

AN EXPERIMENTAL STUDY TO INVESTIGATE THE
DEVELOPMENT OF A RATING SCALE FOR
EVALUATING GAME PLAY IN TENNIS

A Thesis
Presented to
the Faculty of the Women's Physical Education Department
Kansas State Teachers College Emporia

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
Laura L. Clark
May 1964

Thesis
1964
C

JAN 20 1964

LIBRARY

OF THE

SCHOOL

OF THE

Approval of Major Department

Jeanne C. Galley

Approval of Graduate Council

Jessie C. Boyer

209062

ACKNOWLEDGEMENTS

The author would like to express appreciation to the following people for their cooperation throughout the writing of this thesis:

To Miss Jeanne C. Galley of Kansas State Teachers College Emporia for her guidance and encouragement.

To Don Martin of Wichita High School North for his help with the statistics.

To R. O. Johnston of Anthony High School for his valuable aid on the illustrations.

TABLE OF CONTENTS

CHAPTER	PAGE
I. THE PROBLEM, DEFINITIONS OF TERMS USED, METHODS OF STUDY, IMPORTANCE OF THE STUDY, AND LIMITATIONS OF THE STUDY.	1
The Problem	1
Statement of the Problem.	1
Definitions of Terms Used	2
Standardized skill tests.	2
Observational check sheet or rating scale	2
Grading scale	3
Methods of Study.	3
Importance of the Study	4
Limitations of the Study.	5
II. REVIEW OF THE LITERATURE	6
Skill or performance tests.	6
Rating scales	8
Purpose of rating scales.	9
Reliability and validity of rating scales	9
Choosing the items to include in a rating scale	10
Selection of a grading scale.	11

CHAPTER	PAGE
III. PROCEDURES OF THE STUDY	13
Selection of Test Groups.	13
Selection of Judges	14
Selection of the Tests.	15
The Revised Dyer Backboard Test of Tennis Ability	15
The Miller-Broer Forehand Backhand Test of Tennis Ability.	16
Subjective Rating of the Judges	16
The Rating Scale.	16
Methods of Administration	17
IV. DISCUSSION OF THE DATA.	19
Results of the Scores of the Pilot Test	20
Results of Test Scores.	20
Comparison of Test Results.	31
V. SUMMARY AND CONCLUSIONS	37
Summary	37
Conclusions	39
BIBLIOGRAPHY	43
APPENDIX A	44
Results of the Scores of the Revised Dyer Backboard Test of Tennis Ability.	45

CHAPTER	PAGE
Results of the Miller-Broer Forehand Backhand Test of Tennis Ability	46
Results of the Scoring of the Judges' Subjec- tive Rating	47
Results of the Scores of the Rating Sheet . .	48
APPENDIX B	49
Comparison of Raw Scores and Scores on All Tests for the Forty-eight Subjects.	49
APPENDIX C	51
Revised Directions for Dyer Backboard Test of Tennis Ability.	51
Directions for the Miller-Broer Forehand Backhand Test of Tennis Ability	56
Instructions for Subjective Judges.	61
Tennis Check Sheet.	62

LIST OF TABLES

TABLE	PAGE
I. Scores and Computations of the Pilot Study. . .	21
II. Comparison of Agreement of Scores for All Tests	32
III. Comparison of the Results of the Rating Scale to those of the Revised Dyer, Miller-Broer and Subjective Judges	34
IV. Comparison of the Results of the Revised Dyer and Miller-Broer Tests to Those of the Subjective Judges	35
V. A Comparison of the Results of the Revised Dyer Test to Those of the Miller-Broer Test . . .	36

LIST OF FIGURES

FIGURE	PAGE
1. Frequency and Percentage Distributions for the Revised Dyer Backboard Test of Tennis Ability.	26
2. Frequency and Percentage Distributions for the Miller-Broer Forehand Backhand Test of Tennis Ability	27
3. Frequency and Percentage Distributions for the Subjective Ratings of the Judges.	28
4. Frequency and Percentage Distributions for the Rating Scale.	29
5. Completed Score Card for Forehand-Backhand Drives.	60
6. Special Court Markings for Forehand-Backhand Drive Test.	60

CHAPTER I

THE PROBLEM, DEFINITIONS OF TERMS USED, METHODS OF STUDY, IMPORTANCE OF THE STUDY, AND LIMITATIONS OF THE STUDY

Advancement in the educational system of America depends upon evaluation. All progression is based upon some type of testing. There are many tools of evaluation and they are used by educators in every teaching field, however, many of these tools are designed to measure only mental knowledge or specific and isolated motor skills.

Physical educators throughout the country are cognizant of the need for an evaluation tool to be used in measuring the physical performance or ability in the game situation. Since the majority of the time spent in these classes is to develop a degree of skill in one activity or another there is a need for measuring the progress of the students in the activity as a whole. The use of standardized skill tests to measure specific isolated skills is now being practiced by many physical education teachers and coaches as a means of classifying and evaluating, but there is no valid instrument devised to measure actual game play.

I. THE PROBLEM

Statement of the problem. It was the purpose of this study to devise a practical and efficient method of

evaluating beginning and intermediate tennis skill through the use of a rating scale for grading game play. In accomplishing this, the following questions were to be answered:

1. Will there be a significant correlation between the scores on the two standardized skill tests in tennis?
2. Will there be a significant correlation between the scores of the two standardized skill tests and the results of the subjective ratings of the judges?
3. Will there be a significant correlation between the scores of the two standardized skill tests and the results of the rating scale?
4. Will there be a significant correlation between the results of the subjective ratings of the judges and the results of the rating scale?
5. Will the rating scale be a valid instrument for use in grading tennis skill?

II. DEFINITIONS OF TERMS USED

Standardized skill tests. Tests that have been proven to be valid and reliable in testing specific skills in tennis.

Observational check sheet or rating scale. An itemized list containing various skills of the activity involved. These skills are to be observed in the game situation and given a score by a qualified person. In this

study the term rating scale will be used throughout.

Grading scale. An instrument by which the items of the rating scale are scored.

III. METHODS OF STUDY

This study was conducted at Wichita High School North. The rating scale was constructed from specific skills in tennis that were taught at the beginning and intermediate levels. An effort was made to develop a rating scale that would comply to the basic skill objectives desired in a beginning or intermediate tennis class. When the scale was constructed a pilot study by three qualified judges and a group of sixteen high school girls in the after school tennis program was conducted. Each girl was rated by all the judges on her game play which was broken down into various skills. An item analysis was then made by comparing the results of the judges on each item to determine if all items were necessary and to determine the reliability of the rating scale.

After the rating scale was proved as to worth, the actual comparison of the tests began. The forty-eight subjects in the study were given the following tests: the Revised Dyer Backboard Test of Tennis Ability; the Miller-Breer Forehand Backhand Test of Tennis Ability; a rating on the rating scale; and a subjective score of playing ability

by judges. All testing was done in the regular high school physical education class for girls. The tests were given at the end of the instructional period, which was four weeks in duration. The two standardized skill tests, the rating scale score, and the subjective ratings of the judges were given within a week of each other to help eliminate the variable of improvement from practice between tests.

After the pilot testing of the rating scale the reliability of the ratings were checked by computing the deviation from the mean for each set of scores for each subject. After all tests were given the results were compared by the use of z scores. These scores were then grouped and plotted on a Z-scale to show any relationships between the tests.

IV. IMPORTANCE OF THE STUDY

"All authorities are agreed that the highest percentage of the physical education grade should be allotted to skill, because it is such an important objective and the greatest proportion of time is given directly to its development."¹ In today's large classes there is a

¹Marion R. Broer, "Are Our Physical Education Grades Fair?" Journal of Health Physical Education Recreation, 30:49, March 1959.

definite need for a practical method of evaluation that takes no special equipment and does not waste playing time for special testing. "In physical education, a measure is needed of the use that a player makes of his skills and knowledge in a game situation."² Also, such an instrument would enable the teacher to measure her students against the standards and objectives she has set up for them and at their own skill level.

V. LIMITATIONS OF THE STUDY

The findings of this study were limited to eleventh grade girls in high school. Also, due to the girls' ability only beginning and intermediate skills could be rated.

²Gladys M. Scott and Esther French, Measurement and Evaluation in Physical Education (Dubuque: Wm. C. Brown Company Publishers, 1959), pp. 221-228.

CHAPTER II

REVIEW OF THE LITERATURE

Generally there are two types of tests: teaching and research. Teaching tests characteristically possess practical value as teaching devices and research tests are designed for research purposes only.¹ Both the skill tests and the rating scale used in this study would be classified as teaching tests.

Standardized skill tests and rating scales each have individual characteristics that are sometimes identical and sometimes very different, therefore, for sake of clarity a brief summary of each one will be given.

SKILL OR PERFORMANCE TESTS

Skill and performance tests are used extensively by many physical educators today but prior to 1924, most fundamental sport skill test items were included as part of tests of general motor ability. Until then no attempt was made to evaluate the ability or skill used in a particular game. Development of skill tests increased in the 1930's and by 1940, skill tests began to appear in many activity areas.²

¹Carlton R. Meyers and T. Erwin Blesh, Measurement of Physical Education. (New York: P.S. Crofts and Co., 1962), p.10.

²Ibid.

The two tennis skill tests used in this study were the Revised Dyer Backboard Test of Tennis Ability and the Miller-Broer Forehand Backhand Test of Tennis Ability. They were chosen because they coincide with the basic criteria for a good skill test in that they take little time, are easy to administer, take no special equipment, and both tests have fairly high reliability and validity ratings for determination of tennis playing ability.

