1-1-1970

The Reliability and Validity of the modified AAHPER Fitness Test as Applied to Freshmen College Males at the University of North Dakota

Paul O. Soderberg

THE RELIABILITY AND VALIDITY OF THE MODIFIED AAHPER FITNESS TEST AS APPLIED TO FRESHMEN COLLEGE MALES AT THE UNIVERSITY OF NORTH DAKOTA

by

Paul O. Soderberg

Bachelor of Science, Minot State College 1965

A Thesis

Submitted to the Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the Degree of

Master of Science

Grand Forks, North Dakota

January
1970

This thesis submitted by Paul O. Soderberg in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota is hereby approved by the Faculty Advisory Committee under whom the work has been done.

Wm. W. Balonchuk
Chairman

Robert D Clayton

John C. Williams

William Johnson
Dean of the Graduate School

ACKNOWLEDGMENTS

The writer wishes to extend his sincere gratitude and appreciation to Mr. William Bolonchuk for the many hours spent in helpful guidance and assistance in preparation and completion of this study; to Dr. John D. Williams for his advice on the statistical procedures used; to Dr. Robert Clayton for critical analysis and help in preparing the initial draft; and to Mr. Len Marti for the time taken as a member of the examining committee.

Mr. Theodore Keck gave valuable assistance in helping the writer complete the muscle analysis research for the criterion fitness test. Last, the writer wishes to express his thanks to the Computer Center at the University of North Dakota and the employees that helped analyze the data.

" other " on top (center) no( )

page 14 Chapter 3 (4 lines from bottom change to

Tables II, III, IV

number page of tables + graphs

Type Table of Contents page

" Bibliography + BIBLIOGRAPHY

center -

Small caps

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This thesis topic was basically concerned with two phases. The first phase of the study was to determine the reliability of the Modified AAHPER Fitness Test. The second phase of the study was to determine the validity of this same test. Test administrations were given in a one week period.

The reliability was established by the test-retest method. Thirty-seven subjects participated in two test periods. The validity was established by comparing nineteen subjects' results on a criterion test, to their results on the first administration of the Modified AAHPER Test Items.

All items of the Modified AAHPER Test were found reliable at the .01 level, but only three items (shot put, standing broad jump and the pull-up) were found to have high enough coefficients to be acceptable under certain standards. The pull-up item was not significant at the .01 level in the test for validity. The standing broad jump and 600-yard run-walk were found to have low validity coefficients. This suggests that future tests to validate measures of this kind should measure more closely the same criteria.

CHAPTER I

THE PROBLEM AND ITS SCOPE

Introduction

Because of modern innovations that are constantly replacing the tasks ordinarily done by man, the physical fitness level of society is becoming a major concern. With the present-day emphasis on physical fitness, physical educators and the public should be well-informed on the subject. A current method being used nationally to evaluate physical fitness is the American Association of Health, Physical Education, and Recreation Fitness Test, more commonly known as the AAHPER Fitness Test.

The American Association for Health, Physical Education, and Recreation has been deeply concerned with ways to improve the fitness of American boys and girls. One effective contribution has been the Youth Fitness Test Project initiated in 1958. This test was designed to measure the level of fitness of American youth. The test is composed of seven items. These items are the pull-up, sit-up, shuttle run, standing broad jump, 50-yard dash, softball throw, and the 600-yard run walk.

The University of North Dakota Department of Health, Physical Education, and Recreation uses a test similar to the AAHPER Fitness Test for evaluating the fitness level of incoming freshman males. The only difference between the AAHPER Test and the fitness test

administered at the University of North Dakota (the Modified AAHPER Fitness Test) is in one test item. The Modified AAHPER Test includes the shot put in place of the softball throw because of weather restrictions. The tests are constructed to include all the basic areas of fitness, mainly strength, endurance, and speed.

The AAHPER Fitness Test has grown increasingly popular in the past decade. Various articles and television commercials about this test are presented to the public daily. Because of the stress of this fitness test throughout the United States, it is important that the test be evaluated. The purpose of this study was to determine the reliability and validity of the AAHPER Fitness Test as it is administered at the University of North Dakota. Reliability was determined on a test-retest basis. Validity was determined by comparing each item of the Modified AAHPER Test with the items of a similar test. Since the Modified Fitness Test is nearly the same as the AAHPER Fitness Test, the results of this study will be helpful to physical educators. Because of its wide use, it is important that the test measures what it purports to measure. It is believed that the information derived from this study should be used to guide personnel involved in administering this test in their selection of the test and its results. Because of the amount of research available locally and nationally on this topic is small, there is a need for this study. The student and general public should benefit from the end results by learning the reliability, validity, and thus the actual effectiveness of the test. By carefully studying and testing each item of the test, actual knowledge of the reliability and validity of the test will be attained. By

knowing this information the administrator of the test will better be able to use the test in evaluating fitness.

Definition of Essential Terms

Concentric contraction: a contraction where the muscle length shortens.

Dynamic strength: the ability to exert muscular force repeatedly or continuously over time.

Eccentric contraction: a movement whereby the muscle length increases.

Explosive strength factor: the ability to expend a maximum of energy in one or a series of explosive acts. Not necessarily continuous as dynamic strength.

Motor fitness: tests that aim to measure the fitness of the body for strenuous work, and define as components strength, agility, speed, endurance, power, and the like.

Physical fitness: muscular and cardiovascular-respiratory capacity of the body in relation to a specific task.

Reliability: consistency or stability of measuring in a test-retest situation.

Endurance: the capacity to continue maximum effort, requiring prolonged exertion over time.

Prime movers: muscles mainly responsible for a certain action.

Shoulder girdle: consists of the clavicle and the scapula.

Shoulder joint: formed by the articulation of the humerus with the scapula.

Validity: a test is valid if it measures what it purports to measure.

Limitations

The following limitations must be taken into consideration when interpreting the results of this study.

1. No control was exercised over the sleep, diet, and daily habits of the subjects.

2. Exercise by the individuals outside of class was not taken into consideration.

3. Previous testing experiences of the subjects were not considered.

4. The degree of interest on the part of students may have affected results in some cases.

Delimitations

1. The study was limited to 37 male subjects chosen at random from a population of 320 males enrolled in freshmen physical education classes at the University of North Dakota.

