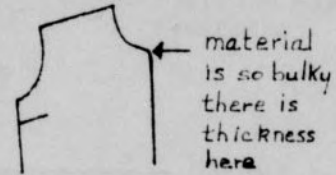
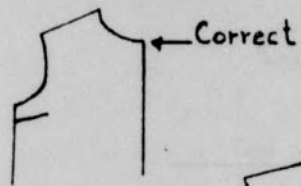
28. Understitching threads are pulled to the wrong side at each end. pulled to the wrong side for one but not for the other end. not pulled to the wrong side.
29. Neckline corners are turned so they approximate right angles. (See attached diagram.) turned so one does, but the other does not approximate a right angle. turned so neither approximates a right angle.
30. The facing does not roll to the right side. not roll to the right side in more than 2 places. does roll to the right side in more than 2 places.

V. DARTS:

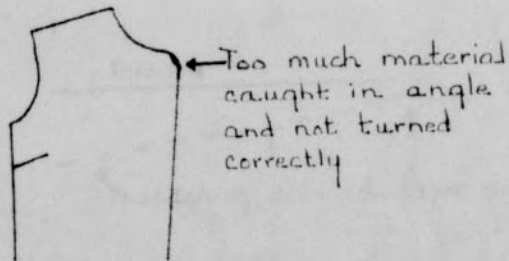
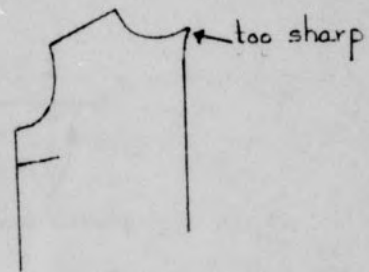
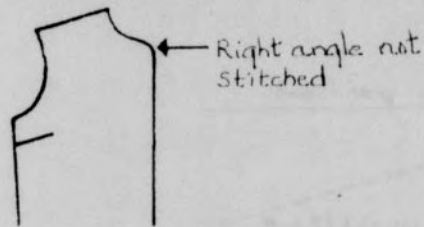
(If the darts are not traced, skip items 31-32.)

31. For each stitching and fold line for  
a) shoulder darts  
b) underarm darts
- there is a single traced straight line for both darts. there is a single traced line for one dart, but not the other. (There may be more than one line, or this line may be crooked for the other dart.) there is not a single traced straight line for both darts. (Again, these lines may be crooked, or there may be more than a single line.)