Studies have shown the Revised Dyer Backboard Test of Tennis Ability to have a reliability of .70 and a validity of .92 for prediction of tennis playing ability. The Miller-Broer Forehand Backhand Test also was designed to measure tennis playing ability and has a reliability of .80 and a validity of .85. According to Scott and French, anything above .85 is considered excellent. Validity coefficients seldom go above .85, reliability coefficients frequently do. From .75 to .85 is considered adequate for many purposes. As reliability coefficients drop below .75, they indicate an inconsistent and poor measuring instrument. A validity coefficient below .60 to .65 indicates poor predictive value but the test may be used to serve other purposes.³

³Clayds H. Scott and Esther French, op. cit., pp. 221-228.

RATING SCALES

A rating in its simplest form is merely the opinion of one individual concerning the trait or ability being rated. It is limited in its accuracy by the experiences and the insight of the rater.⁴

Traditional rating scales were developed primarily for appraisal purposes, that is to determine the total achievement throughout the six weeks, semester or year. The reliability of these tests were low because the concepts being rated were vague and general. Scores ranged widely when graded by two different teachers because of differences of opinion.

When general concepts as personality, skill or ability are rated the results are almost certainly to possess low reliability. When, on the other hand, such concepts are broken down and analyzed into their constituent elements, or defined, reliability is increased. This type of device is known as a graphic rating scale and the reliability is increased markedly when using this type.⁵

⁴Carl E. Willgoose, Evaluation in Health Education and Physical Education (New York: McGraw-Hill Book Company, Inc., 1961), p. 215.

⁵Ernest Tieg, Tests and Measurements in the Improvement of Learning (Boston: Houghton Mifflin Co., 1934) p. 375.

To the teacher who knows for what to look, observation can be a most fruitful evaluation technique. A simple list of things to look for when engaged in appraising pupils makes the difference between a thorough and objective evaluation and one of limited value. The rating scale brings order to observation. It is generally put together after an analysis of pupil behavior in a given situation.⁶ The values or criteria against which to evaluate achievement in many experiences are the standards and rules of the activity. These are also modified according to the goals which are desired.⁷ The rating scale may be used to appraise individual social efficiency on the playground or to measure the outcomes of a unit of work.⁸

Purpose of rating scales: These instruments may be used in physical education to evaluate the ability of a person for the purpose of grading him or classifying him to his proper group.⁹

Reliability and Validity of rating scales. "Experience and several studies have indicated that this type of rating

⁶Willgoose, op. cit., p. 215.

⁷Phillip A. Smithells and Peter E. Cameron, Principles of Evaluation (New York: Harper and Brothers, Publishers, 1962), p. 228.

⁸Willgoose, op. cit., p. 375.

⁹Charles Harold McCloy, Tests and Measurements in Health and Physical Education (New York: F.S. Crofts and Co., 1942), p. 91.

has a reliability of about .80 between the study of two raters only, and its validity is probably high."¹⁰

In an unpublished study, George Wells found that the ratings of a football squad made by three coaches correlated .80 with the ratings of the same squad made by seventy-three students. The intercorrelation of the ratings of the three coaches averaged .90.¹¹

McCloy gives several suggestions for increasing the reliability and validity of these types of tests. Those that pertain to this study are: (1) Ratings that represent a composite of a number of raters are usually more valid than those by one rater, (2) Ratings which are compromised from a number of small items are usually more valid than the rating of one large item, (3) Ratings made by a person of the same sex as the one being rated are usually more valid, (4) To check the reliability of the ratings, correlate the ratings of one individual against the other, (5) Have all raters rate a known group and see how they differ.¹²

Choosing the items to include in a rating scale.

The items may be selected by securing the opinion of

¹⁰Ibid., pp. 93-94.

¹¹Ibid., p. 95.

¹²Ibid.

experts as to what is most important. In another method the researcher may carry on an analysis of all skills that are performed during an activity and tabulate the most frequent and the most important items.

Each item should be carefully defined and delimited: (1) make definitions as objective as possible, (2) state them in terms of observed behavior, (3) sometimes the negative aspect is more readily observed and defined than the positive aspect.¹³

Selection of a Grading Scale. The type of scale used depends upon the type of information to be obtained. Sometimes it is necessary to use numbers, symbols, or descriptive words to grade the subject. Sometimes it is preferable to combine numerical and graphic types in situations where a numerical score is desired.

The number of categories used depends upon the degree of exactness desired. If only two categories are used, the discrimination would show the extremes but the varying degrees of behavior between the extremes would be disregarded. If 100 categories were used, the discrimination from one extreme to the other would be broken down into such minute details, the instrument would be impossible to use. In

¹³Ibid.

general, three to five categories provide adequate data and make it possible for the rater to distinguish accurately between the units of the scale.

Words such as excellent, good, average, poor have little meaning in themselves and should be further defined if they are used. Words and phrases identifying the particular behavior being rated should be used, and they should be specifically defined in the directions to the rater, by describing the particular behaviors to which they refer.¹⁴

¹⁴Marjorie Latchaw and Camille Brown, The Evaluation Process in Health Education Physical Education and Recreation (Englewood Cliffs: Prentice-Hall, Inc., 1961), p. 205.

CHAPTER III

PROCEDURES OF THE STUDY

The research for this study was done at Wichita High School North in the Girls' Physical Education Department. The subjects in this study were all junior girls in high school and students in the regular physical education program. The objectives of this study were (a) to select a pilot group to test the reliability of the items presented on the rating scale; (b) to give each subject the Revised Dyer Backboard Test of Tennis Ability; (c) to give each subject the Miller-Broer Forehand Backhand Test of Tennis Ability; (d) to have each subject rated subjectively by a group of judges; (e) to give each subject a score on the rating scale; (f) to compare the scores on all tests to determine if there is any significant correlation among them; (g) to devise a rating scale suitable for grading tennis skill.

SELECTION OF TEST GROUPS

The sixteen girls that were used as subjects in the pilot testing of the check sheet were selected at random from the after school tennis program. This program included sophomores, juniors, and seniors in high school and all participants were beginning or intermediate tennis players. These girls were used only to check the reliability of the judges'

ratings and the necessity of items on the rating scale.

The forty-eight girls in the actual experiment were students within the regular physical education classes who had just finished a four week unit in beginning and intermediate tennis.

Each girl was given the following tests: The Revised Dyer Backboard Test of Tennis Ability; The Miller-Broer Forehand Backhand Test of Tennis Ability; a subjective rating by judges; and a score on the rating sheet by her regular teacher.

SELECTION OF JUDGES

The judges used in this study were three former physical education teachers. All three have taught physical education previously in public high schools and have taught tennis to this age group. Due to this fact they were all aware of the difficulty of the skills involved and able to grade them more accurately. Also, since they did not know any of the subjects the possibility of grading anything other than skill was somewhat eliminated. Each judge was given a sheet of instructions (see Appendix C) and an opportunity to ask questions before judging the girls.

SELECTION OF THE TESTS

In this study each subject was given four different tests to determine her playing ability. The standardized skill tests that were used were The Revised Dyer Backboard Test of Tennis Ability and The Miller-Broer Forehand Backhand Test of Tennis Ability. Both tests have fairly high validity ratings themselves but a low correlation because they do not measure entirely the same things.¹ A brief explanation of each test will be given in the following paragraphs.

The Revised Dyer Backboard Test of Tennis Ability.

As stated above this test was chosen because of its high validity rating for predicting tennis ability, its reliability, and also for its ease and speed of administration. The Revised Dyer test in this experiment was administered according to instructions in Appendix C.

The Miller-Broer Forehand Backhand Test of Tennis Ability.

This test was chosen because it measures the ability of the students to hit a good forehand and backhand drive. "Since the Dyer test gives only one score as

¹Katharine Fox, "A Study of the Validity of the Dyer Backboard Test and the Miller Forehand Backhand Test for Beginning Tennis Players", The Research Quarterly, March 1953, Vol. 24, pp. 1-8.

a measure of tennis ability it does not indicate what primary area a student's weakness lies."² The author felt it was necessary to include this test because it tested the student's ability to drive the ball deep into back court, which is very desirable in game play. The directions for administration may be found in Appendix C.

Subjective Rating of the Judges. All forty-eight subjects in the test were given a subjective rating on their game play. The judges watched the girls play doubles for four games before giving a rating for any of the players. The author felt this length of time was sufficient to judge a girl's ability on the court. The rating was purely subjective and no check sheets or lists were used. The judges rated the players from one to ten. The scores ten to eight were given for excellent playing, seven to six for good playing, five to three for fair and two to one for poor playing. Each of these terms were defined for the judges in their instructions which may be found in Appendix C.

The Rating Scale. The rating scale was constructed from the various skills used in beginning and intermediate tennis. Each skill was broken down into various parts for ease in rating. While the students played tournament games their instructor rated them on the sheet. Whenever a play

²Ibid.

was made the girl was rated on the various items. A check (✓) was used to represent a skill well done; a minus (-) was used if the skill was done but lacked form; and a zero (0) was used if the skill was failed completely. The girls were rated while playing four games of doubles. After the girls had played four games their rating scores were computed on a basis from one to ten with the following formula:

$$\frac{2\checkmark + -}{\text{total of all marks}} (10) = \text{Score.}$$

The grading scale went from zero, which was a complete failure, to ten which represented a perfect score. Each teacher rated her own class. An example of the rating scale used in this study may be found in Appendix D.

METHODS OF ADMINISTRATION

All forty-eight subjects used in this study had completed a unit in tennis that included instruction and practice on the various beginning and intermediate skills before taking any of the tests. The tests were given within one week of each other to prevent the possibility of improvement between tests.

