Summer 1962

An Analysis of the McDonald Soccer Skill Test as Applied to Junior High School Girls

Bonnie Streck
Fort Hays Kansas State College

AN ANALYSIS OF THE McDONALD SOCCER SKILL TEST AS APPLIED TO JUNIOR HIGH SCHOOL GIRLS

being

A Master's Report Presented to the Graduate Faculty of the Fort Hays Kansas State College in Partial Fulfillment of the Requirements for the Degree of Master of Science

by

Bonnie Streck, B.S.
Fort Hays Kansas State College

Date July 30, 1962 Approved Walter Kroll
Major Professor

Approved Ralph A. Loder
Chairman, Graduate Council

ACKNOWLEDGMENT

The author wishes to acknowledge the assistance of all those whose interest, time and effort made this study possible.

The gratitude of the author is respectfully accorded Dr. Walter Kroll, major advisor, for his guidance, understanding and helpful criticism which have aided materially in the successful completion of this study.

AN ANALYSIS OF THE McDONALD SOCCER SKILL TEST AS APPLIED TO JUNIOR HIGH SCHOOL GIRLS

(Abstract)

by Bonnie Streck

PURPOSE: The major purposes of this study were to (1) study the reliability of the McDonald Soccer Skill Test, (2) set up norms at Brooks Junior High School for the McDonald Soccer Skill Test, and (3) set up a grading scale that would be used after the completion of the soccer unit for the girls at Brooks Junior High School.

RESEARCH METHODS: The subjects used in this study were 265 girls, grades 7-8-9, at Brooks Junior High School who were enrolled in physical education classes during the school year 1961-1962. The McDonald Soccer Skill Test was administered to all subjects after four weeks of soccer participation in class and was then repeated the following week to obtain two sets of raw scores.

The mean, standard deviation, chi square, standard error, linear product moment correlation, paired comparison analysis and standard scores were the statistical measures used on the collected data.

RESULTS: The means of all the trials did improve after each successive trial. Very little difference was noted when

using the average of four trials or the best of three trials. The seventh graders had the lowest mean scores while the ninth graders had the highest mean scores. Each grade improved on test II. The correlation between the two tests was .6248 on the four trials and .5715 on the best of three trials.

The values of t obtained through a comparison of the means of test I, best of three trials to test I, four trials; test II, best of three trials to test II, four trials; test I, best of three trials to test II, best of three trials; and test I, four trials to test II, four trials were significant at the .01 level of confidence. This would seem to indicate that the difference obtained was not due to chance factors. Therefore, it is possible that this difference occurred as a result of the skill practice from taking the test the first time.

The chi square test for normality exceeded the value necessary for significance at the .05 level, indicating either a normal distribution of soccer ability was not present in this sample or that the test was not able to discriminate between different levels of soccer ability.

CONCLUSIONS: The conclusions that have been derived from this study are:

1. All differences obtained were significant when

using the average of the four trials or the best of three trials.

2. The reliability correlation coefficient between test I, four trials and test II, four trials was .6248 and on test I, best of three trials and test II, best of three trials was .5715 and both are low for a reliable test.
3. All of the mean scores improved which could indicate the McDonald Test could be used for the improvement of volleying skill in soccer during the instructional period.
4. The letter grade scales did not show a large range of scores, as in most cases only 2 or 3 kicks would result in a grade change.
5. There are two conclusions possible as a result of the computation of the chi square test. They are: (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill present in the sample of Brooks Junior High School girls was not distributed normally. It would seem logical that the sample obtained was large enough and widely varied enough, according to ability, to obtain a normal distribution of test scores for soccer

ability. Therefore, the data indicates that the test did not discriminate between different levels of soccer ability.

6. The reliability of the McDonald Soccer Skill Test was low as shown in this study and would therefore limit the validity of this test for use with the Brooks Junior High School girls.

7. The ^{T}-scores that were prepared from this study must be used with some caution and only under certain conditions due to the findings of the significant chi square values which showed a departure from the normal curve.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Purpose	1
	Significance of Study	1
	Need for Study	2
	Procedures	3
	Delimitations	4
	Limitations	4
	Organization of Remainder of Report	5
	Summary	5
II.	REVIEW OF RELATED LITERATURE	7
	Introduction	7
	History of Skill Tests	7
	Purposes of a Skill Test	11
	Criteria for Skill Tests	12
	Practical Considerations in Applying Criteria	13
	Available Soccer Skill Tests	13
	Selection of a Soccer Skill Test	17
	Basis for Selection of Test	18
	Summary	18
III.	METHODOLOGY	22
	Introduction	22

CHAPTER PAGE

Sampling 22
Testing Procedures 22
Score Card Data 23
Description of Test 24
Analysis Plan 25
Summary 25
IV. RESULTS AND DISCUSSION 28
Introduction 28
Summary of Raw Score Data 28
Reliability 31
Changes in Test Scores 34
Test for Normality 35
Norms 37
Interpretation of Results 38
Summary 42
V. SUMMARY AND CONCLUSIONS 50
Introduction 50
Purpose 50
Significance 50
Delimitations 51
Limitations 51
History 52
Purposes and Criteria for Skill Tests 52
Selection of Test 53

CHAPTER PAGE

Procedures 54

Results 55

Conclusions 61

BIBLIOGRAPHY 64

APPENDIX 67

APPENDIX A. McDonald Soccer Skill Test and Floor Plan . 68

APPENDIX B. Letter Grades for Brooks Junior High School on the McDonald Soccer Skill Test 72

APPENDIX C. Frequency Polygons for Combined 7-8-9 Grades and Grades 7-8-9 Individually, Test II Best of Three Trials 73

APPENDIX D. Score Card for the McDonald Soccer Skill Test as Administered at Brooks Junior High School 75

LIST OF TABLES

TABLE PAGE

I. Summary of McDonald Soccer Skill Test (Soccer Kicks per 30 Seconds) Administered at Brooks Junior High School, Wichita, Kansas, 1961-1962, Combined 7-8-9 Grades 29

II. Summary of McDonald Soccer Skill Test (Soccer Kicks per 30 Seconds) Administered at Brooks Junior High School, Wichita, Kansas, 1961-1962, Grades 7-8-9, Listed Separately 30

III. Frequency Distribution of Raw Scores for Brooks Junior High School Sample, Test I, Best of Three Trials and Four Trials 69

IV. Frequency Distribution of Raw Scores for Brooks Junior High School Sample, Test II, Best of Three Trials and Four Trials 70

V. T-Scores for Combined 7-8-9 Grades and each Grade Individually for Brooks Junior High School Girls on the McDonald Soccer Skill Test by Cumulative Frequency Method 71

LIST OF FIGURES

FIGURE PAGE

1. Frequency Polygon for Combined 7-8-9 Grades on the McDonald Soccer Skill Test (Soccer Kicks in 30 Seconds) 73

2. Frequency Polygon for Grades 7-8-9 Individually on Test II, Best of Three Trials on the McDonald Soccer Skill Test (Soccer Kicks in 30 Seconds) 74

CHAPTER I

INTRODUCTION

In Chapter I the following areas will be discussed: (1) the major purposes of this study, (2) the significance of this study, (3) need for study, (4) the procedures followed, (5) delimitations, (6) limitations, and (7) the organization of the remainder of this thesis.

Purpose. The major purposes of this study were to (1) study the reliability of the McDonald Soccer Skill Test, (2) set up norms at Brooks Junior High School for the McDonald Soccer Skill Test, and (3) set up a grading scale that could be used after the soccer unit for girls at Brooks Junior High School. The basis for selecting the McDonald Test is discussed in Chapter II.

Significance of Study. Physical education today is to be considered an important subject in the school curriculum. In fact it has now taken on added emphasis as it is graded A,B,C just like all other subjects in the curriculum. No longer are students of physical education willing to let their work be considered as just "exercise". Physical education now has the problem of discovering laws, principles and fundamentals upon which to build such a science. To do this educators must develop experimental conditions, research

minds and ideas, and make full use of testing in day-to-day teaching.[1]

Two of the major complaints by educators today of the physical education program are, (1) Are we really producing a skill or just playing the game? (2) The grading system in physical education is not based upon the results of the skill of the student, but is usually a personality and social adjustment grade.[2]

To actually arrive at a reliable grade after a unit of a particular sport, a skill test must be given and it must have acceptable reliability and validity as a test of this particular skill. The significance of this investigation was to study if the McDonald Soccer Skill Test would have sufficient reliability to be given to junior high school girls and if a grade may be obtained from this particular test to be used for a skill grade in soccer.

Need for Study. Much of the significant progress in physical education during the past three decades has been due partly to the interest of teachers toward problems of a

---

[1]John F. Bovard, Frederick W. Cozens and E. Patricia Hagman, Tests & Measurements in Physical Education (Philadelphia: W. B. Saunders Company, 1949), p. iii.

[2]Mary J. Moriarty, "How Shall We Grade Them?", Journal of Health, Physical Education and Recreation, 25:27, January, 1954.

scientific nature in the field of physical education. The idea that much more progress can be made by increased knowledge of scientific procedures has made advances through the entire professional atmosphere. The most essential procedure to continued professional growth is in the area of measurement. Only as the individual teacher in the field increases his use and knowledge of available tests, his skill in applying them and his ability to evaluate results in view of the limitations of the tests used will any advancement be made in the area of measurement in physical education.[3]

In view of the above statements of Bovard, Cozens and Hagman, the need for this study was to satisfy a problem of general interest to many teachers of physical education. A valid and reliable test of soccer ability that takes a small amount of time to administer and yet has valid results for purposes of grading would be of value in many physical education programs.

Procedures. The McDonald Soccer Skill Test was administered to 265 Junior High School Girls enrolled in physical education classes, during the 1961-1962 school year, after the completion of a four weeks unit of study in the

---

[3]Bovard, Cozens, and Hagman, op. cit., p. 14.

game of soccer. The same test was then repeated one week later, and given in the same manner. Each class was divided into four testing areas and all four areas took the test at the same time.

In taking the test, each student waited approximately five minutes between each kicking trial of the test, in an attempt to eliminate any fatigue factor, rather than taking all four trials in succession.

Delimitations. This study was delimited to the statistical analysis of the one item of wall volley as prescribed by the McDonald Soccer Skill Test. The extent to which this skill test measures the skill of volleying in soccer was thereby delimited to the validity of the McDonald test to measure the skill of volleying in soccer.