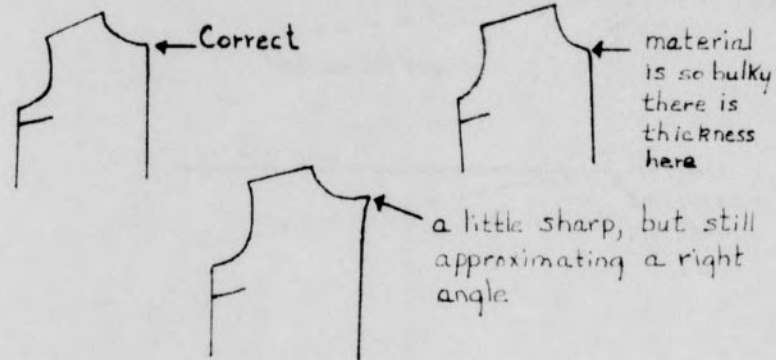
### Approximating Right Angles



### Not Approximating Right Angles

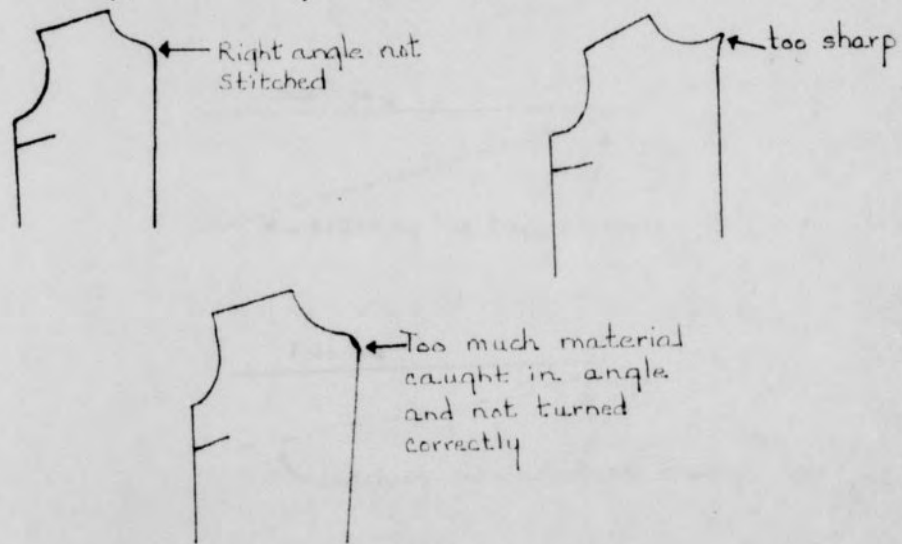


### Approximating Right Angles

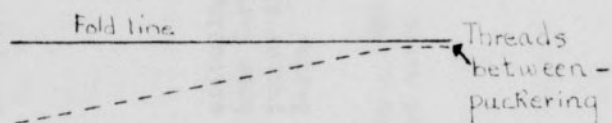
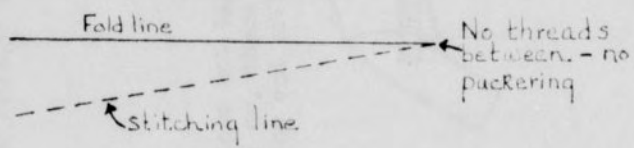



---

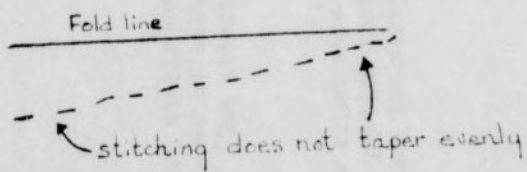
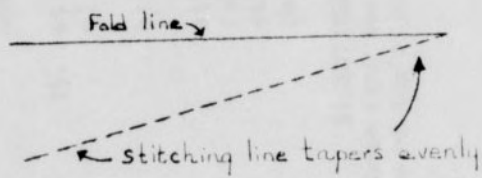
### Not Approximating Right Angles



## Puckering



## Tapering





DARTS: (continued)

6	5	4	3	2	1
---	---	---	---	---	---

32. A small crosswise line or "X" to show where stitching should stop for
- a) shoulder darts
  - b) underarm darts

is found for both darts.

is found for one dart, but not the other. is found for neither dart.

33. At the point of the dart for
- a) shoulder darts
  - b) underarm darts

stitching tapers evenly so there is no pucker at either dart's point. (See attached diagram.)

stitching tapers evenly for one, but not for the other; or, there is a pucker resulting.

stitching does not taper for both darts; or, there is a pucker at the end of both darts.

DIRECTIONS FOR MEASURING IN ITEM 34:

Measure on the right side of blouse from seamline to last stitch as shown at "R" and only that portion that shows on the right side. Use ruler or compare darts and then measure the difference.

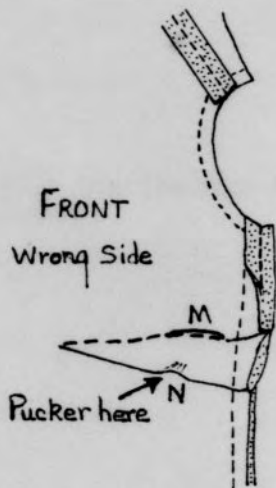


DARTS: (continued)