2. The testing was administered in a one week period to insure that the subjects' physical condition did not change.

Review of Related Literature

The President's Council on Physical Fitness was set up by President Eisenhower in 1956 to give federal encouragement to programs aimed at increasing over-all fitness of American youth. The Council officially adopted and endorsed the use of the AAHPER Youth Fitness Test and recommended testing as the first step toward improvement of

physical fitness. The AAHPER Fitness Test was developed in 1957 and was revised in 1965. The reason for its development was the Association's interest in establishing a scientific basis for measuring the physical fitness of boys and girls. The AAHPER Test Items purport to measure a variety of criteria in determining an individual's physical fitness. The pull-up for boys designates measuring arm and shoulder girdle strength. The sit-up judges efficiency of abdominal and hip flexor muscles. The shuttle run judges speed and change of direction. The standing broad jump judges explosive muscle power of leg extensors. The 50-yard dash judges speed. The shot put item used in the Modified AAHPER Test judges explosive strength and coordination. The 600-yard run-walk judges cardiovascular efficiency.[1]

In order to establish reliability and validity for best performance, several procedures were investigated. Julian C. Stanley[2] reviewed the various types of reliability and validity. The three types of reliability are the coefficient of stability, the coefficient of equivalence, and the coefficient of equivalence and stability. The coefficient of stability is the test-retest coefficient which was used in this study. Stanley also defines concurrent validity as the type which best illustrates the validity used in this study. This type of validity basically means using different contents in two different tests and attaining approximately the same reliability. In

---

[1]American Association for Health, Physical Education, and Recreation, AAHPER Youth Fitness Test Manual (Washington, D. C.: AAHPER, 1965), p. 7.

[2]Julian C. Stanley, Measurement in Today's Schools (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964), pp. 157-63.

this study the validity was established by comparing the results of the Modified AAHPER Test with the results of a designed test. The criteria used in devising the test of validity will be explained in Chapter II.

H. Harrison Clarke[3] states that in determining the validity of tests, the physical educator should evaluate two elements: (1) the degree to which the criterion measure represents the quality being measured; (2) the amount of relationship shown between the test and the criterion.

In justifying reliability after it has been measured, Clarke suggests that physical tests should have a minimum reliability coefficient of approximately 0.90. Clarke also states that there are three major factors that influence the reliability of a measure in statistical works. They are: "the sample itself, the number of cases contained in the sample, and the variability of the distribution." Other factors directly related to the reliability and validity of a test are the type of items utilized and the administrative procedures followed.

Paul Hunsicker and George Grey[4] reported facts on human strength that will aid in determining the reliability and validity of some of the strength items. There is a difference between static strength and dynamic strength, and the relationship between the two is not high. The position of the body is also very important when administering strength tests. Another factor in strength test data

---

[3]H. Harrison Clarke, Application of Measurement to Health and Physical Education (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1959), pp. 29, 35, and 445.

[4]Paul Hunsicker and George Grey, "Studies in Human Strength," Research Quarterly, XXVIII (May, 1957), pp. 109-22.

is the failure to insist on precisely the same testing conditions. Slight differences in the administration of a test may alter the effective forces considerably and the data will reveal differences in test administration rather than differences in strength. By following certain specifications in all administrations of the test items, more dependable results will be attained.

Leonard A. Larson[5] also studied strength factors. He illustrated that dynamic strength is nearly three times more significant in predicting a composite index of motor ability than is static dynamometrical strength. Larson further states that arm strength as indicated by Roger's Strength Index contributes approximately 90% of the total 38% of the composite motor ability criterion. (The item with arm strength, chinning, is a dynamic strength element). These facts would support the use of some of the items of the Modified Test.

A number of comments have been made concerning the worthiness of the Modified AAHPER Fitness Test. Arthur A. Esslinger[6] states that the test does have a number of advantages that explain its popularity. It is simple to administer, does not require expensive equipment, and is not unduly time-consuming. The test items reflect specific qualities such as arm and shoulder girdle strength,

---

[5]Leonard A. Larson, "A Factor and Validity Analysis of Strength Variables and Tests with a Test Combination of Chinning, Dipping, and Vertical Jump," Research Quarterly, XI (December, 1940), pp. 82-96.

[6]Arthur A. Esslinger, "Perspective on Testing," Journal of Health, Physical Education, and Recreation, XXXI (September, 1960), pp. 36-37.

abdominal and hip flexor strength, speed, agility, and endurance. This specificity facilitates identification and thus correction of deficiencies.

Esslinger also states that some facts that physical educators not supporting the AAHPER Test have mentioned. These critics have pointed out that:

> This test does not measure physical fitness in its entirety; that its brevity and simplicity prevent it from being a completely adequate evaluative instrument; that the factors of muscular power and endurance are the ones most important; that the boys' form of the pull-up test does not measure arm strength adequately; that no validation of the test was attempted; that the norms are based on percentile scores and have certain limitations which require they be used with caution.

Florence P. Kendall[7] has also done critical research on the AAHPER Fitness Test. She states the criteria of good tests should be evidence of reliability and validity, ease of administration, economy of time, standardization of directions, and availability of norms. Also, tests should measure important abilities, be interesting and meaningful, be of suitable difficulty, and the factor of safety should be given careful consideration. Kendall states that the pull-ups and bar hang fall into the category of "too difficult." A good test should have no zero scores, but many physical education instructors could not even report on the results of the bar hang because well over half of the students scored zero. The sit-up purports to measure strength of abdominal muscles, but in the manner done, weakness of these muscles can go undetected. Hence the item may not be valid.

---

[7]Florence P. Kendall, "A Criticism of Current Tests and Exercises for Physical Fitness," Journal of American Physical Therapy Association, XXXXV (March, 1965), pp. 188-89.

It was also stated by Kendall that in establishing the norms for the AAHPER Test, none of the participants had a chance to practice for the test items. In administering the test to classes the procedure generally followed is to spend time training for the testing.

Edwin A. Fleishman,[8] in his Examiner's Manual For Fitness Tests, discusses the reliability of four of the items that are included in the Modified AAHPER Fitness Test. These are correlations obtained from two administrations of each test item to the same subjects. These coefficients reflect the extent to which repeated administrations of the test rank these high school students in the same order of performance. Sit-ups, which measure a trunk strength factor, have a reliability of 0.72. The shuttle run, defined as explosive strength, has a reliability of 0.85. Pull-ups, a dynamic strength factor, has a reliability of 0.93. The 600-yard run-walk, which is a type of cardio-vascular endurance, has a reliability of 0.80. It should be stressed that these reliabilities are based on the administration procedures recommended in this particular manual, which were modified to improve reliability. It can be seen that these coefficients tend to be high, indicating considerable stability for the functions measured.

Fleishman, in this same manual, also states some facts on the validity of these test items. The validity was established by correlating results to a similar test. This similar test was devised by Fleishman and the reliabilities were found to be high. The shuttle

---

[8]Edwin A. Fleishman, Examiner's Manual for the Basic Fitness Tests (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964), pp. 23-25.

run has a validity of 0.77, pull-ups 0.81, and sit-ups were found by Fleishman to be inadequate measures of dynamic strength. The testing of validity is primarily concerned with knowing and being able to define the components of fitness and determining how to measure them. This same criteria for validity was used to design the test of validity in this study.