Limitations. The present study is limited to the data collected by administering the McDonald Soccer Skill Test to 265 Junior High School Girls enrolled in physical education class, during the school year, 1961-1962 at Brooks Junior High School. The class enrollment at Brooks is such that some girls come to class two days per week, some three days per week and others five days per week. At no grade level is this consistent but is determined by the size of classes and the students' enrollment in other classes on alternate days of the week. This sample may or may not be representative

of the skill of junior high school girls over a period of time.

Organization of Remainder of Report. Chapter II is related literature, Chapter III is methodology, Chapter IV is results and discussion and Chapter V is the summary and conclusions.

Summary. The purposes of this study were to (1) determine the reliability of the McDonald Soccer Skill Test, (2) set up norms at Brooks Junior High School for the McDonald Soccer Skill Test, and (3) set up a grading scale for the girls at Brooks Junior High School that could be used after the unit on soccer.

The significance of this paper was to study the reliability of the McDonald Soccer Skill Test to be given to junior high school girls after the completion of the unit on soccer. As the physical educators today are faced with a challenging situation of saying that physical education is an important subject like any other in the curriculum, it must be graded like the others and full use of measurement must be made in day-to-day teaching. If a reliable grade is to be arrived at after a unit of a particular sport, a skill test must be given and it must have acceptable reliability and validity as a test of this particular skill.

The need for this study is to find a test for soccer

skill which does not take a great deal of time to administer and yet yields valid and reliable results.

The McDonald Soccer Skill Test was given after four weeks of soccer participation in class. The same test was then repeated one week later. Each class was divided into four testing areas and all four areas took the test at the same time.

This study was delimited to the statistical analysis of the one item of the wall volley. The extent to which this skill test measures the skill of volleying in soccer was delimited to the validity of the McDonald test to measure the skill of volleying in soccer.

This study is limited to the data collected by administering the McDonald Soccer Skill Test to 265 junior high school girls enrolled in physical education classes, during the school year 1961-1962, at Brooks Junior High School in Wichita, Kansas. The class enrollment at Brooks varies for physical education with some girls attending class two days per week, others three days, and others five days per week depending on their other class enrollment. This sample may or may not be representative of the skill of junior high school girls over a period of time.

CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction. The purpose of Chapter II is to give a review of the literature that was used in this study. The chapter will be divided into the following areas: (1) history of skill tests, (2) purposes of a skill test, (3) criteria for skill tests, (4) considerations in applying criteria, (5) selection of a soccer skill test and (6) available soccer skill tests, (7) basis for selection of test.

History of Skill Tests. Bovard, Cozens and Hagman divide the history of measurement work in physical education into five general phases or stages. They are as follows: (1) Anthropometric from 1860-1880, (2) Strength from 1880-1915, (3) Cardiac functional from 1900-1925, (4) Athletic ability from 1904 to present date and (5) Single test or index figure from 1920 up to present date.

Although these periods overlap and are not clear-cut steps in the use of testing measures, they do give us ideas as to how the problem has been attacked and the dates help us to fix the limits as to when a particular test was most used by leaders in physical education.[1]

---

[1]John F. Bovard, Frederick W. Cozens and E. Patricia Hagman, Tests & Measurements in Physical Education (Philadelphia: W. B. Saunders Company, 1949), p. 17.

The Athletic Badge tests developed by the Playground and Recreation Association of America in 1916 were the first to make use of game elements as possibilities for testing in physical education. These tests include; (1) baseball throw for accuracy, (2) baseball throw for distance, (3) basketball throw for distance, (4) volleyball serve, (5) tennis serve, (6) baseball throw and catch and (7) basketball goal throw.[2]

As early as 1916, Reilly's program of rational athletics included certain tests which contained the elements of game activities such as; (1) baseball pitching, (2) basketball goal throw, (3) throwing the basketball, (4) serving in tennis, (5) putting in golf and (6) driving in golf.[3]

In developing the California decathlon, Hetherington included the following new tests; (1) punting a football for distance, (2) soccer kick for goal, (3) running and catching and (4) tennis and volleyball serving.[4]

Many of the tests constructed before 1924 included tests of fundamental sport techniques but there were very few batterys of tests which would measure (or attempt to measure)

---

[2]Bovard, Cozens, and Hagman, op. cit., p. 190.

[3]Ibid., pp. 190-191

[4]Ibid., p. 191.

many of the fundamental skills in a particular sport.[5]

The introduction of statistical methods into test construction appeared about 1930. At that time, a number of tests, including both the classification and achievement types, were devised and are still usable today. Some of these early works, using scientific procedures of test construction include Beall[6] in tennis, Brace[7] in basketball, and Rodgers and Heath[8] in playground baseball and soccer.

One of the publications in 1937 by Cozens, Cubberley and Neilson proposed achievement scales in basketball, baseball, track and field, soccer, softball, archery, field hockey, speedball, swimming, volleyball and tennis. Although the validity and reliability are not reported, they do

---

[5] Ibid., p. 192.

[6] Elizabeth Beall, Essential Qualities in Certain Aspects of Physical Education with Ways of Measuring and Developing Same. Unpublished Master's Thesis, University of California, 1925, cited by Leonard A. Larson and Rachael Yocom, Measurement and Evaluation in Physical, Health, and Recreation Education (St. Louis: C. V. Mosby Company, 1951) p. 206.

[7] David K. Brace, Measuring Motor Ability (New York: A. S. Barnes and Company, 1927) pp. 74-77.

[8] Marjorie L. Heath and Elizabeth G. Rodgers, "A Study in the Use of Knowledge and Skill Tests in Soccer," Research Quarterly, 3:33-53, December, 1932.

present achievement scales which are still useful.

The next group of published tests included those in swimming, volleyball, tennis, basketball, dance, football, gymnastics, field hockey, badminton and softball.

The emphasis during World War II on physical fitness resulted in considerable experimentation on test and standards in a wide variety of physical education activities. The Division of Health and Physical Education of the New York State Education Department in cooperation with the Office of Physical Fitness of the New York State War Council[10] developed a set of standards for boys and young men in the following six areas:

1. Gymnastics, Apparatus and Tumbling Activities.
2. Individual and Dual and Combative Sports.
3. Track and Field Activities.
4. Swimming and Water Safety.
5. Team Sports.
6. One Hundred Yard Obstacle Course.

Evaluative procedures were developed in the following

---

[9] Frederick W. Cozens, Hazel J. Cubberley and N. P. Neilson, Achievement Scales in Physical Education Activities for Secondary School Girls and College Women (New York: A. S. Barnes and Company, 1937).

[10] New York State Physical Fitness Standards for Boys and Young Men, "A Manual for Young Men," and New York State Physical Fitness Standards, "Evaluative Procedures in Physical Activities for Girls and Young Women," (Albany, New York, New York State War Council, Office of Physical Fitness and New York State Education Department, Division of Health and Physical Education, 1944).

activities for girls and young women:

1. Team Sports: Basketball, Field Hockey, Softball, Soccer, Speedball, Volleyball.
2. Individual Sports: Archery, Badminton, Bicycling, Bowling, Croquet, Deck Tennis, Fencing, Handball, Golf, Horseshoes, Horseback Riding, Ice Skating, Paddle Tennis, Roller Skating, Shuffleboard, Skiing, Snow Shoeing, Table Tennis, Tennis, and Tobogganning.
3. Dance: Folk Dance, Modern Dance, Social Dance, Tap Dance.
4. Self Testing Activities: Stunts and Tumbling, Apparatus, and Emergency Skills.
5. Aquatics: Swimming, Boating, Canoeing.
6. Outing and Camping.
7. Fundamentals of Motor Performance.

An award system was established listing "Merit," "Excellent," and "Superior" achievement standards in each area. Those meeting standards in all areas could earn an ALL-ROUND award.

While work on the validation and perfection of these standards, tests, and evaluative procedures was not completed, the program as developed was administratively sound, and served to stimulate interest and effort in improving the physical education programs in the high schools of the state.

Purposes of a Skill Test. A skill test is one of the tools used to measure the progress of the student and to help determine whether or not the objectives of the program are being met. Just as in other branches of education, the tools of measurement in physical education have changed as it has developed. Skill tests are usually given to determine the general ability in a particular sport or specific ability in

one technique important to that sport. The relationship between the specific technique to the game determines the significance of measuring one specific ability.[11]

Criteria for Skill Tests. Tests must possess certain qualities if they are to serve the purposes expected of them. Scott and French[12] have set up the following list as the most pertinent criteria for the selection of a skill test:

1. Test should measure important abilities.
2. Tests should be like the game situation.
3. Tests should encourage good form.
4. Tests should involve one performer only.
5. Tests should be interesting and meaningful.
6. Tests should be of suitable difficulty.
7. Tests should differentiate between levels of ability.
8. Tests should provide accurate scoring.
9. Tests should provide a sufficient number of trials.
10. Tests should be judged by statistical evidence.
11. Tests should provide a means for interpreting performance

---

[11] M. Gladys Scott and Esther French, Measurement and Evaluation in Physical Education, (Dubuque, Iowa: Wm. B. Brown Company, 1959) pp. 1-6.

[12] Ibid., pp. 10-32.

Practical Considerations in Applying Criteria. In applying the above, certain points must be kept in mind. They are:

1. No test will contain all criteria, therefore the purpose of the test will help to determine which criteria is most important.
2. Since a single test cannot test all the skills involved, a small sampling may be adequate.
3. Effective test results are dependent upon maximum effort by each student.
4. Tests are only tools to estimate individual abilities, therefore, the teacher must be skilled in using these tools.
5. With rare exceptions, the test is not the game itself, but if the test measures the same relative abilities as the game, it will have served one purpose, that of determining relative abilities of class members.
6. Since groups differ, so will the norms and achievement scales differ.[13]

Available Soccer Skill Tests. The following is a list of some of the available skill tests for soccer. A brief description follows each test.