6	5	4	3	2	1
		34. Corresponding darts for			
		a) shoulder darts			
		b) underarm darts			
are the same length or miss it by no more than 1/16".		miss being the same length by no more than 1/4".		miss being the same length by more than 1/4".	
		35. At the point of the dart, threads are hand tied so they have for the following darts			
		a) shoulder darts			
		b) underarm darts			
not come loose for either dart.		not come loose for one dart, but they have for the other. (This dart may not have been tied, or it may have come loose.)		come loose for both darts; or, threads are not tied or ends are backstitched.	
		36. Of these threads for			
		a) shoulder darts			
		b) underarm darts			
both are trimmed to at least 3/4".		one is, but one is not trimmed to at least 3/4".		neither is trimmed.	

DARTS: (continued)

6 5 4 3 2 1



37. A pucker and/or pleats resulting from the dart not being pressed before crossed with a seam. (See diagram at "N") for
- shoulder darts
  - underarm darts

are found for neither dart.

are found for one, but not the other.

are found for both darts.

38. Pressing direction of the following darts
- shoulder
  - underarm

are correct for both darts: shoulder, toward the center; underarm, toward the hem.

are correct for one but not the other dart.

are not correct for either dart.

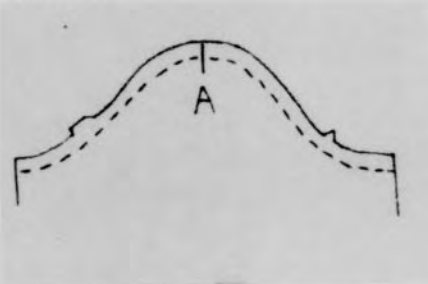
6	5	4	3	2	1
---	---	---	---	---	---

## VI. SLEEVE

(If there is no tracing, skip items 39-41.)

39. On sleeve caps, tracing is

5/8" from the edge for each sleeve.



matches shoulder seam on both sleeves.

5/8" from the edge for one but not for the other; or, tracing is no more than 3/4" and no less than 1/2" from the edge in places on both sleeves.

more than 3/4" or less than 1/2" from the edge on either or both sleeves.

40. The center of sleeve cap (indicated by a short line as at "A")

matches shoulder seam on 1 sleeve but on the other misses by no more than 1/4".

misses matching shoulder seam by more than 1/4" on either or both sleeves.

41. The dots on the sleeve cap

match the dots on the armhole on both sleeves.

match the dots on the armhole for one sleeve but miss on the other by no more than 1/4".

miss matching the dots on the armhole for either or both sleeves by more than 1/4".

SLEEVE: (continued)

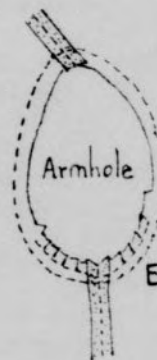
6	5	4	3	2	1
---	---	---	---	---	---

- are found for neither sleeve.
42. Ease threads showing on the right side
- are found for one sleeve but not the other; on this other sleeve, these threads do not show in more than three places.
- are found for either or both sleeves in more than three places.
43. The seam joining sleeve to garment for
- a) left sleeve (as worn)
- b) right sleeve
- is on or very near the marked curve or forming a true armseye curve.
- is on or very near the marked curve in all but three places; or, forming a true armseye curve in all but two places.
- is not on or very near the marked curve in more than three places; or, not forming a true armseye curve in more than two places.
44. Between notches on the sleeve cap for
- a) left sleeve (as worn)
- b) right sleeve
- the ease is distributed
- so there are no more than three pleats and/or three places where there are puckers.
- so there are more than three pleats and/or three places where there are puckers.

DARTS: (continued)

6 5 4 3 2 1

(See diagram "E" for items 45-48.)



45. Between the cut edge and armseye seam on the lower portion of the sleeves between the notches:

there is a row of stitching for both sleeves.

there is a row of stitching for one but not the other sleeve.

there is not a row of stitching for either sleeve.

46. This row of stitching is within 1/4" from the armseye seam toward the cut edge

on both sleeves.

on one but not the other sleeve.

on neither sleeve.