All tests, with the exception of the subjective rating of the judges, were given by the individual teacher to her own class. This was done to prevent any confusion that may have been caused by having a different teacher and to keep the schedule as near normal as possible. Only two teachers were used so interpretation of directions would be

as close as possible. In each administration the teachers read identical instructions and answered questions concerning the directions. The Dyer test was scored by the teacher in charge. The Miller-Broer test was scored by twenty student judges by use of individual score cards. (Appendix C). These girls had previous instructions on how to score the test. Their duty was to record on the score card the exact spot each ball hit on the court and to indicate whether it went over or under the rope. The scores were determined by the author by use of the score cards. The subjective ratings were given by the selected judges previously mentioned and the score on the rating sheet was given by the individual teachers.

CHAPTER IV

DISCUSSION OF THE DATA

Before the four tests used in the study could be compared and an effective rating sheet compiled it was necessary to run an initial test of the proposed rating sheet. This pilot test was done in the after school tennis program.

Three physical education teachers at Wichita High School North rated a group of sixteen girls on their tennis playing ability. This group was composed of sophomore and junior girls in high school with beginning and intermediate tennis skill ability. The girls were rated while playing one set of doubles. When the ratings were figured for each girl using the formula, $\frac{2\sqrt{+-}}{\text{total of all marks}}(10) = \text{Score}$, the results of the judges were then compared to determine the reliability of the rating sheet. This was done by using the average and maximum deviations from the mean for each girl. After the mean score was established for each girl the raw score was then subtracted from the mean to determine the deviation for each score. (Table 1). The average deviation from the mean was then determined by taking an average of these scores and dividing by the mean to arrive at the per cent of deviation. The maximum deviation was computed by dividing the highest raw score by the mean.

RESULTS OF THE SCORES OF THE PILOT TEST

The data shows the ratings of the judges for each girl to compare very closely. Statistically speaking, anything below forty per cent would be considered very good for a comparison of this type. The comparison of the results of the judges in the pilot study showed the average deviation from the mean never to be above twenty-seven per cent and in only three instances was it above twenty per cent. The maximum deviations were very close also with only two scores above forty per cent. This would indicate that the rating sheet was a reliable instrument.

RESULTS OF TEST SCORES

Before a raw score could be determined for each test several procedures had to take place. The raw score for the Revised Dyer Backboard Test of Tennis Ability was determined by taking the sum of the three trials. For the Miller-Broer Forehand Backhand Test of Tennis Ability the total raw score was computed from the sum of the fourteen forehand trials and the fourteen backhand trials. The raw score for the subjective ratings of the judges was the average of the scores of the three judges. The raw scores for the rating sheet were computed by use of the following formula:

$$\frac{2V + \text{total of all marks}}{10} = \text{Score}$$

A comparison of tests would not be possible unless the

TABLE I

SCORES AND COMPUTATIONS OF THE PILOT STUDY

Students	J U D G E S						C O M P A R I S O N O F S C O R E S		
	Hawk		Haskin		Clark		Mean	Average Deviation From Mean	Maximum Deviation From Mean
	Raw Score	Deviation From Mean	Raw Score	Deviation From Mean	Raw Score	Deviation From Mean			
Simmons	6.90	+ .63	5.85	- .42	6.05	- .22	6.27	7 %	10 %
Barnes	3.55	- .03	3.60	+ .02	3.60	+ .02	3.58	1 %	1 %
Walker	3.65	-1.07	5.75	+1.03	4.75	+ .03	4.72	15 %	23 %
Harper	5.60	+ .35	5.15	- .10	5.00	- .25	5.25	5 %	7 %
Bishop	7.90	+1.52	5.15	-1.23	6.10	- .28	6.38	16 %	24 %
Gouchenour	2.15	-1.48	4.65	+1.02	4.10	+ .47	3.63	27 %	41 %
Wells	7.70	+1.68	5.15	- .87	5.20	- .82	6.02	19 %	28 %
Telford	4.65	- .02	3.60	-1.07	5.75	+ .27	4.67	10 %	23 %
Hunsinger	5.35	+1.57	3.50	- .28	2.50	-1.28	3.78	28 %	42 %
Turner	8.75	+1.17	6.15	-1.43	7.85	+ .27	7.58	13 %	19 %
Carter	4.80	- .30	6.10	+1.00	4.40	- .70	5.10	13 %	20 %
Firsten- berger	5.25	+ .03	5.40	+ .18	5.00	- .10	5.22	2 %	3 %
Neek	6.65	- .77	8.50	+2.68	7.10	-1.28	7.42	21 %	36 %
Carpenter	6.10	+ .28	5.30	- .52	6.05	+ .23	5.82	6 %	9 %
Matthews	6.00	+ .08	5.35	- .57	6.40	+ .48	5.92	6 %	10 %
Brown	4.35	+ .28	3.85	- .22	4.00	- .07	4.07	5 %	7 %

raw scores of each test were reduced to a common denominator to give them an equal meaning. After the raw scores were known for all tests they were each given a *s*-score and placed on a *Z* scale.

To arrive at the *s*-score the following procedure was used: The mean score was found for each test, the mean was then subtracted from all scores and this number was then squared. The standard deviation was then calculated by the following formula: $\sqrt{\frac{\text{total of } m^2}{48}}$. When the standard deviation was established the following formula was used to determine the *s*-score: $20\left(\frac{\text{score} - \text{mean}}{SD}\right) + \text{mean} = Z \text{ score}$ (See Appendix A, page 45).

The *s*-scores were put on bar graphs by use of a *Z* scale ranging from five to one hundred and five to determine the distribution of scores for all tests. The graphs were divided into five divisions for ease of comparison. The frequency distributions for each division were converted to percentages and compared to those of the Normal Bell Curve. The distribution of scores for the rating sheet showed the highest correlation with the bell curve than any of the other tests. (See Figures 1-4, page 26).

It was found on the Revised Dyer Backboard Test of Tennis Ability of the forty-eight subjects, ten percent of the scores were at the bottom of the scale with a score of under twenty five; forty per cent of the scores were

between twenty-five and forty-five; twenty-nine per cent were between forty-five and sixty-five; fourteen per cent were between sixty-five and eighty-five and one per cent of the scores were over eighty-five. Statistically speaking the curve formed by this distribution of scores was satisfactory and close to the normal curve described by Bell. The majority of the scores fell into the second bracket and decreased toward the top of the scale which tends to indicate the test was a little difficult for the skill of the majority of the players involved. According to the scores on the Revised Dyer Backboard Test of Tennis Ability most of the players involved were of beginning or intermediate ability. (See Figure 1, page 26).

The Miller-Broer Forehand Backhand Test of Tennis Ability showed nine per cent of the scores under twenty-five; thirty-six per cent of the scores were between twenty-five and forty-five; thirty-three per cent between forty-five and sixty-five; fifteen per cent between sixty-five and eighty-five, and seven per cent of the scores at the top of the scale were over eighty-five. The frequency distribution for scores on this test were well balanced, with nine per cent at the lower end of the scale and the desired seven per cent at the top of the scale, however the curve was rather unusual because the scores were more frequent in the second division from twenty-five and forty-five. Nevertheless, the

the curve formed by the distribution of these scores is very good. (See Figure 2, page 27).

The subjective ratings of the three judges were poorly distributed with seventeen per cent of the scores below twenty-five; twenty-seven per cent from twenty-five to forty-five; twenty-seven per cent between forty-five and sixty-five; twenty-nine per cent from sixty-five to eighty-five; and no scores above eighty-five. This lack of discrimination may have been due to several factors: (1) the rating was purely subjective, therefore difficult to judge a girl's ability as a whole, (2) game play is difficult to judge without employing some method of noting particular skills being used, (3) the length of time was possibly too short for a judgement of this type, (4) the differences of opinions of the judges as to which phases of the game play are most important, and (5) when judging another's ability one tends to never give a perfect score or an absolute zero. Therefore, discrimination with this type of testing is usually poor. (See Figure 3, page 28).

The rating sheet results of the two teachers had the best distribution of scores with twelve per cent of the scores below twenty-five; twenty-three per cent of the scores fell between twenty-five and forty-five; forty-two per cent of the scores were between forty-five and sixty-five; sixteen per cent between sixty-five and eighty-five and

seven per cent of the scores were above eighty-five. This favorable distribution of scores for the rating sheet tends to show how it may be used for discriminating between the playing ability of students presumed to all be very close to the same ability level. The results obtained by use of the objective grading with the rating scale are much more discriminating than those of the subjective judges. This fact alone would suggest the value of a rating scale as adverse to merely subjective observation. (See Figure 4, page 29).