1. McDonald Volleying Test--A test of general soccer ability for college men. The test consists of only one item, kicking against a backboard that is 30 feet wide and 11½ feet high. Approximately 20 subjects may be tested at one station during a 40-minute period. A restraining line is drawn 9 feet from the backboard and parallel to it. Three soccer balls are used; one is placed on the restraining line; the other two are located 9 feet behind this line in the center of the area. The test consists of kicking the soccer

---

[13] Ibid., pp. 32-33

ball against the backboard as many times as possible in 30 seconds. Any type of kick may be used; both ground balls and fly balls which hit the backboard count. To count, however, all balls must be kicked from the ground with the supporting leg behind the restraining line. Rebounds may be retrieved in any manner, including use of the hands. If a ball is out of control, the subject may play one of the spare balls, but must bring the ball, by use of hands or feet to a position at the restraining line before kicking against the backboard (no penalty other than the lost time in getting the ball in position to kick). The score is the number of legal kicks in the time period; the best of four trials is recorded. Reliability correlations were .94 for varsity players, .63 for junior varsity players, .76 for freshman varsity, and .85 for combined groups.[14]

2. Bontz Combination of Soccer Skills--Purpose is to test for general soccer ability. Test was given to fifth and sixth grade boys and girls. The test includes a straight away dribble combined with a side pass to wall and recovery, followed by a continued dribble and kick for goal. This test produced a reliability coefficient of .96, and a validity coefficient of .92 with a subjective rating criterion. The score is the total amount of time it takes to complete the eight trials; each one is timed separately. Four trials are given with the wall on the left, making the pass with the right foot, then four more trials are given with the wall on the right, passing with the left foot. Trials which include errors must be repeated.[15]

---

[14]Lloyd G. McDonald, "The Construction of a Kicking Skill Test as an Index of General Soccer Ability," Unpublished Master's Thesis, Springfield College, Springfield, Massachusetts, 1951, cited by Harrison H. Clark, Application of Measurement to Health and Physical Education, (Prentice-Hall, Englewood Cliffs, N. J., 1959), pp. 342-343.

[15]Jean Bontz, "An Experiment in the Construction of a Test for Measuring Ability in Some of the Fundamental Skills Used by Fifth and Sixth Grade Children in Soccer," Unpublished Master's Thesis, State University of Iowa, Iowa City, Iowa, 1942, cited by Scott and French, Measurement and Evaluation in Physical Education (Dubuque, Iowa: Wm. C. Brown Company, 1959), p. 195.

3. Shaufele Test--Purpose of test is to measure soccer playing ability. Test was given to ninth and tenth grade girls. The test is made up of three items; they are a repeated volley test, in which the soccer ball is kicked successively against a wall; a passing and receiving test which involves dribbling the ball, passing it against a wall, controlling it on the rebound and repeating; and a judgment in passing test which requires a dribble and kick for goal. Coefficients of validity with a subjective rating of player ability as a criterion were .57, .50 and .34 respectively for the three tests, but were .77, .72 and .82 when a total battery criterion was used. Reliability coefficients were .67, .72 and .69. T-scales are available.[16]

4. Vanderhoff Ability Tests in Soccer--Purpose of test is to stimulate students to improve their skill and to show progress made during a unit of soccer. The test consists of a battery of ten items. No attempt has been made to validate the battery but the tests represent important element in the game and include: (1) dribble, (2) trapping, (3) throw-in, (4) place kick for accuracy, (5) punt for distance, (6) volleying (by using forehead, shoulder, hip and knee). (7) throw-down (securing ball from opponent in a 6-yard circle), (8) tackling, (9) corner kick, (10) goal keeper's test (skill in preventing goals). Scoring has been made objective but is set up empirically with a possible 10 points for each test. By validating it and using a scoring scheme worked out by correct statistical procedures, this battery of tests shows possibilities for use in measuring ability to play soccer.[17]

5. Heath and Rodgers Ability Test in Soccer--Purpose of

---

[16]Evelyn F. Shaufele, "The Establishment of Objective Tests for Girls of the Ninth and Tenth Grades to Determine Soccer Ability," Unpublished Master's Thesis, State University of Iowa, Iowa City, Iowa, 1942, cited by Scott and French, Measurement and Evaluations in Physical Education (Dubuque, Iowa: Wm. C. Brown Company, 1959), pp. 187-192.

[17]Mildred Vanderhoff, "Soccer Skill Tests," Journal of Health and Physical Education, 3:42, pp. 54-56, October, 1932.

test is to evaluate ability in soccer skills. Test was given to fifth and sixth grade boys and girls. The test consists of four items; (1) dribble, (2) throw-in, (3) place kick for goal and (4) kicking a rolling ball. Two class periods of approximately 40 minutes are required for testing a class of 50. A validity coefficient of .62 when scores are correlated with a judgment rating and a reliability of .74 were arrived at. T-scales are available for fifth and sixth grad boys.[18]

6. Buchanan Speedball Tests--Purpose of test is to measure fundamental skills in speedball and to provide a test of playing ability. The test was given to high school girls. This test had one item in relation to soccer. It is a dribbling and passing test around five Indian clubs over a course 60 yards long. Three scores are given on this test; a combined dribbling and passing score, a time score for dribbling, and a point score for accuracy in passing. Ten trials, five to the right and five to the left are given. Approximately 8 to 10 subjects may be measured at one station during a 40 minute period.[19]

7. Battery of Soccer Tests by Scott and French--Purpose of this battery of soccer tests is to determine which tests are of the most value in determining soccer ability. The one best test for high school girls is volleying. A good two-item battery of tests is the combination of passing and receiving with volleying. The following equation in combining the two items is used:

2.0 passing and receiving + 1.0 volleying.

---

[18]M. L. Heath and Elizabeth Rodgers, "A Study in the Use of Skill and Knowledge Tests in Soccer," Research Quarterly, 3:35-54, December, 1932.

[19]Ruth E. Buchanan, "A Study of Achievement Tests in Speedball for High School Girls," Unpublished Master's Thesis, State University of Iowa, Iowa City, Iowa, 1942, cited by Weiss, Raymond A. and Marjorie Phillips, Administration of Tests in Physical Education (St. Louis: C. V. Mosby Company, 1954), pp. 253-257.

> An equally good two-item battery is the combination of judgment in passing with volleying. The recommended weighting of scores is:
>
> 2.0 judgment in passing + 1.0 volleying.[20]

Selection of a Soccer Skill Test. The McDonald Soccer Skill Test was chosen for the purpose of this study as it contained, in the writer's estimation, most of the criteria for a good test with the exception of encouragement of good form. In this particular test any type of body form could be used, and a satisfactory grade could be obtained. In many of the tests reviewed the major disadvantage to administering the test was that they were too time consuming.

In recent years there have been comparatively few tests of soccer skill devised, and of those with any reliability and validity, have only norms set up for boys and men.

In review of several of the other tests of soccer skill, the Bontz Test[21] would have been the best test to administer to obtain a good evaluation of all the soccer skills involved, but it would have taken approximately two weeks to administer to the students at Brooks Junior High School because of the size of classes and the number of days

---

[20] Scott and French, op. cit., p. 198.

[21] Bontz, op. cit., p. 195.

per week that the students attend physical education class. Many of the other tests pertained to speedball and speed-a-way skills, with usually only one item, that of volleying, which really pertained to the main skills involved in the game of soccer. Speedball and speed-a-way are games involving soccer, basketball, field hockey and football skills that have been invented in the past twenty years and have taken the place of soccer in many school's activity programs.

Basis for Selection of Test. The McDonald Soccer Skill Test was chosen in preference to the other tests to be given at Brooks Junior High School for the following reasons: (1) small amount of time to be given, (2) contained the one best test item, that of volleying, for determining soccer ability, (3) ease of administration. In view of the available soccer skill tests, most of them would have taken approximately two or three weeks of class time to be given at this particular school due to the size of classes and number of class periods per week. The test item of volleying was included in all of the tests reviewed and was considered the one best test item for determining soccer skill. The scoring of the test was simple and could be easily added and averaged by the students taking the test.

Summary. The five phases of measurement work in the history of physical education are; (1) Anthropometric,

(2) Strength, (3) Cardiac functional, (4) Athletic ability and (5) Single test or index figure. The first use of game elements as possibilities for testing in physical education came in the development of the Athletic Badge Tests by the Playground and Recreation Association of America in 1913. In 1916, Reilly introduced his program of rational athletics which was followed by Hetherington in developing the California decathlon. With the introduction of statistical methods around 1930, a number of sports tests were devised which are usable today. Classification and achievement were included in these tests. During World War II the emphasis was placed on physical fitness which resulted in a wider variety of physical education activities. Evaluative procedures were developed in many new areas because of this trend.

Skill tests are usually given to determine the general ability in a particular sport or specific ability in one technique important to that sport. If tests are to serve the purposes we expect of them, they must possess certain qualities. Some of these qualities are; (1) be like game situation, (2) involve only one performer, (3) be interesting and meaningful, (4) differentiate between levels of ability, (5) provide accurate scoring, (6) be judged by statistical evidence, and (7) provide a means for interpreting performance.

Certain points must be kept in mind when judging the value of a good skill test. Some of these points are; (1) one test cannot possess all of the good characteristics, (2) a test can only sample the many skills involved in an activity, (3) effective results are dependent upon maximum effort of the participant, (4) the test administrator is the key to obtaining good results and (5) groups differ, so the norms and achievement scales must differ.

The McDonald Soccer Skill Test was chosen for the purpose of this study as it contained, in the writer's estimation, most of the criteria for a good test. In most of the tests reviewed, the major disadvantage to the test was the time element. The Bontz Test would have been one of the best tests to administer but it would have taken approximately two weeks to administer due to the large size of classes and number of times per week the girls at Brooks Junior High School attend class. Some of the available tests used for testing soccer skills other than the McDonald Test are; (1) the Bontz Combination of Soccer Skills, (2) Shaufele Test, (3) Vanderhoff Ability Tests, (4) Heath and Rodgers Ability Test and (5) Buchanan Speedball Tests. According to Scott and French a good two-item battery of soccer tests should include the combination of passing and receiving with volleying. The one best test item for girls is volleying.

The McDonald Skill Test was chosen to be given at Brooks Junior High School for the following reasons; (1) small amount of time to be given, (2) contained the one best test item, that of volleying, and (3) ease of administration.

CHAPTER III

METHODOLOGY

Introduction. In Chapter III, the methods and procedures of the study of the McDonald Skill Test will be discussed. The chapter will be divided into the following areas: (1) Sampling, (2) Testing procedure, (3) Score card data, (4) Description of test and (5) Analysis plan.