47. On the lower portion of  
a) right sleeve  
b) left sleeve



DARTS: (continued)

6	5	4	3	2	1
---	---	---	---	---	---

(47 continued)

clipped to or within 1 or 2 threads of this second line of stitching; or, clipped to or within 1 or 2 threads of this second line of stitching for at least 3/4 of the underarm area.

the seam allowance between the notches is

clipped within 1/8" of this second line of stitching; or, clipped within 1/8" of this second line of stitching for at least 3/4 of the underarm area. not clipped.

(Skip item 48 if the seam allowance is not clipped)

48. This clipping is done at least

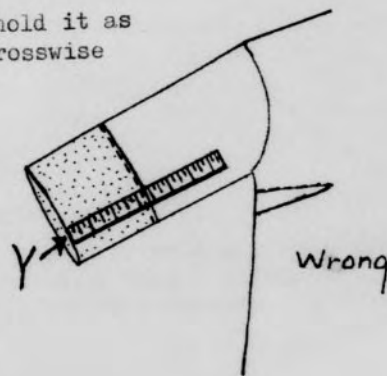
every 1/2" for both sleeves.

every 3/4" for both sleeves; every 1" for both sleeves. or every 1/2" for one and every 3/4" for the other.

VII. SLEEVE HEM:

DIRECTIONS FOR MEASURING IN ITEM 49:

Measure on blouse's wrong side. Use ruler; hold it as shown at "Y", making sure that the ruler's crosswise line coincides with the bottom hemline.



SLEEVE HEM: (continued)

6	5	4	3	2	1
---	---	---	---	---	---

49. The hem on the  
 a) left sleeve (as worn)  
 b) right sleeve

is even in width.

misses being even in width by no more than 1/8". misses being even in width by more than 1/8".

50. The width of both sleeve hems  
 (Use procedure for measuring as in item 49.)

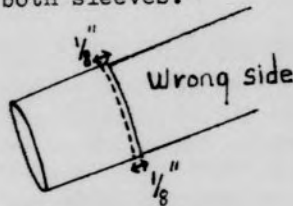
is the same or misses this by no more than 1 or 2 threads.

misses being the same by no more than 1/4". misses being the same by more than 1/4".

51. The stitch of the hem is (See diagram at "U".)

within 1/8" from the turned back edge for both sleeves.

within 1/4" from the turned back edge for either or both sleeves. not within 1/4" from the turned back edge for either or both sleeves.



52. This stitching

catches the fold all the way on both sleeves so free edges of the hem do not show.

catches the fold all the way on 1 sleeve; but, on the other, misses catching the fold in no more than 2 places.

misses catching the fold in more than 2 places on either or both sleeves.

SLEEVE HEM: (continued)

6                      5                      4                      3                      2                      1

1/2" beyond the starting point on both sleeves.

53. This stitching extends approximately

1/2" beyond the starting point on 1 sleeve; but, on the other, approximately 1"; or, does not extend at least 1/4" on 1 sleeve.

more than 1" on both sleeves; or, does not extend at least 1/4" on both sleeves.

are not found for either sleeve.

54. Puckers or diagonal wrinkles resulting from the hem being put in off-grain or by its being stretched

are found for 1 sleeve but not the other.

are found for both sleeves.

VIII. BLOUSE HEM:

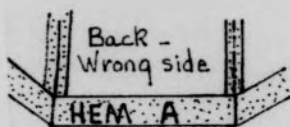
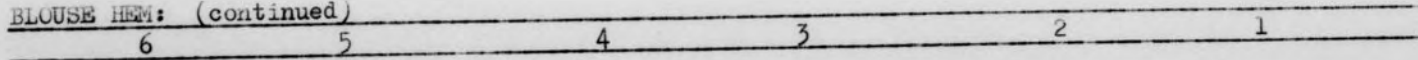
55. The first turning of the hem (Turn front facing wrong side out to check for seeing if there is a bridgestitch line.).

is either on the bridgestitch line or so this bridgestitch line is turned under for the entire hem. (See diagrams at "A".)