Frequency
12-

11-

10-

9-

8-

7-

6-

5-

4-

3-

2-

1-

4.5%

177%

29%

29%

14%

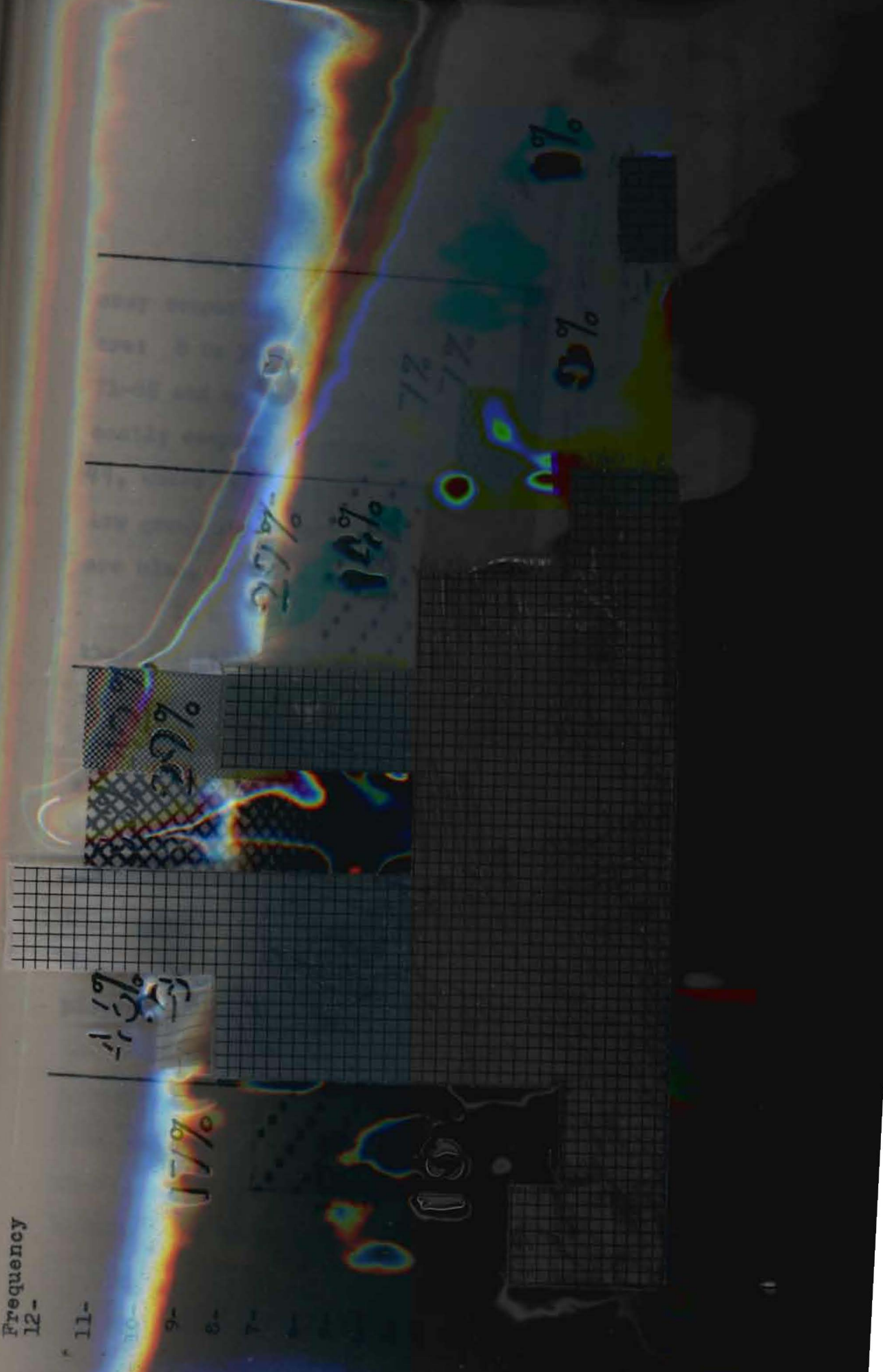
10

7%

7%

3%

0%



frequency

2-

1-

0-

9-

8-

7-



3.10%
3.4%

1.7%

2.9%

0.5%

Frequency

12-

11-

10-

9-

8-

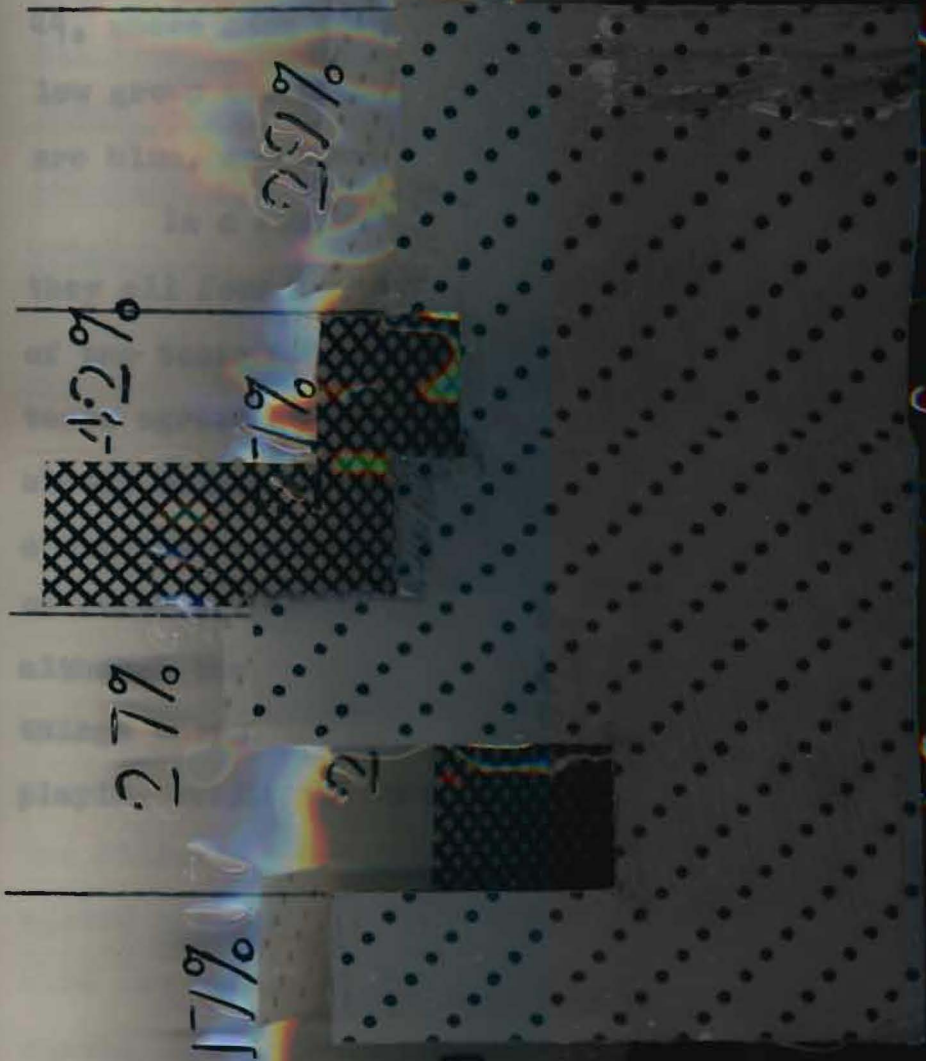
7-

6-

5-

4-

3-



very important
294 8 to
71-85 and
quality group

18

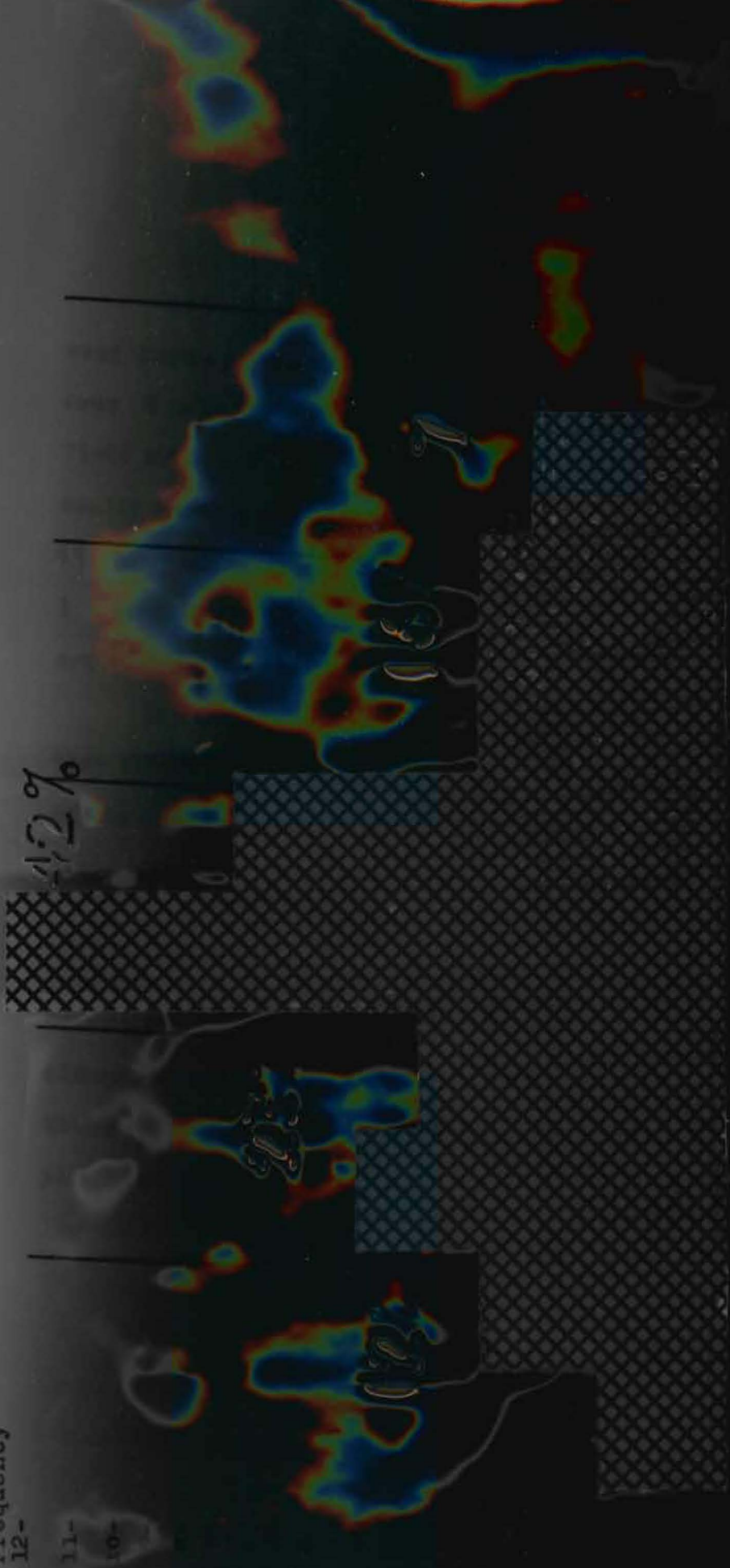
77%

Frequency

12-

11-

10-



Frequency

12-

11-

10-

9-

8-

7-

6-

5-

4-

3-

2-

1-

5 15 25 35 45 55 65 75 85 95 105

Figures 1-4. A comparison of the score distribution for the Revised Dyer Backboard Test, the Miller-Broer Forehand Backhand Test of Tennis Ability, the subjective ratings of the judges, and the rating sheet.