Sampling. The subjects used for this study were 265 girls at Brooks Junior High School, Wichita, Kansas, during the school year 1961-1962, who were enrolled in physical education classes. There were 100 ninth grade girls, 80 eighth grade and 85 seventh grade girls used for this study. The scheduling of physical education at Brooks Junior High School is irregular as most seventh and eighth grade girls attend class three times per week and a few come only two days per week. The ninth grade has some students who come five days per week, some three days and others two days per week depending on their other schedule of classes. All three grades are mixed in the same class period.

Testing Procedure. The McDonald Skill Test was given at Brooks Junior High School to all girls who were enrolled in physical education. The McDonald Test was given after four weeks of soccer participation in class.

This same test was then given one week later to determine the reliability of the test for girls of junior high school age. On the second test the girls were told that they could improve their grade from the first test.

Each class was divided into four equal groups and stationed at four different areas of wall space in the gymnasium. Five spare balls were placed in the center of the area on a mat and if a ball went out of control they could use one of these. Other girls not taking the test returned the balls to the mat.

In taking the test, each girl waited approximately five minutes between each kicking trial of the test, in an attempt to eliminate any fatigue factor, rather than taking all four trials in succession.

Each girl was given a score card to be filled out and the method for averaging the scores was explained. After the final completion of both tests all the girls exchanged score cards and checked for errors in addition and averaging of scores. A score card may be found in Appendix D.

Score Card Data. The data collected on the score card is explained in the following statements. The Test I total was the composite score of all four trials. The average of four was determined by dividing the total of all four scores by four. The average of the best three was determined by

omitting the lowest score and adding the remaining three and dividing by three. The same procedure was used on Test II.

Description of Test. At the command of Ready, Go, the subject begins kicking the soccer ball against the wall for 30 seconds. Any type of kick may be used; both ground balls and fly balls which hit the wall count. To count, however, all balls must be kicked from the ground with the supporting leg behind the restraining line. Rebounds may be retrieved in any manner, including use of the hands. If a ball is out of control, the subject may play one of the spare balls, but must bring the ball by use of hands or feet to a position at the restraining line before kicking against the wall (no penalty other than the lost time in getting the ball back in position to kick). Spare balls are placed 9 feet behind the line and in the center of the testing area. The score is the number of legal kicks in 30 seconds. Four trials are permitted with two practice trials.[1]

---

[1]McDonald, Lloyd G., "The Construction of a Kicking Skill Test as an Index of General Soccer Ability," Unpublished Master's Thesis, Springfield College, Springfield, Mass., 1951, cited by Clark, H. Harrison, Application of Measurement to Health and Physical Education, (Englewood Cliffs, N. J., Prentice-Hall, Inc., 1959), pp. 342-343.

A diagram of the floor plan and description of the test may be found in Appendix A.

On the official McDonald test, a line 11½ feet high from the floor was drawn on the wall and balls kicked higher than this did not count as "fair" kicks. Very few girls kicked the ball this high, so this part of the test was eliminated. The official test was administered to college men and no norms, at any level, have ever been devised for this test.

Analysis Plan. Individual raw scores were used for all analyses. The mean, standard deviation, chi square, standard error, linear product moment correlation, calculation and interpretation of the difference between two means, and standard scores were the statistical measures used on the data.

First the raw scores were analyzed by using the total of all three grades, then each grade was analyzed separately.

A complete discussion of the results of the statistical analysis of the data will be given in Chapter IV.

Summary. The subjects used for this study were 265 girls at Brooks Junior High School, Wichita, Kansas during the school year 1961-1962, who were enrolled in physical education classes. There were 100 ninth graders,

80 eighth graders and 85 seventh graders used in this study. The scheduling of physical education at this particular school is irregular as some girls attend class two days per week, some three days per week and others five days per week.

The McDonald Test was given after four weeks of participation in soccer and the exact same test was then given one week later to determine the reliability of the test for junior high school girls.

Each class was divided into four groups and stationed at four different areas of wall space. In taking the test, each girl waited five minutes between each kicking trial of the test.

All girls were instructed on how to fill in the score card which is shown in Appendix D, with an explanation of what the scores meant.

A description of the test and a diagram of the floor plan for the test are found in Appendix A. The official test was deviated from and on this test the line 11½ feet high from the floor was omitted as very few girls kicked the ball higher than this.

In analyzing the data, individual raw scores were used. The mean, standard deviation, chi square, standard error, linear product moment correlation, calculation and

interpretation of the difference between two means and standard score were the statistical measures used. The raw scores were first analyzed by using the total of all three grades. Then each grade was analyzed separately.

CHAPTER IV

RESULTS AND DISCUSSION

Introduction. The results of the McDonald Soccer Skill Test, which was adminstered to Junior High School Girls at Brooks Junior High School, Wichita, Kansas, during the school year 1961-1962, will be discussed in this chapter. The reliability of the test, changes in test scores between test I and test II, norms for junior high school girls, and interpretation of the results of the tests will be the four areas discussed. At Brooks Junior High School some girls come to class two days per week, other three days and some five days per week. No attempt will be made in this study to determine the value of class time but each grade will be considered separately in setting up the norms for grading purposes. Invidividual raw scores were used in all computations.

Summary of Raw Score Data. In the following two tables, the mean and standard deviation of the scores give a general overview of the results of the test.

TABLE I

SUMMARY OF McDONALD SOCCER SKILL TEST (SOCCER KICKS IN 30 SECONDS) ADMINISTERED TO BROOKS JUNIOR HIGH SCHOOL GIRLS, WICHITA, KANSAS, 1961 - 1962
COMBINED 7-8-9 GRADES

Test I	Mean	Standard Deviation	Test II	Mean	Standard Deviation
Trial 1	11.4		Trial 1	12.6	
2	12.1		2	13.4	
3	12.5		3	13.2	
4	12.9		4	13.5	
$\bar{X}_4$	12.3	2.24	$\bar{X}_4$	13.3	2.12
$\bar{X}_3$	13.0	2.32	$\bar{X}_3$	13.9	2.16

Note: $\bar{X}_4$ - the mean of four trials
$\bar{X}_3$ - the mean of the three best trials

TABLE II

SUMMARY OF McDONALD SOCCER SKILL TEST (SOCCER KICKS IN 30 SECONDS) ADMINISTERED TO BROOKS JUNIOR HIGH SCHOOL GIRLS, WICHITA, KANSAS, 1961 - 1962, GRADES 7-8-9 LISTED SEPARATELY

Test I	Mean	Standard Deviation	Test II	Mean	Standard Deviation
Grade 7 $\bar{X}_4$	12.0	2.12	Grade 7 $\bar{X}_4$	12.5	1.79
$\bar{X}_3$	12.3	1.97	$\bar{X}_3$	12.8	1.86
Grade 8 $\bar{X}_4$	12.3	2.19	Grade 8 $\bar{X}_4$	13.4	1.94
$\bar{X}_3$	12.7	2.33	$\bar{X}_3$	14.1	2.20
Grade 9 $\bar{X}_4$	12.7	2.34	Grade 9 $\bar{X}_4$	13.5	2.36
$\bar{X}_3$	13.7	2.31	$\bar{X}_3$	14.4	2.22

Note: $\bar{X}_4$ - the mean of four trials
$\bar{X}_3$ - the mean of the three best trials

By observation of Table I, it can be seen that the mean of all the trials did change somewhat. On test I, trial 1, with a mean of 11.4 and on test II, trial 4, 13.5 resulted in an improvement of 2.1 kicks. On the composite score of test I, four trial average, a mean of 12.3 was arrived at and on test II, four trials, a mean of 13.3 resulted in a 1.0 kick difference. On the best of three trials a .9 difference was found. Thus to determine which would be the most satisfactory method of averaging the total trials, the best of three would result in a higher average, but very little difference was found in using four trials or the best of three. The psychological effect which might have some motivating factor on the subject taking the test would be one reason for using the best of three trials.

In Table II, each grade was listed separately and the seventh graders had the lowest mean while the ninth graders had the highest mean. Each grade improved on test II. The seventh grade had a .5 improvement on both the best of three and the four trial averages. The eighth grade had a 1.1 improvement on the four trial average and 1.4 on the best of three. The ninth grade had a .8 improvement on the four trial average and .7 on the best of three trial average.

Reliability. The reliability of a test means that the test will repeat itself. With the same person giving the

test, using the same methods with the same students, the same scores will be obtained both times.[1]

According to Remmers and Gage[2] "if a test is valid it must also be reliable, whereas a reliable test may not be valid." Therefore, an unreliable test if correctly administered may be more valid than a reliable test administered for purposes for which it is unsuitable. Reliability is included within validity; it is essential but not sufficient to it. In order for a test to be valid it must have some degree of reliability, but the opposite is not always true.

Reliability may be determined by three different methods. These include (1) stability which is accomplished by giving the same test twice with a time interval between adminstrations; (2) internal consistency which is accomplished by obtaining two test scores from the same test, and (3) equivalent forms, which are different forms of the same test but are as similar as possible, administered with a very short time interval between test administrations.

The correlation of test I, four trials and test I, best of three trials for the combined 7-8-9 grades resulted

---

[1] Carl E. Willgoose, Evaluation in Health Education and Physical Education, (McGraw-Hill Book Company, New York, 1961), pp. 24-25.

[2] H. H. Remmers and N. L. Gage, Educational Measurement and Evaluation, (Harper and Brothers, New York, 1943), pp. 132.

in a r of .8095. On test II, four trials and test II, best of three trials for the combined 7-8-9 grades, a correlation of .7810 resulted. The correlation between the two tests was considerably lower with a .6248 correlation on four trials and a .5715 on the best of three trials. Therefore, only a slight difference resulted in four trials compared to three trials. An interpretation of these differences necessitates a discussion of internal consistency and stability concepts of reliability.

According to Lindquist[3] the degree of reliability a test has indicates how consistent and how stable this test is in relation to its measuring the same thing twice regardless of what it is intended to measure. There are no set standards for reliability coefficients to determine "high" degree of reliability or a "low" degree of reliability. This will have to depend on what the test scores are used for, that is, if they are considered reliable enough to be of some value. Another point to be considered is, if a test is low in reliability it would also be low in validity.