In his over-all evaluation of the AAHPER Fitness Test, Fleishman[9] found, according to his criteria for statistical reliability, that several of these tests, individually, come out very well. However, a fitness battery cannot be evaluated solely in terms of individual tests; rather, the key issues are the comprehensiveness and efficiency of the tests "as a battery." According to these criteria, the seven AAHPER Test Items measure only three factors well. These are dynamic strength in the pull-up, explosive strength in the standing broad jump and the shot put, and stamina in the 600-yard run-walk. The trunk strength factor is measured only imperfectly by the sit-ups test, which was found to have a lower factor loading (validity) and reliability coefficient than leg lifts. Furthermore, the explosive strength factor is overemphasized in the AAHPER battery, since four of the seven tests fall under this factor. Measures of static strength, extent flexibility, dynamic flexibility, gross body coordination, and gross body equilibrium are not covered by the AAHPER battery. Therefore, Fleishman concluded that the AAHPER Test does not measure physical fitness as he defines it.

---

[9]Edwin A. Fleishman, The Structure and Measurement of Physical Fitness (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964), p. 150.

McCloy and Young[10] have also done studies of reliability on items similar to those found in the AAHPER Test. They found the thirty-foot shuttle run has a validity rating of 0.829 and a reliability of 0.932 based on three trials. However, the shuttle run item of the AAHPER Test has only two repeats. They also stated that for purposes of research, tests of "breaking" strength and tests of "active" contraction of specific muscle groups as they are related kinesiologically to the endurance under consideration are recommended. Reliability ratings of tests may be increased both by the increase of the number of competent raters and by an increase in the number of relevant items being rated. The validity for a rating is increased if the raters are given instruction and practice in how to use the selected rating device, and if they discuss their ratings with one another. The validity of a rating is frequently lowered by what is called the "halo effect." This means placing the validity items in the nearly perfect category. These facts were considered in the administration of the test items in this study.

Arthur E. Field[11] studied a topic concerning a reliability analysis of the AAHPER Youth Fitness Test Items. This basically was involved with determining if practicing the items helped an individual's score on the test. Fifty-seven male University of British

---

[10]Charles H. McCloy and Norma D. Young, Tests and Measurements in Health and Physical Education (New York: Appleton-Century-Crofts, Inc., 1954), pp. 27, 28, and 80.

[11]Arthur E. J. Field, "A Reliability Analysis of the American Association for Health, Physical Education, and Recreation Youth Fitness Test Items," (Unpublished M.P.E. thesis, University of British Columbia, 1964). Cited in Completed Research in Health, Physical Education, and Recreation, VII (1965), p. 41.

Columbia students took the AAHPER Test four times in four weeks. An analysis of variance was prepared for each item and Fisher's t ratios and correlation coefficients were calculated for all possible combinations between trials. The results indicated the amount of practice necessary to raise the reliability of individual items to a satisfactory level. Standard errors of measurement were reported for each item. Slight changes occurred from the first test administration to the last.

The International Committee on Standardization of Physical Fitness Tests[12] has proposed some physical fitness measurement standards that will aid the reliability and validity ratings of a test. This committee is concerned primarily with methodology of measurements and analysis of scientific data. Some factors concerning this study deserve mentioning. It is very probable in the endurance run test that the scores would vary according to the amount and intensity of vocal encouragement given while the run is in progress, and pacing by another runner or runners would almost certainly affect the times. In running events, stop watch timing, even when carefully done, is subject to ccnsiderable error. Attention, interest, and effort may all contribute to results. Special importance must be placed on the degree and amount of external motivation from the tester on others. Also, the subject must be fully acquainted with the details of the test technique and understand clearly the objectives of the test program.

---

[12]International Committee on Standardization of Physical Fitness Tests, Tentative Physical Fitness Measurement Standards, First Report on Physical Fitness Measurement Standards prepared at the Sandefjord, Norway 1966 Meeting.

Montoye, Cunningham, Marmis, and Kozar[13] have all worked with the reliability of the multitrial items of the AAHPER Youth Fitness Test. The purpose of their investigation was to study the test-retest reliability of the AAHPER Test with a view to obtaining evidence concerning the appropriateness of the prescribed number of trials. The youth fitness test was administered to 2,060 Michigan school children, ages 9-18. The test was administered carefully. For each of the multiple trial items (standing broad jump, shuttle run, and the 50-yard dash), the differences between means of the trials were tested for significance and correlation coefficients for each pair of trials were computed. The calculations were done for each age and sex separately. The results indicated that the number of trials in the test items should be changed as follows:

Standing broad jump: Instead of 3 trials, 2 would be sufficient.

Shuttle run: Instead of 2 trials, at least 3 are necessary.

50-yard dash: Two, as prescribed are necessary.

These results show that the reliability and validity of the AAHPER Test could be improved by better test procedures.

Pollock, Rothermel, and Cureton[14] studied the AAHPER Fitness Test score changes as a result of an eight-week sport and physical

---

[13]Henry J. Montoye, David A. Cunningham, Cary Marmis, and Andrew J. Kozer, "Reliability of Multitrial Items of the AAHPER Youth Fitness Test," Abstracts of Research Papers 1968 (Published by AAHPER, 1968), p. 71.

[14]Michael L. Pollock, Bradley L. Rothermel, and Thomas K. Cureton, "AAHPER Fitness Test Score Changes as A Result of an Eight-Week Sport and Physical Fitness Program at the University of Illinois," Abstract of Research Papers 1967 (Published by AAHPER, 1967), p. 82.

fitness program at the University of Illinois. The purpose of this investigation was to determine the effects of a program of gymnastics, swimming, track and field, conditioning activities, and endurance running on fitness performance as measured by the AAHPER Youth Fitness Test Battery. Data were collected on 65 boys, ages 7 - 14, who were enrolled in the University of Illinois summer fitness school (1966) and 22 boys from the same age group who acted as controls. The data were collected in the initial and final weeks of an 8-week program in accordance with the directions in the manual, and precautions were taken to ensure that each test was administered to both groups by the same examiner and on the same day. Test-retest data were collected for each of the seven items and reliability coefficients were calculated. Based on the results of this investigation, the following conclusions were cited:

1. The changes which transpired in items purporting to measure muscular strength and endurance, power, and cardio-respiratory endurance were significantly greater for the boys in the organized program.

2. No changes appeared for either group in the items which purport to measure speed, agility, and coordination.

From this study it can be noted that the subjects participating in organized training programs can increase their scores on the AAHPER Fitness Test. However, the testing periods set up for this thesis study were scheduled close together and scores were not affected by the subjects' condition.