COMPARISON OF TEST RESULTS

The Z scale was divided into three major groups for easy comparison of the test results. The range of the groups are: 0 to 30, indicating below average; 31-70, average; and 71-85 and up, above average. Each girl's test scores may be easily compared by reference to the chart in Appendix B, page 49, where each group is indicated by color. The scores in the low group are shaded yellow, the scores of the middle group are blue, and those of the high group are red.

In a comparison of all four tests it was found that they all four agreed thirty-three per cent of the time. Three of the tests agreed thirty-five per cent of the time and two tests agreed thirty-two per cent of the time. The tests were always within one division of one another, and at no time did a student receive a high rating on one test and a low rating on another. (Table II). These results would indicate that although the tests were not always measuring exactly the same things they were fairly close in their ability to measure the playing ability of the girls as a whole.

TABLE II

COMPARISON OF AGREEMENT OF SCORES:

FOR ALL TESTS

Rating scale

Tests	Number of Subjects	Per Cent of Agreement
All four agree	16	33
Three agree	17	35
Two agree	15	32
Total	48	100

The rating scale was compared with the other three tests in the study and the Revised Dyer Backboard Test of Tennis Ability was found to have the highest correlation of the three. The Dyer test and the rating sheet agreed seventy-three per cent of the time. The rating sheet results were higher than those of the Revised Dyer test thirteen per cent of the time and lower fourteen per cent of the time, which indicates both tests are rating essentially the same things. The Miller-Broer Forehand Backhand Test of Tennis Ability agreed with the rating sheet only sixty-two per cent of the time and tended to rate the students higher than the rating sheet. The subjective ratings of the judges had a low correlation with the rating scale. The results agreed only

forty-six per cent of the time and the results of the judges were higher than the rating scale thirty-one per cent of the time. Statistically, this would indicate the two tests were not measuring the same things. The author feels this is probably true since the rating scale had definite skills to be measured but the three judges just watched the game play and gave a rating. Due to differences in individual opinions of the judges they possibly did not rate exactly the same skills. (Table III)

In a comparison of the two standardized skill tests with the subjective ratings of the judges it was found that neither test had a very high correlation with the judges' results. The Revised Dyer Backboard Test of Tennis Ability agreed only sixty-four per cent of the time and the Miller-Broer Forehand Backhand test agreed only sixty-seven per cent of the time. The Miller-Broer test showed a close ratio between the number of scores higher and lower than the judges, which indicated these two tests measure approximately the same things. However, the subjective ratings of the judges had the lowest overall correlation with all the tests which indicates it was possibly the least valuable of all the tests. (Table IV).

TABLE III

COMPARISON OF THE RESULTS OF THE RATING SCALE
TO THOSE OF THE DYER, MILLER-BROER
AND SUBJECTIVE JUDGES

Tests	Comparison of Scores	Number of Subjects	Per Cent of Agreement
DYER VS. RATING	Number Agree	35	73
	Number Rating Higher	6	13
	Number Rating Lower	7	14
	TOTAL	48	100
MILLER- BROER VS. RATING	Number Agree	30	62
	Number Rating Higher	6	13
	Number Rating Lower	12	25
	TOTAL	48	100
SUBJECTIVE JUDGES VS. RATING	Number Agree	22	46
	Number Rating Higher	11	23
	Number Rating Lower	15	31
	TOTAL	48	100

TABLE IV

COMPARISON OF THE RESULTS OF THE DYER AND MILLER-BROER
TESTS TO THOSE OF THE SUBJECTIVE JUDGES

Tests	Comparison of Scores	Number of Subjects	Per Cent of Agreement
DYER VS. JUDGES	Number Agree	29	61
	Number Judges Higher	12	25
	Number Judges Lower	7	11
	TOTAL	48	100
MILLER- BROER VS. JUDGES	Number Agree	32	67
	Number Judges Higher	7	14
	Number Judges Lower	9	19
	TOTAL	48	100

When the two standardized skill tests were compared they had low correlation and agreed only sixty-four per cent of the time. This indicates they are measuring different things, a fact which Katharine Fox established from her study in December, 1952.³ (Table V)

TABLE V

A COMPARISON OF THE RESULTS OF THE DYER TEST
TO THOSE OF THE MILLER-BROER TEST

Tests	Comparison of Scores	Number of Subjects	Per Cent of Agreement
DYER VS. MILLER- BROER	Number Agree	31	64
	Number Dyer Higher	9	19
	Number Dyer Lower	8	17
	TOTAL	48	100

³Katharine Fox, "A study of the Validity of the Dyer Backboard Test and the Miller Forehand Test for Beginning Tennis Players." The Research Quarterly of the American Association for Health, Physical Education and Recreation, 24:1-8, March, 1953.

CHAPTER V

SUMMARY AND CONCLUSIONS

SUMMARY

I. Statement of Purpose

The purpose of this study was to devise a rating scale that could be used as a practical and efficient method of grading game play of beginning and intermediate tennis players.

II. Background Material

In reviewing the literature, it was found that no studies of this type had been made in tennis, however, rating scales have been used successfully in other areas. The standardized skill tests used were checked for their reliability and validity in predicting tennis skill and both were found to be good tests. The material concerning the construction and use of rating scales was of considerable importance in developing the one used in this study.

III. Procedures of the Study

The experimentation for this study was done at Wichita High School North in the girls' physical education department. The subjects for the pilot test of the rating sheet were taken

from the after school program. These girls were sophomores, juniors and seniors in high school. The rating sheet was tested for its reliability and validity. The girls were all beginners or intermediates and were rated on the rating sheet by three qualified judges. The results of each judge were then compared with the other judges' results on the basis of the average deviation from the mean. It was determined from this pilot study that the rating sheet was reliable and valid.

The actual testing was all done during the regular physical education classes at the end of a four week instruction period for beginning tennis. The forty-eight subjects for this study were beginning or intermediate tennis players and each was given the following tests to determine her playing ability: The Revised Dyer Backboard Test of Tennis Ability; the Miller-Broer Forehand Backhand Test of Tennis Ability; a subjective rating by three judges and a rating on the rating scale. All tests with the exception of the subjective ratings of the judges were given by the individual teachers to their own classes and within one week of each other.

IV. Test Results

The results of the judges in the pilot test of the rating scale were shown to be very close and to have a high reliability. These results were compared to the mean of the

three scores to determine the deviation of each score.

The actual test results were compared by the use of z-scores. These were then plotted on a Z scale and illustrated on bar graphs for easy comparison. (Figures 1-4, page 26). It was found by comparing the distribution of the test scores with that of the Bell curve, the rating scale correlated most favorably and the score distribution for the subjective ratings of the judges were lacking in discrimination.

A comparison of the results for each test to those of the other tests showed them to be in complete agreement thirty-three per cent of the time. The Revised Dyer Backboard Test of Tennis Ability and the rating scale correlated most closely by agreeing on scores seventy-three per cent of the time. This close correlation would suggest that both tests were measuring essentially the same things. The Miller-Broer Forehand Backhand Test of Tennis Ability and the subjective ratings of the judges agreed with the rating scale sixty-two per cent and forty-six per cent, respectively. The two standardized skill tests had low correlation with the subjective ratings of the judges and also with each other.

CONCLUSIONS

On the basis of the data presented in this study the following conclusions appear warranted: (1) the results of the two standardized skill tests show a low correlation of

sixty-four per cent, (2) the correlation of the scores for the two standardized skill tests with those results of the subjective ratings of the judges was insignificant and agreed only sixty-four per cent of the time, (3) The Revised Dyer Backboard Test of Tennis Ability had a significant correlation with the rating scale of seventy-three per cent while the Miller-Broer Forehand Backhand Test of Tennis Ability correlates only sixty-two per cent of the time, (4) the lowest correlation was that between the rating sheet and the ratings of the subjective judges, which was forty-six per cent, (5) the rating scale prepared seems to be both reliable and valid in predicting beginning and intermediate tennis playing ability.

It is the author's opinion that the rating scale prepared for this study would be a valuable teaching aid to fairly evaluate tennis playing ability. The rating scale is simple and easy to use while students are playing tournaments or practice games. There is no need to set aside time and space for special tests and to disrupt the normal class schedule. The results may be posted and students may figure their own scores if the teacher desires. The rating scale also enables the teacher to emphasize what she thinks most important in playing the game of tennis.