In comparing the r of .8095 for test I, four trials and test I, best of three trials, and on test II, four trials and test II, best of three trials with r of .7810 there was

---

[3]Lindquist, E. F., A First Course in Statistics, (Cambridge, Massachusetts: The Riverside Press, 1942), pp. 217-224.

a relatively high degree of internal consistency. This was due to the fact the same scores were used except in the case of the best of three trials, the lowest score of the four trials was omitted.

The difference of the r between stability and internal consistency was larger than was to be expected of a test that is reliable; therefore, it must be concluded that the McDonald Soccer Skill Test is not a reliable test for junior high school girls. This would also indicate that this test could not be very high in validity for junior high school girls.

Changes in Test Scores. In experimental studies when a difference occurs that cannot be attributed to chance factors or sampling variations, it is often of great importance.[4] In this study the writer had no basis for believing that the mean score of test I would be greater than the mean score of test II or that just the opposite would be true. The writer was only interested in any difference that did occur between the two means, regardless in which direction the difference occurred. The t test was the method used to determine this factor.

The .01 level of confidence was used as the level to

---

[4]Edwards, Allen L., Statistical Methods for the Behavioral Sciences, (New York: Rinehart and Co., Inc., 1956) p. 249.

be considered statistically significant in this study. For twelve degrees of freedom, a t which results in a value larger than 3.06 is necessary for significance at the .01 level. The values of t obtained from this study were:

Test I - 3 trials v.s. Test II - 3 trials - t - 6.976

Test I - 4 trials v.s. Test II - 4 trials - t - 8.175

Test I - 4 trials v.s. Test I - 3 trials - t - 8.178

Test II - 4 trials v.s. Test II - 3 trials - t - 6.833

All of the values of t exceeded the value necessary for significance at the .01 level. Therefore, we can assume that chance factors did not cause the observed differences between test scores considered. It is possible that these significant values of t occurred as a result of the skill practice from taking the test the first time which could have caused the improvement in skill on the second trial of the test.

Test for Normality. The normal curve is used as a basis for determining the extent of the deviation of an obtained frequency distribution. Judgment must be made by comparison, whether or not the deviations are large enough to prevent the use of the distribution as normal. The inspectional test for normality, which is the frequency polygon, was first graphed for combined 7-8-9 grades and then for each grade separately. These are shown in Appendix

C on two separate frequency polygons and appear to be bell-shaped.

The chi square test as a statistical test for normality was computed to test the hypothesis that the sampling of scores obtained from the Brooks Junior High School girls on the McDonald Soccer Skill Test had a distribution that was not significantly different from a normal curve. The following chi square values resulted:

Grade 7 30

Grade 8 77

Grade 9 74

Combined 7-8-9 grades . 52

The .05 level of confidence was used as the level to be considered statistically significant in this study. A chi square of 18.307 for 10 degrees of freedom is required for significance at the .05 level. The chi square obtained for grade 7 was 30; grade 8, 77 and grade 9, 74; on the combined grades 7-8-9 a chi square of 52 resulted. All of the chi square values obtained exceeded the value necessary for significance at the .05 level. Therefore,we reject the hypothesis at the .05 level that the distribution on the McDonald Soccer Skill Test was not significantly different from normal on each of the grades and for combined grades 7-8-9.

Two possibilities for these significant chi square values might be: (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill present in the sample of Brooks Junior High School girls was not distributed normally. It would seem feasible that the sample of 265 girls obtained from Brooks Junior High School with grade 7-8-9 would have been large enough and widely varied enough, according to ability, to obtain a normal distribution of test scores for soccer ability. Therefore, considering the desirable factor of the test discriminating between different levels of ability, it would indicate that the McDonald Soccer Skill Test does not accomplish this for the junior high school girls represented in this study.

Norms. As no norms have ever been devised for girls of junior high school age on the McDonald Soccer Skill Test, one of the purposes of this study was to set up ^{T}-scores which may be used as a grading scale for each particular grade level and also one grading scale which could be used for junior high school girls and not separating each particular grade level but retains them as a whole unit. The ^{T}-scores may be found in Table V. In using the ^{T}-scores that were prepared, they must be used with some caution and

only under certain conditions due to the findings of the significant chi square values which indicated lack of normality in the obtained distribution of test scores. This caution must be used regardless of the cause of the significant chi square.

The letter grade scores that are given in Appendix B are in the proportion of 7% - A, 24% - B, 38% - C, 24% - D and 7% - F in regard to the distribution of the test scores from the Brooks Junior High School girls sample.

Interpretation of Results. The comparison of the coefficients of correlation between stability and internal consistency were the methods used to establish reliability for the McDonald Soccer Skill Test. Only rarely will the two scores be the same due to the inherent qualities within these two methods.

In relation to the other soccer skill tests discussed in Chapter II, the reliability coefficients for the McDonald Soccer Skill Test[5] was .94 when administered to varsity players, .63 for junior varsity players, .76 for freshmen varsity, and .85 for combined groups. The Bontz Test[6] had a

---

[5]McDonald, op. cit., pp. 342-343.

[6]Bontz, op. cit., p. 195.

r of .96, the Shaufele Test[7] .67, and the Heath and Rodgers Test[8] .74. Therefore, with a r of .62 on four trials and .57 on the best of three trials for the Brooks Junior High School sampling, the reliability coefficient was lower in respect to all of the other tests but not too much lower than the Shaufele Test and the junior varsity players in the original McDonald Test.

If the method used to establish reliability is considered, it is conceivable that the coefficients found on the McDonald Soccer Skill Test for the Brooks Junior High School girls (.80 four trials and .78 best of three) through the method of internal consistency are comparable to those reported in other soccer skill tests. However, if the method of reliability was through stability, (.62 four trials and .57 best of three trials) the coefficients obtained from the Brooks sample are much lower in comparison to these same tests. The type of reliability was not specified in the various soccer tests reviewed, and therefore the writer assumes that the method of reliability established was by stability.

Based on this data it must be concluded that the McDonald Soccer Skill Test was not as reliable as the other

---

[7] Shaufele, op. cit., pp. 187-192.

[8] Heath and Rodgers, op. cit., pp. 35-54.

previous experience in soccer and therefore would have a larger range for improvement.

The t test was used for this study to indicate any difference between the means of test I, best of three trials to test I, four trials; test II, best of three trials to test II, four trials; test I best of three trials to test II, best of three trials; and test I, four trials to test II, four trials. On the values of t obtained as shown on page 35 the results were significant at the .01 level of confidence. On the basis of the statistical data we can assume that chance factors did not cause this large of a difference. It may be possible that this difference occurred as a result of the skill practice from taking the test the first time. In order to determine whether the observed improvement in soccer volleying is due to the skill practice obtained from the first test, a series of studies would have to be carried out.

If the observed improvement occurred as a result of practice alone one could reason that the subjects would continue to show an increase of improvement at the same rate. Thus a study designed to test the same individuals over a specific period of time could be attempted. It would be expected that a time interval of one week between test administrations for the entire unit of soccer should result in an increase of improvement at the same rate.

The chi square test was computed to test the hypothesis that the sampling of scores obtained in this study had a distribution that was not significantly different from a normal curve. The .05 level of confidence was used as the level to be considered statistically significant. All of the chi square values obtained exceeded the value necessary for significance at the .05 level. Therefore, the hypothesis was rejected. Two possibilities for this might be: (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill present in the sample of Brooks Junior High School girls was not distributed normally. It would seem logical that the sample obtained was large enough and widely varied enough, according to ability, to obtain a normal distribution of test scores for soccer ability. Therefore, the data indicates that the test did not discriminate between different levels of soccer ability.

Summary. The computations for all of the statistical data were made from individual raw scores. According to Table I for combined 7-8-9 grades, the mean of all the trials did change somewhat. On test I, trial 1, with a mean of 11.4 and on test II, four trials, 13.5 resulted in an improvement of 2.1 kicks. On the composite score of test I,

four trial average, a mean of 12.3 was arrived at and on test II, four trials, a mean of 13.3 resulted in a 1.0 kick difference. On the best of three trials a .9 difference was found. Thus to determine which would be the most satisfactory method of averaging the total trials, the best of three would result in a higher average, but very little difference was found in using four trials or the best of three.

In Table II for 7-8-9 grades listed separately, the seventh graders had the lowest mean while the ninth graders had the highest mean. Each grade improved on test II. The seventh grade had a .5 improvement on both the best of three and four trial averages. The eighth grade had a 1.1 improvement on the four trial average and 1.4 on the best of three. The ninth grade had a .8 improvement on the four trial average and .7 on the best of three trial average.

The reliability of a test means that the test will repeat itself. With the same person giving the test, using the same methods with the same students, the same score will be obtained both times. In order for a test to be valid it must have some degree of reliability, but the opposite is not always true.

Reliability may be determined by three different methods. These are (1) stability which is accomplished by giving the same test twice with a time interval between administrations; (2) internal consistency which is accom-

plished by obtaining two test scores from the same test, and (3) equivalent forms, which are different forms of the same test but are as similar as possible, administered with a very short time interval between test administrations.

The correlation of test I, four trials and test I, best of three trials for the combined 7-8-9 grades resulted in a r of .8095. On test II, four trials and test II, best of three trials for the combined 7-8-9 grades, a correlation of .7810 resulted. The correlation between the two tests was lower with a .6248 correlation on four trials and a .5715 on the best of three trials. Only a small difference resulted in four trials compared to three trials. There are no set standards for reliability or a "low" degree of reliability. This will have to depend on what the test scores are used for, that is, if they are reliable enough to be of some value.

In comparing the r of .8095 for test I, four trials and test I, best of three trials, and on test II, four trials and test II, best of three trials with a r of .7810 there was a relatively high degree of internal consistency. This was due to the fact the same scores were used except in the case of the best of three trials, the lowest score of the four trials was omitted.

The difference of the r between stability and internal consistency was larger than was to be expected of a test that is reliable; therefore, it must be concluded that the

McDonald Test is not reliable for junior high school girls. This also indicates that this test could not be very high in validity for junior high school girls.

The t test was the method used to determine any difference that didoccur between the two means, regardless in which direction the difference occurred. In this study the .01 level of confidence was used as the level to be considered statistically significant. All of the values of t exceeded the value necessary for significance at the .01 level. Therefore, we can assume that chance factors did not cause the observed difference between test scores considered. It is possible that these significant values of t occurred as a result of the skill practice from taking the test the first time which could have caused the improvement in skill on the second trial of the test.