Doolittle and Bigbee[15] have conducted research on the effectiveness of the 600-yard run-walk item. Experience with the 600-yard run-walk led many investigators to question its accuracy. It was reported in the results by Doolittle and Bigbee that the correlation between the 600-yard run-walk and the maximum oxygen intake test (a proven test for cardio-vascular efficiency) was 0.62. Because of this low correlation the twelve-minute run-walk, a test of cardio-respiratory fitness, was evaluated and compared with the 600-yard run-walk. The study was conducted by using 153 ninth grade boys. A test-retest of the twelve-minute run-walk, which yielded a correlation coefficient of 0.94, indicated that it has a high reliability. A correlation coefficient (0.90) between the twelve-minute run-walk and the maximum oxygen intake indicates its validity. Both of these values were found by Doolittle and Bigbee. Their final results showed that the twelve-minute run-walk was more valid and reliable than the 600-yard run-walk in testing cardiorespiratory fitness.

Although many suggestions have been made to improve the administration of the AAHPER Test, these suggestions were not followed in this study. This study is an evaluation of the Modified AAHPER Test and the improved methods of administration stated previously merely aided the writer in his evaluation procedures. The results attained in this study directly apply to the methods of the Modified AAHPER Test as it is administered at the University of North Dakota. The reliability and validity coefficients attained

---

[15]T. L. Doolittle and Rollin Bigbee, "The Twelve-Minute Run-Walk: A Test of Cardiorespiratory Fitness of Adolescent Boys," Research Quarterly, XXXIX (October, 1968), pp. 491-95.

in the study were used to determining the worthiness of the test. The research concerning the criteria for developing the designed test for validity is in Chapter II.

Summary

From the review of related literature various conclusions can be made. The more important conclusions are:

1. The type of reliability used in this study is the coefficient of stability.

2. The type of validity used in this study is content validity.

3. Validity is determined by the degree to which the criterion measure represents the quality being measured, and the amount of relationship shown between the test and the criterion.

4. Three factors that influence reliability are the sample itself, the number of cases contained in the sample, and the variability of the distribution.

5. Body position is very important when administering strength tests and administrative procedures must be precise.

6. Dynamic strength is nearly three times more significant in predicting a composite index of motor ability than is statis dynamometrical strength.

7. The pull-up item is "too difficult" and the sit-up purports to measure strength of abdominal muscles, but in the manner done, weakness of these muscles can go undetected.

8. Fleishman found that the sit-up is an inadequate measure of dynamic strength.

9. The seven AAHPER Test Items only measure three criterion well, according to Fleishman. They are dynamic strength, explosive strength, and stamina.

10. Tests of "breaking" strength and tests of "active" contraction of specific muscle groups are recommended for research studies.

11. Certain amounts of practice are necessary to raise the reliability of individual items of the AAHPER Test significantly.

12. Stop watch timing is subject to considerable error and considerable importance must be placed on the degree and amount of external motivation.

13. Research shows that the number of trials in the multi-trial items of the AAHPER Test should be changed.

14. The twelve-minute run-walk is more valid and reliable than the 600-yard run-walk in testing cardiorespiratory fitness.

15. The Modified AAHPER Test was administered as directed in the AAHPER Test Manual to enable the writer to establish the reliability and validity of that test.

CHAPTER II

METHODOLOGY

The first phase of this study was to determine the reliability of the Modified AAHPER Test by the test-retest method. The second phase was to establish validity. This was done by designing a test with the same criteria as the AAHPER Test and correlating the results of the two tests.

The experimental group in this study was selected from a population comprised of freshmen males enrolled in the physical education service program at the University of North Dakota. Sixty subjects were selected by a random method from a population of 320. Names of the 320 males were written on cards. The cards were shuffled and spread over a floor area. Sixty of the cards were then selected randomly from the 320 without replacement. All subjects were between the ages of 18 and 20 inclusive.

Names of the subjects selected for the sample were distributed to all classes involved and the testing procedures and purposes were explained to these subjects. All subjects were stimulated to participate by being offered two cuts in their respective classes. Other than this motivational device, participation in the testing was voluntary. Approximately 12% of the total population participated in the test-retest administration of the Modified AAHPER Test. Only 6% of the total population took part in the criterion test. The low

percentage of participation can be accounted for by interfering schedules, illness, and motivational factors. Nineteen subjects participated in all three test periods, 37 participated in both administrations of the Modified AAHPER Test, and 13 other subjects participated in only one AAHPER Test. These 13 scores were not used in the study because they could not be compared to other test scores.

The tests were administered by nine test assistants, each of whom had previously administered the Modified AAHPER Test. Six of the nine testers were currently enrolled in a test and measurement class and the remaining three were graduate assistants enrolled in physical education at the University of North Dakota. All the testers were present to administer their respective items at all three test periods to insure consistent results in the testing procedures.

The test periods for establishing the reliability and validity of the Modified AAHPER Test Items were all administered in a one week period. The times are further discussed in administration. This was done to insure the fact that the subjects remained in the same physical condition. Stiffness or other affects on the subject did not interfere with the results. Other pertinent factors involved in the test administration are stated later in this chapter.

The Modified AAHPER Fitness Test was administered as described in the AAHPER Youth Fitness Test Manual.[16] This description includes the purpose, equipment needed, rules, and scoring procedures for each test item except the shot put. The shot put item will be fully explained here.

---

[16]AAHPER, op. cit., pp. 16-22.

The purpose of the shot put item is for judging coordination and explosive power.[17] The equipment needed for the shot put item was a twelve-pound shot put, marking stakes, a tape measure, and a throwing circle with a diameter of seven feet. The Fieldhouse at the University of North Dakota was suited for the administration of the test item. The subject threw the shot put while remaining in the required circle. The point of landing was then marked with a small stake. If the subject's second or third throw is farther, the stake was moved accordingly so that after three throws the pupil's best was recorded. The best distance was recorded immediately after the third throw to eliminate any errors. Rules that were followed:

1. If the subject stepped out of the circle, the throw was not counted and was not repeated.

2. Three puts were allowed.

3. The subject had to put the shot. No under or overhand throwing was permitted.

4. In throwing the subject could take a one step, step-across, but was instructed to use the same procedures in all tests.

5. The distance was measured from the center of the throwing ring to the stake for ease in measuring. Scoring for the shot put was done by recording the best of three tries to the nearest foot and inches.

The criterion test items were selected according to the criteria that their corresponding items on the Modified AAHPER Fitness Test

---

[17]Payton Jordan and Bud Spencer, Champions in the Making (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1968), p. 127.

purported to measure. A complete description of the designed fitness test including purpose, equipment, rules and scoring procedures for each item follow. Each item is accompanied by an illustration.