As a result of this study the author would like to recommend the following: (1) similar studies on the use of

the prepared rating scale using a larger sample of subjects, (2) similar studies to develop rating scales in other sports areas, (3) a study correlating the results of the prepared rating scale with those of a tennis knowledge test, (4) a study using the prepared rating scale to evaluate the ability of advanced players, (5) the use of the rating scale developed in this study for grading and evaluation of beginning and intermediate tennis players.

1934
of Health
1934-1935

and some
and some
of the
1934-1935

1934-1935
The
1934-1935

Revision of the
1934-1935

of the validity of
the Miller-Foreman
1934-1935

BIBLIOGRAPHY

BIBLIOGRAPHY

- Broer, Marion R. "Are Our Physical Education Grades Fair?", Journal of Health Physical Education Recreation, 30:49, March 1959.
- Broer, Marion R. and Donna Mae Miller. "Achievement Tests for Beginning and Intermediate Tennis," The Research Quarterly of the American Association for Health, Physical Education and Recreation, 21:309-311, October, 1950.
- Dyer, Joanna Thayer, "Revision of the Backboard Test of Tennis Ability," The Research Quarterly of the American Association for Health and Physical Education, 9:29-32, March, 1938.
- Fox, Katharine, "A Study of the Validity of the Dyer Backboard Test and the Miller Forehand Backhand Test for Beginning Tennis Players," The Research Quarterly of the American Association for Health, Physical Education and Recreation, 24:1-8, March, 1953.
- Latchaw, Marjorie and Camille Brown, The Evaluation Process in Health Education, Physical Education and Recreation. Englewood Cliffs: Prentice-Hall, Inc., 1961.
- McCloy, Charles Harold, Tests and Measurements in Health and Physical Education. New York: P.S. Crofts and Co., 1942.
- Myers, Carlton R. and T. Erwin Blesh, Measurement of Physical Education. New York: P.S. Crofts and Co., 1962.
- Scott, Gladys M. and Esther French, Measurement and Evaluation in Physical Education. Dubuque: Wm. C. Brown Company Publishers, 1959.
- Smithells, Phillip A. and Peter E. Cameron, Principles of Evaluation. New York: McGraw-Hill Book Company, Inc., 1961.
- Tiegs, Ernest, Tests and Measurements in Improvement of Learning. Boston: Houghton Mifflin Co., 1934.
- Willgoose, Carl E., Evaluation in Health Education and Physical Education. New York: McGraw-Hill Book Company, Inc., 1961.

TW
TEST

COPO =	(Serr- Kann)	(Serr- Kann)
491	201.5X	2
	272.53	289
	209.47	
	89.87	
	74.93	
	82.95	
	131.49	
	38.83	
	20.07	
	12.11	
	51.00	

12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

APPENDIX

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

APPENDIX A

RESULTS OF THE SCORES OF THE DYER REVISED
TEST OF TENNIS ABILITY

Raw Score	Frequency	Score- Mean	(Score- Mean) ²	(Score- Mean) ² ~	(Score- Mean) ² x20	Z Score
10	2	-17.48	305.55	1.91	38.2	11.8
11	1	-16.48	271.59	1.80	36.0	14.0
13	2	-14.48	209.67	1.58	31.6	18.4
18	3	- 9.48	89.87	1.04	20.8	29.2
19	2	- 8.48	71.91	.93	18.6	31.4
20	4	- 7.48	55.95	.82	16.4	33.6
21	1	- 6.48	41.99	.71	14.2	35.8
22	1	- 5.48	30.03	.60	12.0	38.0
23	1	- 4.48	20.07	.49	9.8	40.2
24	3	- 3.48	12.11	.38	7.6	42.4
25	6	- 2.48	61.50	.27	5.4	44.6
27	1	- .48	.23	.05	1.0	49.0
28	1	+ .52	.27	.06	1.2	51.2
29	3	+ 1.52	2.31	1.17	3.4	53.4
30	1	+ 2.52	6.35	1.28	5.6	55.6
31	4	+ 3.52	12.39	.38	7.6	57.6
32	1	+ 4.52	20.43	.49	9.8	59.8
33	1	+ 5.52	30.47	.60	12.0	62.0
34	2	+ 6.52	42.51	.71	14.2	64.2
35	3	+ 7.52	56.55	.82	16.4	66.4
36	1	+ 8.52	72.59	.93	18.6	68.6
37	1	+ 9.52	90.63	1.04	20.2	70.2
39	1	+11.52	132.71	1.26	25.2	75.2
42	1	+14.52	210.83	1.59	31.8	81.8
54	1	+26.52	703.31	2.90	58.0	108.0

APPENDIX A

RESULTS OF THE MILLER-BRGER FOREHAND, BACKHAND
TEST OF TENNIS ABILITY

Raw Score	Frequency	Score- Mean	(Score- Mean) ²	(Score- Mean) ²	(Score- Mean) ² x20	Z Score
6	1	-40.60	1648.36	1.68	37.6	12.4
10	1	-36.60	1339.56	1.70	34.0	16.0
12	1	-34.60	1197.16	1.61	32.2	17.8
20	1	-26.60	707.56	1.23	24.6	25.4
25	1	-21.60	466.56	1.00	20.0	30.0
26	2	-20.60	424.36	.96	19.2	30.8
27	3	-19.60	384.16	.91	18.2	31.8
28	1	-18.60	345.96	.86	17.2	32.8
30	1	-16.60	275.56	.77	15.4	34.6
31	1	-15.60	243.36	.72	14.4	35.6
36	1	-10.60	112.36	.49	9.8	40.2
37	2	- 9.60	92.16	.45	9.0	41.0
38	4	- 8.60	73.96	.40	8.0	42.0
45	1	- 1.60	2.56	.07	1.4	48.6
47	1	+ .40	.16	.00	0.0	50.0
48	2	+ 1.40	1.96	.06	1.2	51.2
53	1	+ 6.40	40.96	.30	6.0	56.0
54	2	+ 7.40	54.76	.34	6.8	56.8
55	1	+ 8.40	70.56	.39	7.8	57.8
56	2	+ 9.40	88.36	.44	8.8	58.8
57	1	+10.40	108.16	.48	9.6	59.6
58	2	+11.40	129.96	.53	10.6	60.6
59	1	+12.40	153.76	.58	11.6	61.6
60	1	+13.40	179.56	.62	12.4	62.4
62	1	+15.40	237.16	.71	14.2	64.2
63	3	+16.40	268.96	.76	15.2	65.2
65	1	+18.40	338.56	.85	17.0	67.0
67	1	+20.40	416.16	.95	19.0	69.0
74	1	+27.40	750.76	1.27	25.4	75.4
75	1	+28.40	806.56	1.32	26.4	76.4
76	1	+29.40	864.36	1.36	27.2	77.2
84	1	+37.40	1398.76	1.74	34.8	84.8
86	1	+39.40	1552.36	1.83	36.6	86.6
94	1	+47.40	2246.76	2.20	44.0	94.0
00	1	+53.40	2851.56	2.48	49.6	99.6

APPENDIX A

RESULTS OF THE SCORING OF THE
JUDGES' SUBJECTIVE RATING

Raw Score	Frequency	Score- Mean	(Score- Mean) ²	(Score- Mean) ²	(Score- Mean) ² x20	Z Score
3.66	1	-2.78	7.73	1.72	34.4	15.6
4	4	-2.44	5.95	1.51	30.2	19.8
4.33	3	-2.11	4.45	1.30	26.0	24.
4.66	2	-1.78	3.17	1.10	22.0	28.
5	2	-1.44	2.07	.89	17.8	32.2
5.33	1	-1.11	1.23	.69	13.8	36.2
5.66	6	-.78	.61	.48	9.6	40.4
6	2	-.44	.19	.27	5.4	44.6
6.33	4	-.11	.01	.07	1.4	48.6
6.66	3	+.22	.48	.14	2.8	52.8
7	4	+.56	.31	.35	7.0	57.0
7.33	1	+.89	.79	.55	11.0	61.0
7.66	1	+1.22	1.49	.75	15.0	65.0
8	5	+1.56	2.43	.96	19.2	69.2
8.33	2	+1.89	3.57	1.17	23.4	73.4
8.66	3	+2.22	4.93	1.37	26.4	76.4
9	4	+2.56	6.55	1.58	31.6	81.6

APPENDIX A

RESULTS OF THE SCORES OF THE RATING SHEET

Raw Score	Frequency	Score - Mean	(Score - Mean) ²	(Score - Mean) ²	(Score - Mean) ² x 20	Z Score
3.15	2	-2.48	6.15	1.77	35.4	14.6
3.55	2	-2.08	4.33	1.48	29.6	20.4
3.90	2	-1.73	2.99	1.23	24.6	25.4
3.95	3	-1.68	2.82	1.20	24.0	26.
4.00	1	-1.63	2.65	1.16	23.2	26.8
4.35	1	-1.28	1.64	.98	18.2	31.8
4.45	1	-1.18	1.39	.84	16.8	33.2
4.75	2	-.88	.77	.63	12.6	37.4
4.85	1	-.78	.61	.56	11.2	38.8
5.00	1	-.63	.40	.45	9.0	41.0
5.20	1	-.43	.18	.31	6.3	43.8
5.30	2	-.33	.11	.24	4.8	45.2
5.35	2	-.28	.08	.20	4.0	46.0
5.45	2	-.18	.03	.13	2.6	47.4
5.50	1	-.13	.02	.09	1.8	48.2
5.60	1	-.03	.00	.02	0.4	49.6
5.65	1	+.02	.00	.01	0.2	50.2
5.75	1	+.12	.04	.09	1.8	51.8
5.90	2	+.27	.07	.19	3.8	53.8
6.10	2	+.47	.22	.34	6.8	56.8
6.15	1	+.52	.27	.37	7.4	57.4
6.20	1	+.57	.32	.41	8.2	58.2
6.25	1	+.62	.38	.44	8.8	58.8
6.35	1	+.72	.52	.51	10.2	60.2
6.45	1	+.82	.67	.59	11.8	61.8
6.55	1	+.92	.85	.66	13.2	63.2
6.75	1	+1.12	1.25	.80	16.0	66.
6.80	2	+1.17	1.37	.84	16.8	66.8
6.95	1	+1.32	1.74	.94	18.8	68.8
7.60	1	+1.97	3.88	1.41	28.2	78.2
7.85	1	+2.22	4.93	1.59	31.8	81.8
7.90	1	+2.27	5.15	1.62	32.4	82.4
8.05	1	+2.42	5.86	1.73	34.6	84.6
8.10	1	+2.47	6.10	1.76	35.2	85.2
8.30	2	+2.67	7.13	1.90	38.0	88.0