The chi square test as a statistical test for normality was computed to test the hypothesis that the sampling of scores obtained in this study were not significantly different from a normal curve. The .05 level of confidence was used as the level to be considered statistically significant. All of the chi square values obtained exceeded the value necessary for significance at the .05 level. Therefore we reject the hypothesis at the .05 level that the distribution on the McDonald Soccer Skill Test was not significantly different from normal on each of the grades and for combined

grades 7-8-9. Two possibilities for these significant chi square values might be; (1) the test was not able to discriminate between different levels of soccer ability in the sample according to normal distribution or (2) that the range of ability in soccer skill present in the sample of Brooks Junior High School girls was not distributed normally. It would have seemed feasible that the sample of 265 girls in grades 7-8-9 would have been large enough and widely varied enough, according to ability, to obtain a normal distribution. Therefore it would indicate that the McDonald Soccer Skill Test did not discriminate between different levels of ability.

T-scores were set up for grading purposes for each individual grade and also for the combined 7-8-9 grades. The T-scores may be found in Table V. Some caution must be used with these T-scores due to the significant chi square values obtained which indicated lack of normality in the distribution of test scores.

The letter grade scores that are given in Appendix B are in the proportion of 7%-A, 24%-B, 38%-C, 24%-D, and 7%-F in relation to the distribution of the test scores from the Brooks Junior High School girls sample.

In an interpretation of results, the comparison of the coefficients of correlation between stability and internal consistency were the methods used to establish

reliability for the McDonald Soccer Skill Test. In relation to the other soccer skill tests studied the Brooks Junior High School sampling was lower than all the other tests. The type of reliability was not established in the soccer tests reviewed, but the writer assumes that the method of stability was used. If the method of internal consistency was used the Brooks sampling would be comparable to those reported in other soccer skill tests. Based on this data it must be concluded that the McDonald Soccer Skill Test was not as reliable as the other soccer skill tests reviewed.

This test when given to Brooks Junior High School girls would not be considered a good test because of its low reliability which would also indicate that this test would not have a high degree of validity and therefore would not be a good test for determining a letter grade for soccer ability. It is possible that the McDonald Test could be useful as a practice technique and could be of some value in the teaching of soccer as shown in the comparison on an improvement basis of the 265 girls who took the test. Of these 167 improved their score, 55 did poorer and 43 remained the same. This is also indicated by the mean score improvement of the combined 7-8-9 grades of .9 kicks on the best of three trials and a 1.0 improvement on the four trial method.

In an analysis of the mean of each grade separately, some improvement was shown at each grade level, with the

eighth grade showing the greatest amount of improvement with 1.4 kicks and the lowest amount of improvement was noted in the seventh grade with a .5 kicks. One might logically expect the greatest amount of improvement to be shown at the seventh grade level.

The t test was used for this study to indicate any difference between the means. On all of the values of t obtained the results were significant at the .01 level of confidence. On the basis of the statistical data we can assume that chance factors did not cause this large of a difference. It might possibly result from the skill practice of taking the test the first time. In order to determine this resultant improvement, a series of studies would have to be carried out.

The chi square test was computed to test the hypothesis that the sampling of scores obtained had a distribution that was not significantly different from a normal curve. The .05 level of confidence was used for statistical significance. All of the chi square values exceeded the .05 level. Therefore, the hypothesis was rejected. Two possiblities for this might be; (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill of the Brooks Junior High School girls was not distributed normally. It would seem that the

sample obtained was large enough and widely varied enough, according to ability, to obtain a normal distribution for soccer ability. Therefore, the data indicates that the test did not discriminate between different levels of soccer ability.

CHAPTER V

SUMMARY AND CONCLUSIONS

Introduction. Chapter V will give a complete summary of the entire study of the McDonald Soccer Skill Test administered to Brooks Junior High School Girls, Wichita, Kansas, during the school year 1961-1962. Also the conclusions that have been arrived at as the result of this study will be discussed in this chapter.

Purpose. The purposes of this study were to, (1) determine the reliability of the McDonald Soccer Skill Test, (2) set up norms at Brooks Junior High School for the McDonald Soccer Skill Test, and (3) set up a grading scale for the girls at Brooks Junior High School that could be used after the unit on soccer.

Significance. The significance of this paper was to study the reliability of the McDonald Soccer Skill Test to be given to junior high school girls after the completion of the unit on soccer. As the physical educators today are faced with a challenging situation of saying that physical education is an important subject like any other in the curriculum, it must be graded like the others and full use of measurement must be made in day-to-day teaching. If a reliable grade is to be arrived at after a unit of a

particular sport, a skill test must be given and it must have acceptable reliability and validity as a test of this particular skill.

The need for this study was to find a test for soccer skill which did not take a great deal of time to administer and yet had valid and reliable results.

The McDonald Soccer Skill Test was given after four weeks of soccer participation in class. The same test was then repeated one week later. Each class was divided into four testing areas and all four areas took the test at the same time.

Delimitations. This study was delimited to the statistical analysis of the one item of the wall volley. The extent to which this skill test measures the skill of volleying in soccer was delimited to the validity of the McDonald test to measure the skill of volleying in soccer.

Limitations. This study was limited to the data collected by administering the McDonald Soccer Skill Test to 265 Junior High School Girls enrolled in physical education classes, during the school year 1961-1962, at Brooks Junior High School in Wichita, Kansas. The class enrollment at Brooks varies for physical education with some girls attending class two days per week, others three days and the remainder come five days per week depending on their other

class enrollment. This sample may or may not be representative of the skill of junior high school girls over a period of time.

History. The history of measurement work in physical education may be divided into five phases; (1) Anthropometric, (2) Strength, (3) Cardiac functional, (4) Athletic ability and (5) Single test or index figure. The first use of game elements as possibilities for testing in physical education came in the development of the Athletic Badge Tests by the Playground and Recreation Association of America in 1913. In 1916, Reilly introduced his program of rational athletics which was followed by Hetherington in developing the California decathlon. With the introduction of statistical methods around 1930, a number of sport tests were devised which are still usable today. These tests included both the classification and achievement types. During World War II the emphasis was placed on physical fitness which resulted in a wide variety of physical education activities. Evaluative procedures were developed in many new areas because of this trend.

Purposes and Criteria for Skill Tests. Tests of sport skills are usually given to determine the general ability in a particular sport or specific ability in one technique essential to successful performance. The significance of a

specific ability as a measure of success in playing the game is dependent upon its relative importance to all of the techniques required in the game. If tests are to serve the purposes we expect of them, they must possess certain qualities. Some of these qualities are; (1) be like game situations, (2) involve only one performer, (3) be interesting and meaningful, (4) differentiate between levels of ability, (5) provide accurate scoring, (6) be judged by statistical evidence and (7) provide a means for interpreting performance.

Certain points must be kept in mind when judging the value of a good skill test. Some of these points are (1) one test cannot possess all of the good characteristics, (2) a test can only sample the many skills involved in an activity, (3) effective results are dependent upon maximum effort of the participant, (4) the test administrator is the key to obtaining good results and (5) groups differ, so the norms and achievement scales must differ.

Selection of Test. The McDonald Soccer Skill Test was chosen for the purpose of this study as it contained, in the writer's estimation, most of the criteria for a good test. In most of the tests reviewed, the major disadvantage to the tests was the time element. The Bontz Test would have been one of the best tests to administer but it would have taken approximately two weeks to administer due to the large size of classes and number of times per week the girls at Brooks

Junior High School attend class. Some of the available tests used for testing soccer skills other than the McDonald Test are; (1) the Bontz Combination of Soccer Skills, (2) Shaufele Test, (3) Vanderhoff Ability Tests, (4) Heath and Rodgers Ability Test and (5) Buchanan Speedball Tests. According to Scott and French a good two-item battery of soccer tests should include the combination of passing and receiving with volleying. The one best test item for girls is volleying.

The McDonald Skill Test was chosen to be given at Brooks Junior High School for the following reasons; (1) small amount of time to be given, (2) contained the one best test item, that of volleying and (3) ease of administration.

Procedures. The subjects used for this study were 265 girls at Brooks Junior High School, Wichita, Kansas, during the school year 1961-1962, who were enrolled in physical education classes. There were 100 ninth graders, 80 eighth graders and 85 seventh graders used in this study.

The McDonald Test was given after four weeks of participation in soccer and the exact same test was then given one week later to determine the reliability of the test for junior high school girls.

Each class was divided into four groups and stationed at four different areas of wall space. In taking the test, each girl waited five minutes between each kicking trial of

the test. All girls were instructed on how to fill in the score card which is shown in Appendix D, with an explanation of what the scores meant.

A description of the test and a diagram of the floor plan for the test are found in Appendix A. The offical test was deviated from and on this test the line 11½ feet high from the floor was omitted as very few girls kicked the ball higher than this.

Results. In analyzing the data, individual raw scores were used. The mean, standard deviation, chi square, standard error, linear product moment correlation, calculation and interpretation of the difference between two means and standard score were the statistical measures used. The raw scores were first analyzed by using the total of all three grades. Then each grade was analyzed separately.

According to Table I for combined 7-8-9 grades, the mean of all the trials did change somewhat. On test I, trial 1, with a mean of 11.4 and on test II, four trials, 13.5 resulted in an improvement of 2.1 kicks. On the composite score of test I, four trial average, a mean of 12.3 was arrived at and on test II, four trials, a mean of 13.3 resulted in a 1.0 kick difference. On the best of three trials a .9 difference was found. Thus to determine which would be the most satisfactory method of averaging the total trials, the best of three would result in a higher average,

but very little difference was found in using four trials or the best of three.

In Table II for 7-8-9 grades listed separately, the seventh graders had the lowest mean while the ninth graders had the highest mean. Each grade improved on test II. The seventh grade had a .5 improvement on both the best of three and four trial averages. The eighth grade had a 1.1 improvement on the four trial average and a 1.4 on the best of three. The ninth grade had an .8 improvement on the four trial average and .7 on the best of three trial average.

The reliability of a test means that the test will repeat itself. With the same person giving the test, using the same methods with the same students, the same score will be obtained both times. In order for a test to be valid it must have some degree of reliability, but the opposite is not always true.