1. Pull-up on the universal gym:

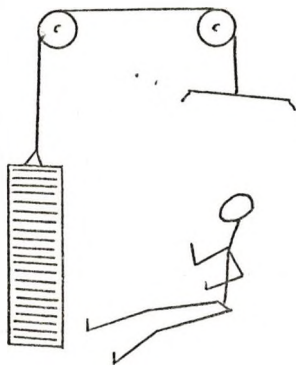

Diagram A.

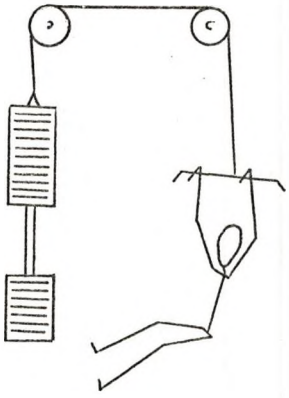

Diagram B.

Purpose: To judge arm and shoulder girdle strength.

Equipment: Scale to weigh the subjects; the universal gym at the University of North Dakota Gymnasium.

Description: The subject was first instructed to weigh himself. Three-fourths of the subjects weight was placed on the resistance part of the universal gym (located at the left side of the diagrams above). Three-fourths of the subjects weight would insure that everyone would be able to complete at least one repetition. The subject will sit on the floor with legs in a spread eagle position. The pull-up bar was then lowered to the subject and was grasped with a pronated grip (Diagram B). The testers hold the subject's legs down so that his body does not rise with each repetition. The subject was instructed to hold his back erect and pull the bar down below chin level using only arm and shoulder strength. The subject does as many repetitions as possible.

Rules: 1. Allow one trial unless it is obvious that the subject has had an unfair chance.

2. The only body movement should be in the arms. If any other movement was used to pull the bar down the repetition was not counted.

Scoring: Record the number of correctly completed pull-ups to the nearest whole number.

2. Inclined sit-up:

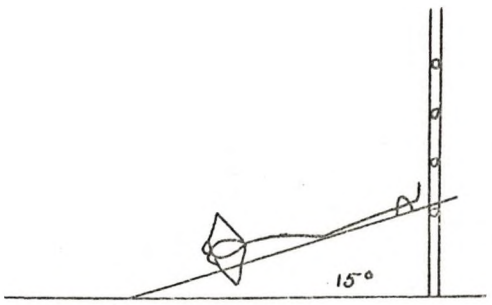

Diagram C.

Purpose: To judge the efficiency of abdominal and hip flexor muscles.

Equipment: Inclined board and wall ladder.

Description: The inclined board is placed on the lowest bar of the wall ladder (approximately 15° incline). The pupil lies on his back on the inclined board and inserts his feet under the holding strap at the top of the board (Diagram C). His hands are placed on the back of the neck with the fingers interlaced. Elbows are retracted. The subject sits up, turning the trunk to the left and touching the right elbow to the left knee, returns to starting position, then sits up turning the trunk to the right and touching the left elbow to the right knee. The exercise is repeated, alternating sides.

Rules: 1. The fingers must remain in contact behind the neck throughout the exercise.

2. The knees must be on the board during the sit-up but may be slightly bent when touching the elbow to the knee.

3. When returning to starting position, elbows must be flat on the board before sitting up again.

Scoring: One point is given for each complete sit-up. No score should be counted if the sit-up is not done correctly. Subjects were instructed to do as many repetitions as possible.

3. Shuttle run:

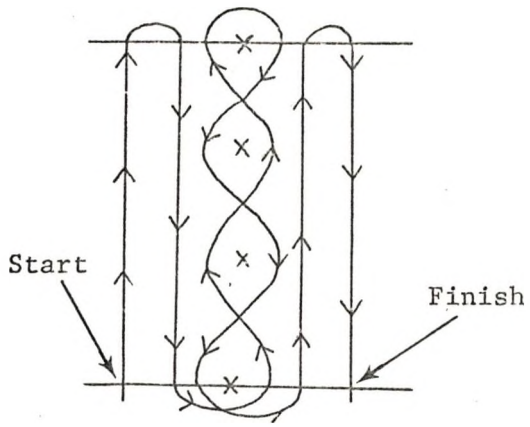

Diagram D.

Purpose: To judge speed and change of direction.

Equipment: Four chairs and a stop watch.

Description: In the shuttle run item chairs are placed at the starting line and at ten, twenty, and thirty foot intervals (the chairs are designated by x's in Diagram D). On the signal "Ready?" "Go!" the subject runs the length of the course (30 feet), steps over the line and returns to the starting point. He then continues by going around the chair at the starting point and weaves in and out of the chairs going around the last chair and returning to the starting line by weaving back through the chairs. The subject completes the test by returning to the 30-foot line, stepping over it and running back across the finish line. The path that the subject follows is designated by the arrowed line in Diagram D above.

Rules: Allow two trials with some rest between.

Scoring: Record the time of the better of two trials to the nearest tenth of a second.

4. Vertical Jump:

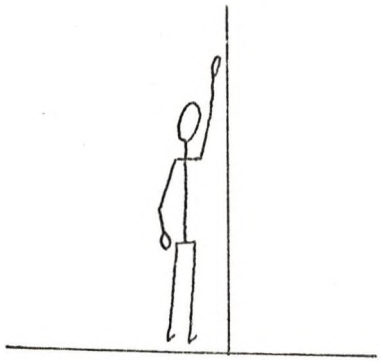

Diagram E.

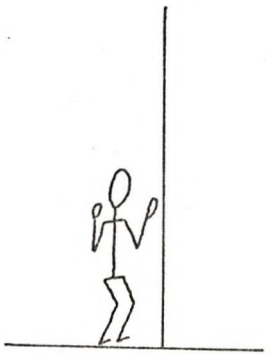

Diagram F.

Purpose: For judging explosive muscle power of leg extensors.

Equipment: An erasing towel, chalk, and a marked surface to jump and mark on that has been measured to the nearest inch.

Description: In the side of the standing vertical jump the subject stands with either right or left foot and shoulder against the wall (Diagram E). With chalk in hand (placed between the index and middle fingers at the first joint) the subject reaches as high as possible on the marking board and places a mark. The subject then jumps and places another mark on the board at the highest possible point holding the chalk in the same manner described. Before jumping the feet are placed at a natural distance apart and the arms are in a swinging motion. The body is flexed at the knee and hip joints and no other movement was permitted (Diagram F). The jump was accomplished by simultaneously extending the knees and swinging the arms upward and then placing a mark at the highest point.