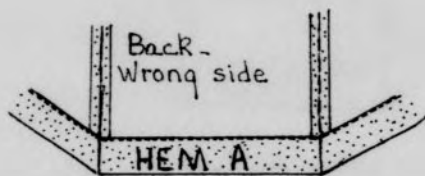
does not allow the bridgestitch line to be seen on the topside of the hem in more than three places. (This bridgestitch line was not used for turning in these places.) (See diagram at "B's" where bridgestitching shows.)

is not on the bridgestitch line because there is not one; or, allows the bridgestitch line to be seen on the topside of the hem in more than three places.

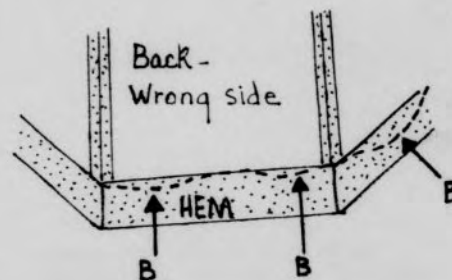
BLOUSE HEM: (continued)



Bridge stitching line  
not showing because  
turned under.



Bridge stitching line  
showing at fold of hem



is  $1/4$ " from the raw edge.

is no less than  $1/8$ " and  
no more than  $1/2$ " from  
the raw edge.

is less than  $1/8$ " or more  
than  $1/2$ " from the raw  
edge.

are not showing.

57. Free edges of the hem

are not showing except  
for 1 place.

are showing in more than  
1 place.

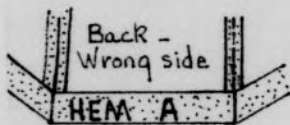
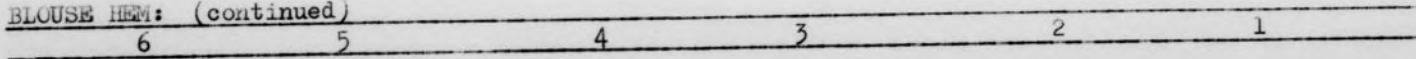
within  $1/8$ " from this first  
turned edge. (See diagram  
at "D".)

58. The hem is stitched

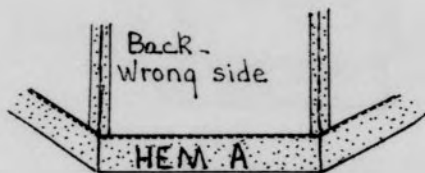
within  $3/16$ " from the  
first turned edge.

within  $1/4$ " from the first  
turned edge.

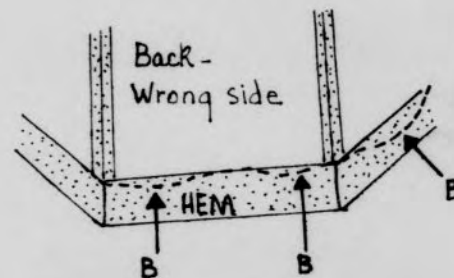
BLOUSE HEM: (continued)



Bridge stitching line  
not showing because  
turned under.



Bridge stitching line  
showing at fold of hem.



56. The first turning of the hem  
(Hold against window pane to see this.)

is  $1/4$ " from the raw edge.

is no less than  $1/8$ " and  
no more than  $1/2$ " from  
the raw edge.

is less than  $1/8$ " or more  
than  $1/2$ " from the raw  
edge.

57. Free edges of the hem

are not showing.

are not showing except  
for 1 place.

are showing in more than  
1 place.

58. The hem is stitched

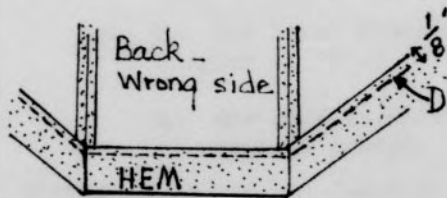
within  $1/8$ " from this first  
turned edge. (See diagram  
at "D".)

within  $3/16$ " from the  
first turned edge.

within  $1/4$ " from the first  
turned edge.