Reliability may be determined by three different methods. These are (1) stability which is accomplished by giving the same test twice with a time interval between administrations; (2) internal consistency which is accomplished by obtaining two test scores from the same test, and (3) equivalent forms, which are different forms of the same test but are as similar as possible, administered with a very short time interval between test administrations.

The correlation of test I, four trials and test I,

best of three trials for the combined 7-8-9 grades resulted in a r of .8095. On test II, four trials and test II, best of three trials for the combined 7-8-9 grades, a correlation of .7810 resulted. The correlation between the two tests was lower with a .6248 correlation on four trials and a .5715 on the best of three trials. Only a small difference resulted in four trials compared to three trials. There are no set standards for reliability coefficients to determine "high" degree of reliability or a "low" degree of reliability. This will have to depend on what the test scores are used for, that is, if they are reliable enough to be of some value.

In comparing the r of .8095 for test I, four trials and test I, best of three trials, and on test II, four trials and test II, best of three trials with a r of .7810 there was a relatively high degree of internal consistency. This was due to the fact the same scores were used except in the case of the best of three trials, the lowest score of the four trials was omitted.

The difference of the r between stability and internal consistency was larger than was to be expected of a test that is reliable; therefore, it must be concluded that the McDonald Test is not reliable for junior high school girls. This also indicates that this test could not be very high in validity for junior high school girls.

The t test was the method used to determine any

difference that did occur between the two means, regardless in which direction the difference occurred. In this study the .01 level of confidence was used as the level to be considered statistically significant. All of the values of t exceeded the value necessary for significance at the .01 level. Therefore, we can assume that chance factors did not cause the observed differences between test scores considered. It is possible that these significant values of t occurred as a result of the skill practice from taking the test the first time which could have caused the improvement in skill on the second trial of the test.

The chi square test as a statistical test for normality was computed to test the hypothesis that the sampling of scores obtained in this study were not significantly different from a normal curve. The .05 level of confidence was used as the level to be considered statistically significant. All of the chi square values obtained exceeded the value necessary for significance at the .05 level. Therefore we reject the hypothesis at the .05 level that the distribution on the McDonald Soccer Skill Test was not significantly different from normal on each of the grades and for combined grades 7-8-9. Two possibilities for these significant chi square values might be; (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range

of ability in soccer skill present in the sample of Brooks Junior High School girls was not distributed normally. It would have seemed feasible that the sample of 265 girls in grades 7-8-9 would have been large enough and widely varied enough, according to ability, to obtain a normal distribution. Therefore, it would indicate that the McDonald Soccer Skill Test did not discriminate between different levels of ability.

T-scores were set up for grading purposes for each individual grade and also for the combined 7-8-9 grades. The T-scores may be found in Table V. . Some caution must be used with these T-scores due to the significant chi square values obtained which indicated lack of normality in the distribution of test scores.

The letter grade scores that are given in Appendix B are in the proportion of 7%-A, 24%-B, 38%-C, 24%-D, and 7%-F in relation to the distribution of the test scores from the Brooks Junior High School girls sample.

In an interpretation of results, the comparison of the coefficients of correlation between stability and internal consistency were the methods used to establish reliability for the McDonald Soccer Skill Test. In relation to the other soccer skill tests studied the Brooks Junior High School sampling was lower than all of the other tests. The type of reliability was not established in the soccer tests reviewed, but the writer assumes that the method of stability

was used. If the method of internal consistency was used, the Brooks sampling would be comparable to those reported in other soccer skill tests. Based on this data it must be concluded that the McDonald Soccer Skill Test was not as reliable as the other soccer skill tests reviewed.

This test when given to Brooks Junior High School girls would not be considered a good test because of its low reliability which would also indicate that this test would not have a high degree of validity and therefore would not be a good test for determining a letter grade for soccer ability. It is possible that the McDonald Test could be useful as a practice technique and could be of some value in the teaching of soccer as shown in the comparison on an improvement basis of the 265 girls who took the test. Of these 167 improved their score, 55 did poorer and 43 remained the same. This is also indicated by the mean score improvement of the combined 7-8-9 grades of .9 kicks on the best of three trials and a 1.0 improvement on the four trial method.

In an analysis of the means of each grade separately, some improvement was shown at each grade level, with the eighth grade showing the greatest amount of improvement with 1.4 kicks and the lowest amount of improvement was noted in the seventh grade with a .5 kicks. One might logically expect the greatest amount of improvement to be shown at the seventh grade level.

The t test was used for this study to indicate any differences between the means. On all of the values of t obtained the results were significant at the .01 level of confidence. On the basis of the statistical data we can assume that chance factors did not cause this large of a difference. It might possibly result from the skill practice of taking the test the first time. In order to determine this resultant improvement, a series of studies would have to be carried out.

The chi square test was computed to test the hypothesis that the sampling of scores obtained had a distribution that was not significantly different from a normal curve. The .05 level of confidence was used for statistical significance. All of the chi square values exceed the .05 level. Therefore, the hypothesis was rejected. Two possibilities for this might be; (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill of the Brooks Junior High School girls was not distributed normally. It would seem that the sample obtained was large enough and widely varied enough, according to ability, to obtain a normal distribution for soccer ability. Therefore, the data indicates that the test did not discriminate between different levels of soccer ability.

Conclusions. The conclusions that have been derived

from the study of the McDonald Soccer Skill Test as administered to Brooks Junior High School Girls, during the school year 1961-1962 are:

1. All Differences obtained were significant when using the average of the four trials or the best of three trials.
2. The reliability correlation coefficient between test I, four trials and test II, four trials was .6248 and on test I, best of three trials and test II, best of three trials was .5715 and both are low for a reliable test.
3. All of the mean scores improved which could indicate the McDonald Test could be used for the improvement of volleying skill in soccer during the instructional period.
4. The letter grade scales did not show a large range of scores, as in most cases only 2 or 3 kicks would result in a grade change.
5. There are two conclusions possible as a result of the computation of the chi square test. They are: (1) the test was not able to discriminate between different levels of soccer ability in the sample according to a normal distribution or (2) that the range of ability in soccer skill present in the sample of Brooks Junior High School girls was not

distributed normally. It would seem logical that the sample obtained was large enough and widely varied enough, according to ability, to obtain a normal distribution of test scores for soccer ability. Therefore, the data indicates that the test did not discriminate between different levels of soccer ability.

6. The reliability of the McDonald Soccer Skill Test was low as shown in this study and would therefore limit the validity of this test for use with the Brooks Junior High School girls.
7. The ^{T}-scores that were prepared from this study must be used with some caution and only under certain conditions due to the findings of the significant chi square values which showed a departure from the normal curve.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bovard, John F., Frederick W. Cozens and E. Patricia Hagman. Tests and Measurements in Physical Education. Philadelphia: W. B. Saunders Company, 1949.

Brace, David K. Measuring Motor Ability. New York: A. S. Barnes and Company, 1927.

Clarke, H. Harrison. Application of Measurement to Health and Physical Education. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1959.

Cozens, Frederick W., Hazel J. Cubberley and N. P. Neilson. Achievement Scales in Physical Education Activities. New York: A. S. Barnes and Company, 1937.

Edwards, Allen L. Statistical Methods for the Behavioral Sciences. New York: Rinehart and Company, Inc., 1956.

Heath, Marjorie L. and Elizabeth G. Rodgers, "A Study in the Use of Knowledge and Skill Tests in Soccer," Research Quarterly, 3:33-53, December, 1932.

Larson, Leonard A. and Rachael Yocom. Measurement and Evaluation in Physical, Health, and Recreation Education. St. Louis: C. V. Mosby Company, 1951.

Lindquist, E. F. A First Course in Statistics. Cambridge, Massachusetts: The Riverside Press, 1942.

Mathews, Donald A. Measurement in Physical Education. Philadelphia: W. B. Saunders Company, 1958.

_____. Measurement and Evaluation Materials in Health, Physical Education and Recreation, Washington, D. C.: AAHPER, A Department of the NEA, 1950.

Moriarty, Mary J., "How Shall We Grade Them?", Journal of Health, Physical Education and Recreation, 25:27, January, 1954.

Ostle, Bernerd. Statistics in Research. Ames, Iowa, Iowa State University Press, 1960.

_____. New York State Physical Fitness Standards for Boys and Young Men, "A Manual for Young Men," and New York State Physical Fitness Standards, "Evaluative Procedures in

Physical Activities for Girls and Young Women," Albany, New York, New York State War Council, Office of Physical Fitness and New York State Education Department, Division of Health and Physical Education, 1944.

Remmers, H. H. and N. L. Gage. Educational Measurement and Evaluation. New York: Harper and Brothers, 1943.

Scott, Gladys M. and Esther French. Better Teaching Through Testing. New York: A. S. Barnes and Company, 1945.

Scott, Gladys M. and Esther French. Evaluation in Physical Education. St. Louis: C. V. Mosby Company, 1950.

Scott, Gladys M. and Esther French. Measurement and Evaluation in Physical Education. Dubuque, Iowa: Wm. C. Brown Company, 1959.

Snedecor, George W. Statistical Methods. Ames, Iowa: Iowa State University Press, 1959.

Vanderhoof, Mildred, "Soccer Skill Tests," Journal of Health and Physical Education, 3:42, 54-56, October, 1932.

Weiss, Raymond A. and Marjorie Phillips. Administration of Tests in Physical Education. St. Louis: C. V. Mosby Company, 1954.

Willgoose, Carl E. Evaluation in Health Education and Physical Education. New York: McGraw-Hill Book Company, 1961.

APPENDIX

APPENDIX A

McDONALD SOCCER SKILL TEST AND FLOOR PLAN

At the command of Ready, Go, the subject begins kicking the soccer ball against the wall for 30 seconds. Any type of kick may be used; both ground balls and fly balls which hit the wall count. To count, however, all balls must be kicked from the ground with the supporting leg behind the restraining line. Rebounds may be retrieved in any manner, including use of the hands. If a ball is out of control, the subject may play one of the spare balls, but must bring the ball by use of hands or feet to a position at the restraining line before kicking against the wall (no penalty other than the lost time in getting the ball back in position to kick). Spare balls are placed 9 feet behind the line and in the center of the testing area. The score is the number of legal kicks in 30 seconds. Four trials are permitted with two practice trials.