Rules: 1. Three trials were allowed.
2. Subjects were not permitted to step into the jump.
3. Subjects must be in contact with the wall surface when placing the initial mark.

Scoring: Subtract the initial mark from the highest of the three trials and record to the nearest inch the best effort.

5. 50-yard dash with a running start:

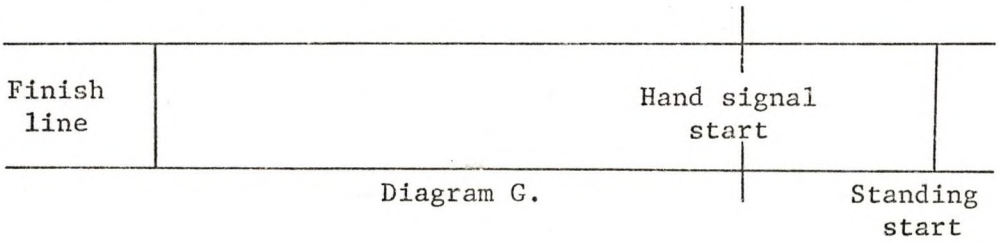

Diagram G.

Purpose: To judge speed.

Equipment: Stopwatch and track area.

Description: The test was administered to one subject at a time. The subject started the race ten yards behind the starting line using the type of start most comfortable for him. He starts at his own will and tries to reach his maximum speed in the ten yard distance to the starting line. When the runner passes the starting line the starter drops his arm from above his head. The stopwatch is started as soon as the timer sees the starter's arm move downward and stopped when the subject crosses the finish line. Diagram G above shows the lay-out of the track area used.

Rules: The score is the amount of time between the starter's signal and the instant the subject crosses the finish line.

Scoring: The time was recorded in seconds to the nearest tenth of a second.

6. Medicine ball put:

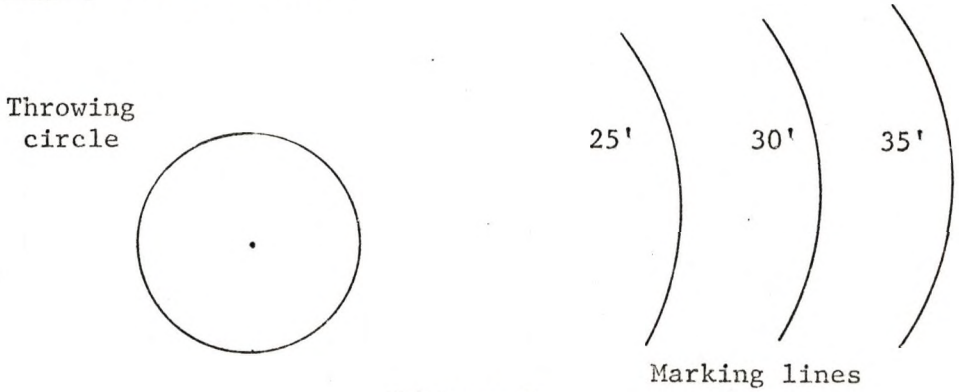

Diagram H.

Purpose: To judge coordination and explosive power.

Equipment: Medicine ball (2 feet 5¼ inches in circumference and weighing 8 pounds), marking stakes, a tape measure, and a throwing circle with a seven foot diameter.

Description: The shot put area in the University of North Dakota Fieldhouse was used for the test area (the area is shown in Diagram H on previous page). The subject was instructed to throw the ball in the same manner that the shot put was thrown. The subject remained in the circle throughout the throw or the throw was disqualified. The point of landing was marked after the first throw. If the second or third throw was farther, the stake was moved accordingly so that, after three throws, the stake is at the point of the subject's best throw.

Rules: 1. Three puts were allowed.
2. The medicine ball could not be thrown. Only the put was permitted.
3. The distance was measured from the center point of the circle to the stake.
4. Subjects stepping out of the circle and throwing illegally were disqualified for that particular throw.

Scoring: The best of the three trials was recorded to the nearest inch.

7. Twelve--minute run walk:

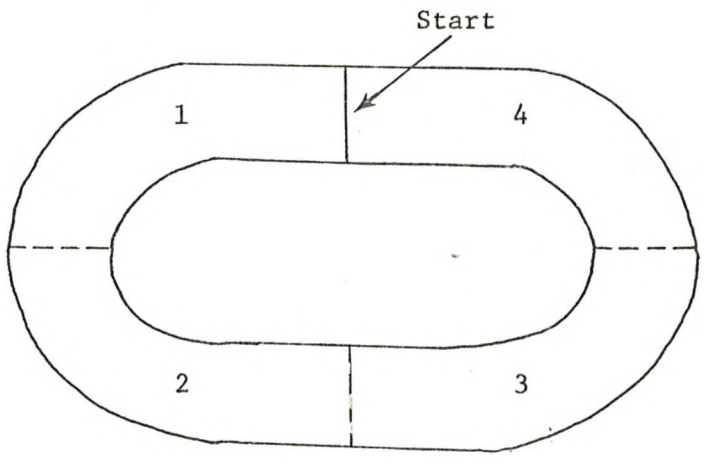

Diagram I.

Purpose: To judge coordination and explosive power.

Equipment: Stopwatch and the track area in the University of North Dakota Fieldhouse marked into four equal divisions as shown in Diagram I on previous page.

Description: The subject used a standing start. At the signal "Ready?" "Go!" the subject starts running. The subjects were instructed to cover as much distance as possible in a twelve minute period. Subjects were also instructed to increase pace during the final minute, if they were able, and that it was essential to put forth their best effort.

Rules: If walking was necessary, the instructions were not to walk more than one-fourth of a lap at a time. The subjects were instructed to stop on the sound of the whistle.

Scoring: The distance in laps was recorded to the nearest one-fourth lap.

The selection of the criterion items for testing validity was determined by a muscle analysis and discussion. The pull-up, sit-up, standing broad jump, and shot put items were selected by describing the muscles involved in these particular items. Mr. Arnold Keck,[18] a physical therapy instructor at the University of North Dakota, aided the writer in describing the muscle analysis. In order to support this muscle analysis the text by Rasch and Burke,[19] Kinesiology and Applied Anatomy, was used. The items used for testing the validity of the shuttle run, 50-yard dash, and the 600-yard run-walk were justified through general discussion with some supporting quotes.

The pull-up item of the Modified AAHPER Test is designed to judge arm and shoulder girdle strength. The pull-up on the universal

---

[18]Arnold W. Keck: Currently trainer and instructor of physical therapy at the University of North Dakota.