APPENDIX B

COMPARISON OF RAW SCORES AND SCORES ON ALL
TESTS FOR THE FORTY-EIGHT SUBJECTS₁

Students	Miller-Broer		Dyer Revised		Judges' Average		Rating Sheet		Agreement of Test Scores
	Raw		Raw		Raw		Raw		
Carolyn Arnold	30	34.6	13	18.4	6	44.6	3.55	20.4	2
Rosemary Balzer	26	30.8	30	55.6	5.66	40.4	6.20	58.2	4
Barbara Baty	63	65.2	20	33.6	7	57.0	5.60	49.6	4
Susie Bauman	94	94.0	25	44.6	8	69.2	6.55	63.2	3
Susan Bowen	37	41.0	42	81.8	6	44.6	8.30	88.0	2
Diane Butcher	58	60.6	39	75.2	7	57.0	8.10	85.2	2
Suzanne Butcher	38	42.0	25	44.6	8.66	76.4	6.80	66.8	3
Ruth Carpenter	54	56.8	13	18.4	4	19.8	6.25	58.8	2
Charlotte Cooley	55	57.8	27	49.0	5	32.2	4.45	33.2	4
Karen Cupit	53	56.0	24	42.4	5.44	24.0	5.65	50.2	3
Lynn Deepke	63	65.2	31	57.6	8	69.2	7.85	81.8	3
Karen Dowler	84	84.8	25	44.6	9.33	81.2	5.50	48.2	2
Linda Ellis	100	99.6	35	66.4	9	81.2	5.30	45.2	2
Marion Ford	36	40.2	23	40.2	7.33	69.2	5.45	47.4	4
Carol Frederick	29	31.8	25	44.6	7	69.2	8.05	84.6	3
Sheliah Garrett	31	36.6	22	38.0	5.66	32.2	3.90	25.4	4
Beverly Gilden	28	32.8	25	44.6	6.33	57.0	4.35	31.8	4
Marilyn Goodgion	45	48.6	29	53.4	5	32.2	3.95	26.0	3
Ruth Hill	10	16.0	10	11.8	4.66	32.2	3.15	4.6	3
Janet Heywood	20	25.4	18	29.2	5.44	44.6	3.55	20.4	3
Gloria Hoyle	37	41.0	18	29.2	4.33	19.8	5.00	41.0	2
Paula Hunt	38	42.0	20	33.6	4	19.8	3.95	26.0	2
Donna Johnson	76	77.2	29	53.4	6.33	44.6	6.10	56.8	3
Shirley Kuhns	75	76.4	54	108.0	8.33	73.4	7.60	78.2	4
Vicki Laurentz	38	42.0	24	42.4	6.66	52.8	6.15	57.4	4
Linda Lepley	62	64.0	31	57.6	6.66	52.8	6.95	68.8	4

APPENDIX B (continued)

Students	Miller-Broer		Dyer Revised		Judges' Average		Rating Sheet		Agreement of Test Scores
	Raw		Raw		Raw		Raw		
Shirley Lindstrom	27	31.8	31.	57.6	8.33	73.4	6.45	61.8	3
Connie Lyon	58.	60.6	35	66.4	7.66	65.0	4.75	37.4	4
Dorothy McGowan	74	75.4	29	53.4	8	69.2	7.90	62.4	2
Beth Maddox	56	58.8	11	14.0	4.66	28.0	5.90	53.8	2
Janice Millspaugh	27	31.8	20	33.6	5.66	40.4	5.35	46.0	4
Teressa Mitchel	57	59.6	32	59.8	8	69.2	5.20	43.8	4
Cecelia Moceno	65	57.0	21	35.8	5.66	40.4	3.95	26.0	4
Pat Morton	26	30.8	18	29.2	4	19.8	4.75	37.4	2
Pat Myers	60	62.4	36	68.6	7	57.0	6.80	66.8	4
Susan Nagel	54	56.8	37	70.2	8	69.2	6.35	60.2	4
Angelina Ornelas	12	17.8	19	31.4	4	19.8	5.90	53.8	2
Annette Peters	6	12.4	19	31.4	3.66	15.6	3.90	25.4	3
Diana Phipps	56	58.8	33	62.0	8.66	76.4	6.10	56.8	3
Sharon Presley	25	30.0	20	33.6	6.33	48.6	3.15	14.6	2
Beverly Robl	59	61.6	35	66.4	9	61.6	6.75	66.0	3
Anna Rudrow	47	50.0	34	64.2	9	61.6	8.30	68.0	2
Marilyn Smith	53	65.2	25	44.6	5.66	40.4	5.45	47.4	4
Jacque Wells	38	42.0	24	42.4	6.33	48.6	4.85	38.8	4
Donna Wheeler	67	69.0	31	57.6	5.66	40.4	5.35	46.0	4
Sandy Winser	48	51.2	28	51.2	6.66	52.8	5.30	45.2	4
Nancy Woodson	48	51.2	10	11.8	5.33	36.2	4.00	26.8	2
Opal Wootton	66	66.6	34	64.2	8.66	76.4	5.75	51.8	2

- ¹Red - Indicates above average range (any score above 71)
 Blue - Indicates average range (30-70)
 Yellow- Indicates below average range (any score below 30)

APPENDIX C

REVISED DIRECTIONS FOR DYER BACKBOARD
TEST OF TENNIS ABILITY

- Equipment. - 1. Backboard or wall, approximately ten feet in height and allowing about fifteen feet in width per person taking the test at one time. Two players taking the test at once has been found to be a very satisfactory arrangement. This allows for adequate supervision by the administrator.
2. On this wall a plainly visible line three inches in width, to represent the net, should be drawn so that the top is three feet from the ground.
3. A restraining line, five feet from the base of the wall, should be drawn on the floor.
4. Stop watch with second hand.
5. Two balls and a racquet per player. It is desirable that the balls be in good condition, although it is not essential that they be exactly new. The racquet should be without flaws.
6. Box for extra balls, about 12 inches long, 9 inches wide and 3 inches deep, placed on the floor where the restraining line joins the side at the left for right handed players and right for left handed players.
7. One pencil for group of four players.
8. Score card per player.

Organization. - Divide the group to be tested into units of four players each, and number them from one to four. Provide each player with a score card on which she writes her name. Then read the following description of the test to the group.

"The Backboard Test consists in rallying a tennis ball against the wall. The object of the test is to cause the ball to strike the wall on or above the net line as many times as you can in 30 seconds. (Pause) When I say 'Go!' start the test immediately. Drop the ball and let it hit the floor once, then put it in play against the wall. Continue to play it to the wall until I say, 'Stop!' at the end of 30 seconds. There is no limit to the number of times the ball may bounce before you hit it. You may volley the ball. The ball need not touch the floor before you play it except at the start and when a new ball is being put in play. You may use any stroke or combination of strokes. You must play all balls from behind this restraining line (indicate the line clearly). You may cross the line to retrieve balls, but any hits made while in such a position do not count. You may use any number of balls. If for any reason you lose control of the ball in play, do not try to retrieve it. Take another ball from this box (indicate clearly) and put it in play as you did at the start. Each ball striking the wall on or above the net line before the word 'Stop' counts as a hit and scores one point.

You will each be given three trials today. The final score on the test is the sum of the scores on the three trials."

Demonstrate the following points:

1. Two balls in hand.
2. Start test by dropping ball, letting it hit floor at least once, then play it.
3. Rally a few times, showing volley.
4. Cross restraining line to retrieve a ball, a low hit to keep it in play and retreat for next shot.
5. Make a wild shot to show how taking another ball saves time. Put this new ball in play as at the start.

Read the following paragraph, making certain that each person understands the test procedure and her duties.

"In each group:

"No. 1 takes the test. At the signal, 'Ready?' she stands anywhere behind the restraining line with her racquet and two balls prepared to start the test at the word 'Go!'

"No. 2 counts the number of balls which strike the wall on or above the net line before the word 'Stop!' and enters them on the score card opposite the appropriate trial number. If any infringements are reported by No. 3 these are deducted before the score for the trial is recorded. A ball striking coincident with the word 'Stop!' does not count.

No. 3 watches the player in relation to the restraining line. She reports to the scorer at the end of the trial the

number of hits, if any, made while the player was standing closer to the wall than the restraining line.

"No. 4 collects the balls of her group before the start of a trial and puts them in the box. During the trial she collects and returns to the box any balls going out of play.