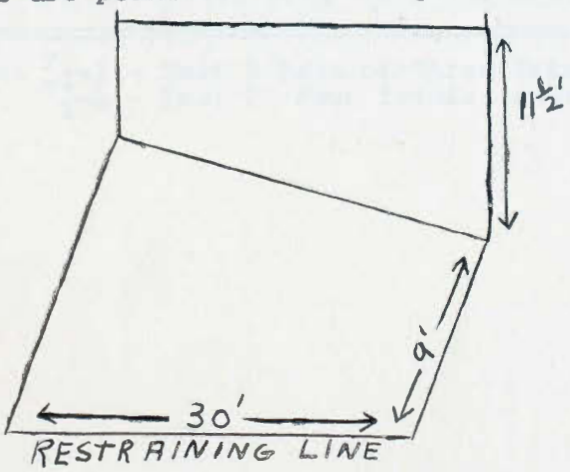


TABLE III

FREQUENCY DISTRIBUTION OF RAW SCORES FOR BROOKS JUNIOR HIGH SCHOOL SAMPLE, TEST I, BEST OF THREE TRIALS AND FOUR TRIALS

Combined 7-8-9 Grades			Seventh Grade		Eighth Grade		Ninth Grade	
Score	T_1-3	T_1-4	T_1-3	T_1-4	T_1-3	T_1-4	T_1-3	T_1-4
20	2	0	0	0	0	0	2	0
19	0	0	0	0	0	0	0	0
18	6	5	0	0	1	1	5	2
17	8	4	1	0	2	0	4	5
16	16	7	4	1	3	2	8	7
15	30	26	5	7	10	11	15	8
14	42	29	15	18	14	8	14	12
13	55	46	13	17	21	18	21	12
12	46	62	19	10	13	17	16	23
11	25	36	12	10	4	8	9	15
10	16	24	9	9	4	6	3	10
9	8	12	4	8	1	3	2	3
8	8	6	3	3	4	1	0	2
7	2	6	0	2	2	3	1	0
6	1	2	0	0	1	2	0	1

Note: T_1-3 - Test I Best of Three Trials, Averaged
T_1-4 - Test I Four Trials, Averaged

TABLE IV

FREQUENCY DISTRIBUTION OF RAW SCORES FOR BROOKS JUNIOR HIGH SCHOOL SAMPLE, TEST II, BEST OF THREE TRIALS AND FOUR TRIALS

	Combined 7-8-9 Grades		Seventh Grade		Eighth Grade		Ninth Grade	
Score	T_{2-3}	T_{2-4}	T_{2-3}	T_{2-4}	T_{2-3}	T_{2-4}	T_{2-3}	T_{2-4}
20	1	0	0	0	1	0	1	0
19	4	4	0	0	2	0	3	4
18	7	3	0	0	3	2	4	1
17	12	11	1	0	3	4	7	2
16	35	14	5	3	12	6	16	10
15	48	46	12	7	15	13	20	12
14	45	46	13	19	12	15	20	26
13	41	43	16	13	13	12	10	20
12	40	42	17	17	11	17	12	7
11	19	30	14	15	4	6	3	9
10	6	16	2	7	3	3	2	3
9	3	7	1	2	0	1	2	4
8	4	2	4	2	1	1	1	1
7	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	1

Note: T_{2-3} - Test II Best of Three Trials, Averaged
T_{2-4} - Test II Four Trials, Averaged

TABLE V

T-SCORES FOR COMBINED 7-8-9 GRADES AND EACH GRADE INDIVIDUALLY FOR BROOKS JUNIOR HIGH SCHOOL GIRLS ON THE McDONALD SOCCER SKILL TEST BY THE CUMULATIVE FREQUENCY METHOD

Combined 7-8-9 Grades		Seventh Grade	Eighth Grade	Ninth Grade
Raw Score	T-Score	T-Score	T-Score	T-Score
20			75	76
19	74		70	70
18	70		66	66
17	67	75	63	62
16	63	67	59	58
15	58	61	54	53
14	54	56	49	48
13	49	51	45	43
12	45	46	40	39
11	40	40	36	35
10	35	35	33	33
9	30	34	28	30
8	26	30	20	24
7	21			

APPENDIX B

LETTER GRADES FOR BROOKS JUNIOR HIGH SCHOOL GIRLS ON THE McDONALD SOCCER SKILL TEST

COMBINED 7-8-9

17 or above - A
15-16 - B
12-14 - C
10-11 - D
9 or below - F

SEVENTH GRADE

16 or above - A
14-15 - B
12-13 - C
10-11 - D
9 or below - F

EIGHTH GRADE

18 or above - A
16-17 - B
13-15 - C
11-12 - D
10 or below - F

NINTH GRADE

18 or above - A
16-17 - B
14-15 - C
12-13 - D
11 or below - F

Figure 1

FREQUENCY POLYGON FOR COMBINED 7-8-9 GRADES
ON THE McDONALD SOCCER SKILL TEST
(SOCCER KICKS IN 30 SECONDS)

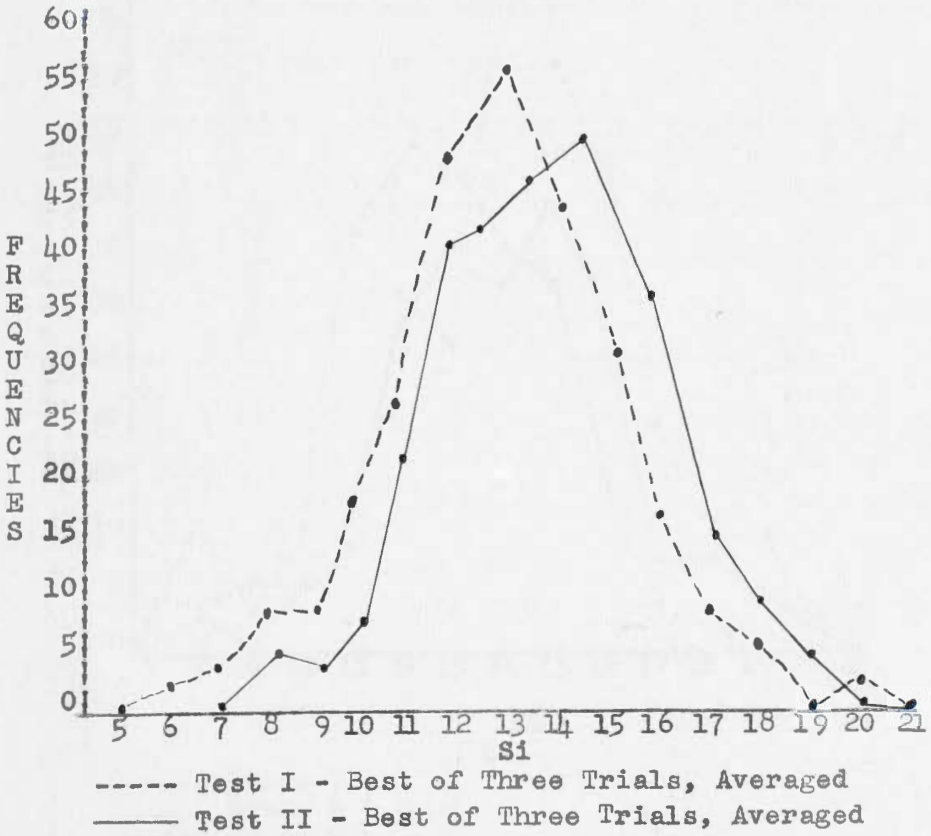
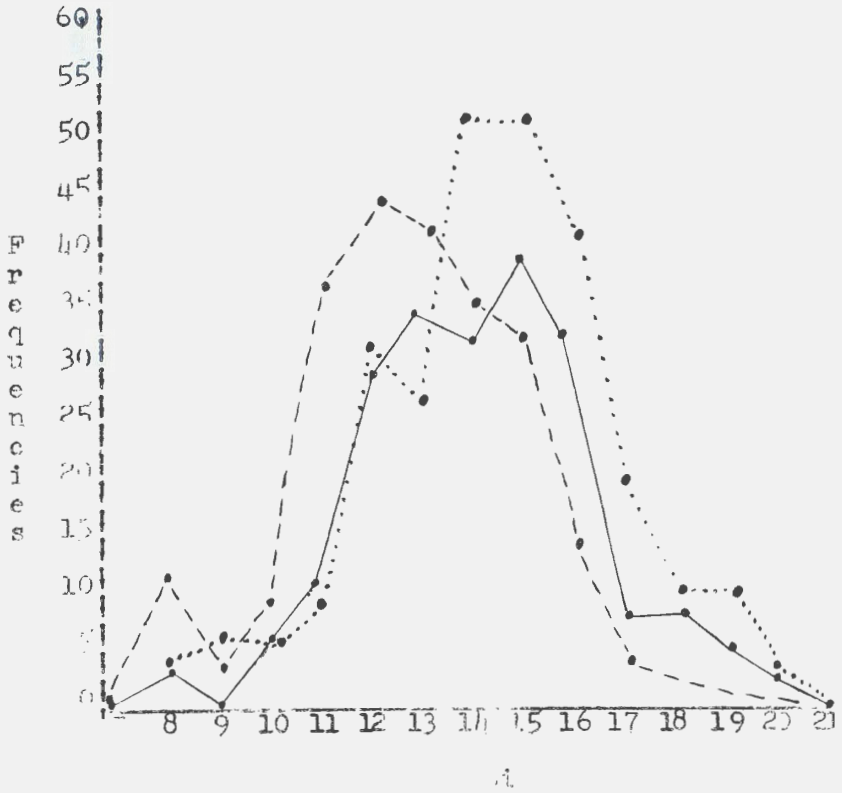


Figure 2

FREQUENCY POLYGON FOR GRADES 7-8-9 INDIVIDUALLY
TEST II - BEST OF THREE TRIALS
ON THE McDONALD SOCCER SKILL TEST
(SOCCER KICKS IN 30 SECONDS)



APPENDIX D

SCORE CARD FOR McDONALD SOCCER SKILL TEST AS ADMINISTERED AT BROOKS JUNIOR HIGH SCHOOL

SCORE CARD

NAME_____ GRADE_____

CLASS PERIOD_____

TEST I TRIALS	1	2	3	4

TOTAL _____

Average of Four _____

Average of Best Three _____

LETTER GRADE _____

TEST II TRIALS	1	2	3	4

TOTAL _____

Average of Four _____

Average of Best Three _____

LETTER GRADE _____