[19]Philip J. Rasch and Roger K. Burke, Kinesiology and Applied Anatomy (Philadelphia: Lea and Febiger, 1967), pp. 164, 181, 196, 225, 252, 269, 293, 319, and 345.

gym was selected as the item for testing validity because of the similarity in the shoulder-arm action. The joint movements present in the AAHPER pull-up are located at the shoulder girdle, shoulder joint, elbow joint, and wrist. The exercise starts with the shoulder joint flexed, the shoulder girdle elevated, and the elbow and wrist joints extended. The primary movements of the exercise are shoulder girdle depression, shoulder joint extension, and elbow and wrist flexion. All the joint actions are concentric contractions when the body is pulled up and eccentric when the body is let down.

The prime movers in shoulder girdle depression are the subclavius, the pectoralis minor, and the trapezius. The shoulder joint is extended primarily by the pectoralis major, the latissimus dorsi, and the teres major. Elbow flexion involves the biceps brachii, the brachialis, and the brachioradialis contracting. Wrist flexion primarily involves the flexor carpi radialis and the flexor carpi ulnaris contracting. The flexors of the hand are also involved, but will not be discussed because of their specificity.

The criterion item involves the same muscle and joint actions that are described above. The two most prominent differences between the tests are body position and gravity. In the criterion item the subject is seated on the floor with two spotters holding him in this position. This eliminates any other body movements which might aid in completing a repetition, such as kicking the legs. The AAHPER Item requires the subject to return to the starting position by letting gravity gradually pull his body down to full extension. In the designed item, the weight of the resistance (weight of the universal

gym) pulls the subject's arms up to the starting position. The contraction of muscles in both items are concentric in flexion and eccentric in extension.

The sit-up criterion item was done with the same test instruction as the sit-up item in the Modified AAHPER Test. The only difference between the two items was that the criterion item was administered on a board inclined at a 15 degree angle. The two basic body movements involved in this test, trunk and hip flexion, are the same for both items. Thoracic and lumbar spine flexion have the rectus abdominus, external oblique, and internal oblique as their prime movers. The prime movers in hip flexion are the psoas, iliacus, rectus femoris, and pectineus. All of the muscles listed here have concentric movements when the subject is curling up for the sit-up and eccentric action when the subject reclines.

The inclined sit-up involves a greater distance of movement to overcome gravity than the level sit-up. In the level sit-up gravity helps to flex the hip when the body is at a 90 degree angle to the floor. The inclined sit-up involves 15 more degrees of movement before the upper body is perpendicular to the floor. Thus, in performing the inclined sit-up the muscles for hip flexion work over 15 more degrees of movement before gravity aids the contraction than in the level sit-up. The reason for fewer repetitions in the inclined sit-up is that the muscles have to work over a greater range of motion before gravity aids the movement than they have to in the level sit-up.

The shuttle run used in the AAHPER Test comprises a standing start, speed, and the ability to change direction; the devised item includes these criteria even though it is administered in a different manner. The basic difference between the two test items is that the AAHPER Item includes picking up the two blocks whereas the criterion item does not, and the criterion item utilizes a "weaving" action that was included to further test the subject's ability to change direction. The same muscle groups and skills are basically involved in both test items.

The standing broad jump and the vertical jump judge the explosive power of the leg extensors. The same joint actions and muscle contractions are present in both items. The primary joints involved are the shoulder, trunk, hip, knee, and ankle joints. The primary movers in shoulder flexion are the deltoid and the pectoralis major. Trunk extension utilizes the erector spinal group and the deep posterior spinal group as its prime movers. Hip extension consists mainly of the contraction of the gluteus maximus, biceps femoris, semitendinosus, and the semimembranosus. The prime movers in knee extension are the rectus femoris, vastus lateralis, vastus intermedius, and vastus medalis. The ankle joint is plantar flexed by the contraction of the gastrocnemius and the soleus. Both test items involve the preceding muscles with their concentric contractions. It should be noted that the only difference between the two test items is that the standing broad jump involves angular velocity and the vertical jump vertical velocity.

In devising the item similar to the 50-yard dash a speed item free from the subjects error in starting was desired because it

minimized the error caused by this factor. The 50-yard dash with a running start eliminates the necessity of the subjects ability to start consistently in each test administration. The timing in both items was done by starting the watch on a hand signal. Use of the 50-yard dash as an item for testing speed on fitness tests is supported by Larson and Yocom[20] when they state that "any of the runs from 30 yards to 100 yards may be used as an index of speed."

The devised item used to test the validity of the shot put was the medicine ball put. This item was chosen because it may be executed in exactly the same manner as the shot put. The basic differences between the two items were the weight and size of the balls. The ability to apply maximum force, as stated by Jordan and Spencer,[21] depends not only upon strength, but also upon speed, for the distance a shot will travel depends on the velocity the ball attains on release. Therefore, the body levers should be coordinated so that each makes a maximum contribution to speed. In analyzing this statement it is seen that the components measured as explosive power and coordination.

The prime movers are the same in both test items. The shoulder girdle is abducted by the pectoralis minor and the serratus anterior. The shoulder joint is horizontally abducted with internal rotation by the pectoralis major, the latissimus dorsi, the teres major, and the subscapularis. Elbow extension is created by the contraction of the triceps brachii. Wrist and hand flexion primarily involves the

---

[20]Leonard A. Larson and Rachael D. Yocom, Measurement and Evaluation in Physical, Health, and Recreation Education (St. Louis, Mo.: The C. V. Mosby Co., 1951), p. 170.

[21]Jordan and Spencer, op. cit., p. 127.

flexor carpi radialis and the flexor carpi ulnaris. Trunk rotation from right to left involves the right external oblique and the left internal oblique muscles. Trunk extension comprises the contraction of the erector spinal group, the deep posterior spinal group, and the semispinalis capitis and cervicis. Hip and knee extension and plantar flexion of the ankle have the same primary movers that were stated in the jumping items. These muscle and joint actions involve concentric contraction of the prime movers stated above. All muscle groups were involved in both test items.

In selecting a validity item for the 600-yard run-walk, an item proven to test cardiorespiratory fitness was chosen. This item was the twelve minute run-walk. Doolittle and Bigbee[22] stated on the basis of their investigation, that the distance an individual can cover, by running and/or walking, in twelve minutes is a highly reliable and valid indicator of his cardiorespiratory fitness, and further, that it is more valid and reliable in this respect than the 600-yard run-walk test. Since both purport to measure cardiorespiratory fitness in a similar manner this is a suitable item for determining validity. Statistical evidence confirming the validity of this item were stated in Chapter I.

Test Administration

The Modified AAHPER Fitness Test and the devised test were both administered in the facilities provided at the University of North Dakota Department of Health, Physical Education, and Recreation. The shuttle run items and the jumping items were administered in

---

[22]Doolittle and Bigbee, op. cit., p. 494.