"Each person takes the test in rotation. After No. 1 has had her first trials she assumes the duties of No. 2 while the latter takes the test; No. 3 and No. 4 remain the same. While No. 3 takes the test, No. 4 scores the hits. No. 1 and No. 2 assume the duties of No. 3 and No. 4 respectively. When No. 4 takes the test, No. 3 scores hits, and No. 1 and No. 2 remain the same. After each person in the entire group being tested has had one trial, the test is repeated in the same order until everyone has had three trials in all."

Answer questions. This organization will consume about ten minutes. Great care should be exercised in these preliminaries to make certain that the test procedure is clearly understood. The testing will then take place smoothly and accurately.

The examiner then assumes a position to the rear of the players with the stop watch, and begins testing the No. 1's who are to take the test at one time, usually one or two. Numbers 2, 3 and 4 of these groups will follow, and then No. 1 of the next two groups, and so on until all have had one trial, after which the test is repeated twice in the same order.

APPENDIX C

DIRECTIONS FOR THE MILLER-BROER

FOREHAND BACKHAND TEST

A. Equipment

1. One regulation court.
2. One regulation net with a rope stretched 4 feet above the top of the net. Note: The most effective drives are those that are hit with a good deal of force into the backcourt. This restraining rope is a device to measure, to a degree, the force of the drive. A ball passing between the net and this restraining rope and landing in the backcourt area must have been hit with more force than a ball going high (over the rope) and landing in the same area. Driver constructed a tennis test which made use of the restraining rope placed 7 feet above the top of the net.⁴ However, during experimentation it was observed that a restraining rope placed that high did not discriminate between players of varying ability; i.e., it was possible for a player to hit the

⁴Helen Irene Driver, Tennis for Teachers, Phila., Penn.: W. B. Saunders Co., 1941, p. 163.

balls slowly and with little force 7 feet high and have them hit in the backcourt and therefore score as high as a player who hit fast low drives.

3. One racket and 15-20 balls in good condition.
4. Score sheets for each player (see Figure I) and pencils.
5. Special court markings (see Figure II).
 - a. Two chalk lines drawn across the court 10 feet inside the service line and 9 feet outside the service line and parallel to it.
 - b. Two chalk lines drawn across the court 5 feet and 10 feet respectively outside the baseline and parallel to it.
 - c. Chalked numbers in the center of each area to indicate its scoring value.

B. Test

1. The player taking the test stands behind the baseline, bounces the balls to herself, hits the balls and attempts to place them in the back 9 feet of the opposite court.
2. Each player is allowed fourteen trials on the forehand and fourteen trials on the backhand.
3. In order to score the values as shown on Figure II,

balls must go between the top of the net and the rope and land in the designated area or on lines bounding the area (balls landing on a line receive the highest score for that area.)

4. Balls which go over the rope score one-half the value of that area in which they land.
5. If the player misses the ball in attempting to strike it, it is considered a trial.
6. Let balls are taken over.

C. Scoring²⁵

1. The number of each trial is marked on the score card diagram in the same relative position as the ball landed on the court. (See Figure I).
2. Each ball hit is scored 2-4-6-8-6-4-2-, depending upon the area in which it lands. Note: Each ball going over the rope is scored one-half the value of the area in which it lands (this may be indicated by circling the ball on the scoring diagram).

²⁵Since the test was designed to measure the ability to place the drives in the back court, it was necessary to give some scoring value to those areas which are just outside the desired area. Trials which hit a few feet beyond the back 9 feet of the court should not be penalized any more than those which hit a few feet short—even though those which are short hit within the court.

3. The total score equals the sum of fourteen balls on the forehand and fourteen balls on the backhand.²

²Marion R. Broer and Donna Mae Miller, "Achievement Tests for Beginning and Intermediate Tennis," The Research Quarterly of the American Association for Health, Physical Education and Recreation, Vol. 21-No. 3 (October, 1950.) pp. 309-311.

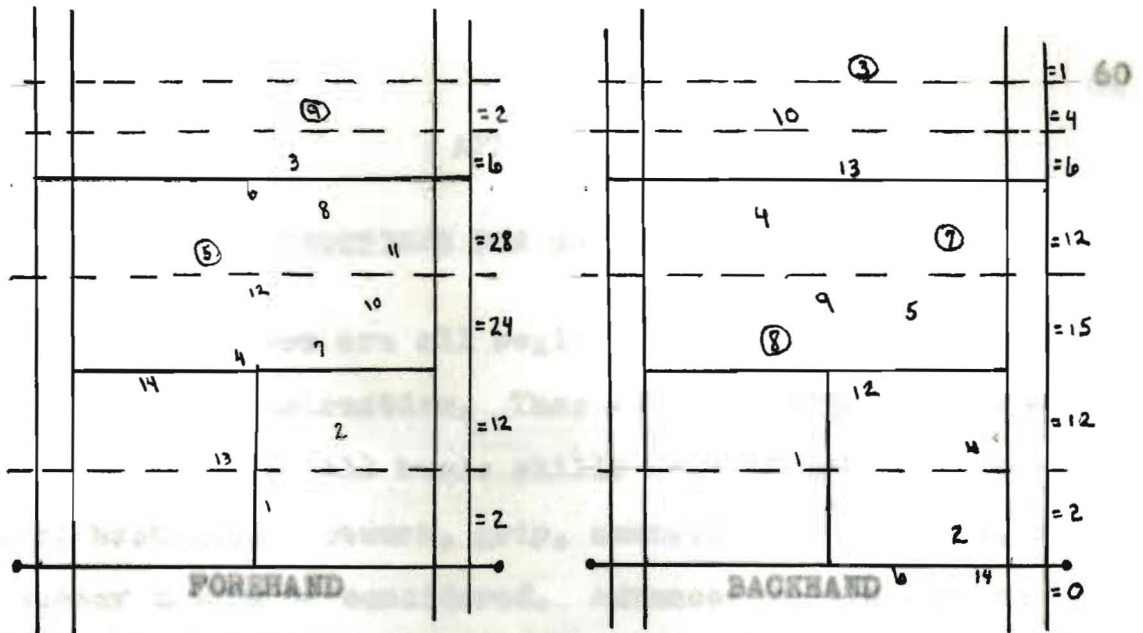


Figure 5. Completed score card for forehand-backhand drives

The trial number is marked in the same relative position on the score card as the ball lands on the court. The circle around the number indicates that the ball went over the restraining rope and therefore receives only half the scoring value of the area in which it hits.³

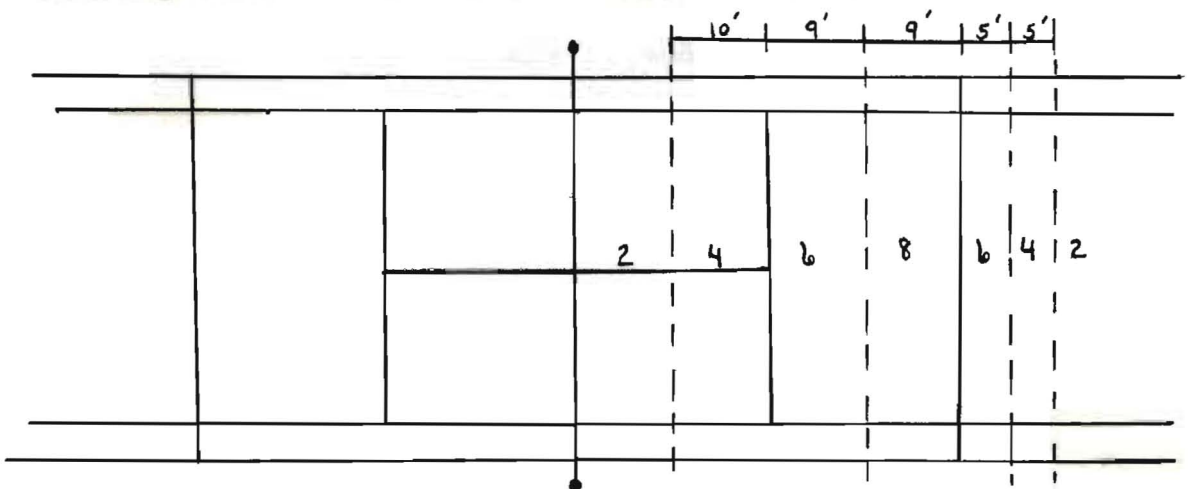


Figure 6. Special court markings for forehand-backhand drive test

³Ibid., p. 310.

APPENDIX C

INSTRUCTIONS FOR SUBJECTIVE JUDGES

The players are all beginners in tennis and have had four weeks of instruction. They will be judged for general playing ability. All basic skills such as service, forehand, backhand, footwork, grip, contact with the ball, and accuracy should be considered. Advanced skills such as placement, strategy, lob, smash, etc., should not be considered.

Each player should be given a score between 0-10. The following scale is the basis for rating:

- 0 - 3 poor
- 4 - 7 fair
- 8 - 10 good

APPENDIX D

TENNIS CHECK SHEET

	NAME			
GRIP				
heel of hand on butt of racket				
FOOTWORK				
returns to home position				
weight shift				
STROKING				
side to net when stroking				
distance from ball when stroking				
use of correct stroke (forehand-backhand)				
in bounds				
SERVICE				
toss				
follow through				
point of contact on ball				
accuracy				

- ✓ **GOOD** - above average. Executes the skill correctly with good form and thinks quickly. Well coordinated.
- **FAIR** - Average, does the skill but may not have good form or balance. Gets the job done, not necessarily accurate or fast. Coordination is lacking.
- **POOR** - Below average or failed. Does not execute the skill at all.

SCORING -
$$\frac{2\checkmark + - (10)}{\text{total of all marks}} = \text{Score}$$