BLOUSE HEM: (continued)

6 5 4 3 2 1



59. Stitching of hem

does not run off to single thickness of fabric.

does not run off to single thickness of fabric in more than 2 places.

does run off the single thickness of fabric in more than 2 places.

60. This stitching of hem extends to

the front facing.

within 1/8" from the front facing on either or both sides.

within more than 1/8" from the front facing on either or both sides.

61. Looking on right side, this stitching of hem is

straight or almost straight.

crooked in no more than 5 places.

crooked in more than 5 places.

62. At each end of hem stitch, threads are

secured.

are secured for 1, but not the other.

are not secured.



BLOUSE HEM: (continued)

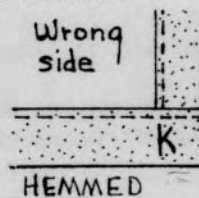
6	5	4	3	2	1
---	---	---	---	---	---

63. For the lower edges of extended front facing, (See diagram at "K".)

a) both edges are faced instead of hemmed.

a) one edge is faced; one edge is hemmed.

a) both edges are hemmed.



b) both are trimmed to within 1/4" (Hold against window pane to see.).

b) one is trimmed to within 1/4" but the other is not trimmed to within 1/4".

b) neither is trimmed to within 1/4".

c) both are pressed so they lie flat.

c) one is pressed so it lies flat; the other is not pressed or if pressed does not lie flat.

c) neither are pressed so they lie flat.

64. At the ends of the lower edges of extended front facing, for the corners, (See page 14 a for attached diagram.)

both approximate right angles.

one does, but the other does not approximate a right angle.

neither approximated a right angle.

BLOUSE HEM: (continued)

6	5	4	3	2	1
		65. The width of the bottom hem (Use ruler; measure largest width; measure smallest width; compute difference.)			
differs by no more than 1 or 2 threads.		differs by no more than 1/4".		differs by more than 1/4".	
		66. The edge (fold) of the hem is			
well-pressed and sharp for the entire hem.		well-pressed and sharp for at least two-thirds of the hem.		well-pressed and sharp for less than two-thirds of the hem.	
		67. Diagonal wrinkles or pleats from stretching fabric while hemming			
are not found.		are not found in more than 2 places.		are found in more than 2 places.	
		<u>IX. GENERAL APPEARANCE</u>			
		68. Trace markings on right side of garment			
are not found.		are not found in more than 2 places.		are found in more than 2 places.	
		69. Tracing color is			
a good choice of yellow, green, blue, orange, or red.		(no middle range description)		not a good choice of these colors.	
		70. Thread			
blends with the fabric.		(no middle range description)		does not blend with the fabric.	

GENERAL APPEARANCE: (continued)

	6	5	4	3	2	1
			71. The blouse			
is free from soil.			is soiled in 1 place.		is soiled in more than 1 place.	
			72. The garment is			
a) not scorched.			a) scorched in 1 place.		a) scorched in more than 1 place.	
b) not shiny on the right side.			b) shiny on the right side in 1 place.		b) shiny on the right side in more than 1 place.	
			73. Imprints of darts or seams on the right side			
are not found.			are not found in more than 2 places.		are found in more than 2 places.	
			74. Ends of threads showing on the right side (excepting easing threads)			
are not found.			are not found in more than 1 place.		are found in more than 1 place.	
			75. All threads on the wrong side are (excepting darts)			
trimmed to within 1".			trimmed to within 1 1/2"; or, one is not trimmed at all.		not trimmed to within 1 1/2"; or, more than one is not trimmed at all.	
			76. Seams pulling out that show on the right side (Look especially at: neckline and facing corners, side, under-arm sleeve, armseye seams.)			
are not found.			are not found in more than 1 place.		are found in more than 1 place.	