Gymnasium I. The sit-up and pull-up items were administered in Gymnasium II. The designed pull-up and the inclined sit-ups were administered in the Weight Training Room. The running items, the shot put, and the medicine ball put were all administered in the Fieldhouse. The subjects completed the four items upstairs before proceeding to the Fieldhouse for the remaining three test items.

In order to establish validity and reliability three test periods were required. The Modified AAHPER Test was administered twice. The testing dates were December 3, 1968, and December 5, 1968. The tests were administered from 3:00 to 6:00 P.M. The subjects could take the test at any time during the testing period, but once they started they were instructed to continue in a predetermined order. Thirty-seven subjects completed a test and retest administration of this test and the results obtained were used to determine the reliability of the Modified AAHPER Test. No make-up tests were administered.

The criterion test was administered on December 7, 1968, from 9:00 to 12:00 A.M. The three test periods were scheduled in a one week period to minimize the training effect that time might have on the results. The results of the criterion test were compared with with the first administration of the Modified AAHPER Fitness Test to establish validity. Nineteen subjects participated in the criterion test.

The subjects were asked to report to Gymnasium I to begin all three tests. When they arrived, they received a score card. A sample of the score card is present in Appendix A. The subjects completed

the shuttle run and standing broad jump, then reported to Gymnasium II for the pull-up and sit-up items. After the first four items the subjects reported to the Fieldhouse where they participated in the 50-yard dash, shot put, and 600-yard run-walk respectively.

The order for the criterion items was the same except the inclined sit-up and the pull-up on the universal gym were done in the Weight Training Room. It might be noted here that the instructions for administering the designed items are similar to the corresponding items in the Modified AAHPER Test. This was done to insure that the results did not fluctuate because of administration procedures. Testing assistants were reminded to be careful in their testing procedures so that the results were kept consistent.

Statistical Procedures

The reliability coefficients were calculated from the test-retest data. After the results from the two Modified AAHPER Tests were obtained they were listed on a separate sheet of paper. Each subject was assigned an identification number and his corresponding test scores listed. The same procedure was used in stating the results of the 19 subjects that completed the criterion test. Since the computer was used for calculating the results, the completed listing of results (data) were then placed on coding sheets. Use of the computer required that the items measured in inches be converted to decimal form. The items that had to be converted were the standing broad jump, the vertical jump, the shot put, and the medicine ball put. The coding sheets were then submitted to the Computer Center at the University of North Dakota, where the information was punched on data cards.

The program used in computing the values desired for this study has been filed in the Computer Center Library. The title of the program is "Means, Standard Deviations, Variances, and Correlation Coefficients." The Product-Moment formula was used in this program for computing the correlation coefficients. An IBM 360/30 Computer was used to calculate the results. The program was written in Fortran IV (E-Level) language. Both the reliability and validity coefficients were determined in the above manner.

The null hypothesis stated the reliability and validity of the Modified AAHPER Test to be zero. The alternate hypothesis is that if the coefficients are above the designated level of significance the items are related. The null hypothesis was tested by comparing the correlation coefficients to a value obtained from a "t" table. Edwards [23] describes using the table for finding the values of the correlation coefficient for different levels of significance. The degrees of freedom are equal to the number of subjects minus two, and the level of significance is .01. A one-tailed test was used. The correlation coefficients for validity must be above 0.528 before the null hypothesis can be rejected. The correlation coefficient for reliability must be above 0.381 before the null hypothesis can be rejected. The mathematical procedure and the results obtained are located in Appendices D, E, and F.

---

[23]Allen L. Edwards, Experimental Design in Psychological Research (Chicago: Holt, Rinehart, and Winston, 1960), pp. 79 and 362.

CHAPTER III

ITEM ANALYSIS TO DETERMINE RELIABILITY AND VALIDITY

The results obtained from the test-retest of the seven Modified AAHPER Test Items are shown in Appendix B. The results obtained from the 19 subjects that took the designed test and the Modified AAHPER Test are shown in Appendix C. The results contained in these two appendices were used to calculate the correlation coefficients for the reliability and validity of the Modified AAHPER Fitness Test. The means, standard deviations, and variances were also calculated using these results. The left hand column contains the identification number of each subject. All conversions were made by rounding off the inch to the nearest tenth. When the inch was exactly between two tenths, the rounding was alternated between the two tenths.

The reliability and validity coefficients that were calculated by the computer are listed in Tables I and II. These tables also contain the means and standard deviations attained. These values were taken from Appendices E and F. The significance factor stated in these tables was determined for validating the correlation coefficients. The test of significance was used to estimate the probability of the calculated coefficients. The following material is a statement of the statistical results of the seven items in the Modified AAHPER Test. There were three test periods that gave four sets of results. The first two figures stated for the means and standard deviations are

for the first two administrations of the Modified AAHPER Test taken from Table I. The second two figures are the means and standard deviations of the 19 subjects that took the first Modified AAHPER Test and the criterion test. These figures were taken from Table II. The reliability and validity coefficients were taken from Tables I and II respectively.

Sit-up

The means obtained from the four sit-up tests were 76.8, 78.4, 71.8, and 35.7. The standard deviations were 30.0, 33.0, 26.4, and 24.9. The correlation coefficients for reliability and validity were both found to be 0.74. Since the observed values of reliability and validity are greater than the hypothetical value the null hypothesis is rejected at the .01 level.

Pull-up

The means of the four pull-up tests were 8.76, 8.68, 9.3, and 17.1. The standard deviations were found to be 3.7, 3.9, 2.2, and 3.7. The pull-up item was found to have a reliability of 0.91 and a validity of 0.41. The reliability coefficient shows that the test is highly reliable. The validity coefficient is below the acceptable value and the null hypothesis is retained.

Standing broad jump

The standing broad jump items obtained means of 7.5, 7.5, 7.7, and 1.8. The standard deviations were 0.7, 0.7, 0.7, and 0.2. The standing broad jumps reliability and validity coefficients were

calculated tc be 0.90 and 0.63. Both values are greater than the hypothetical value and the null hypothesis is rejected at the .01 level.

Shuttle run

The means obtained in the shuttle run were 9.1, 9.1, 9.0, and 16.5 for the three test administrations. The standard deviations were 0.5, 0.5, 0.4, and 0.8. The reliability coefficient was found to be 0.74, and the validity coefficient 0.72. Both of these values are significant at the .01 level.

50-yard dash

The means obtained in the 50-yard dash were 6.7, 6.6, 6.7, and 5.8. The standard deviations were 0.4, 0.3, 0.3, and 0.3. The coefficient of reliability for the 50-yard dash was found to be a low 0.59. The validity coefficient was 0.63. Since both of these values are greater than the hypothetical value, the null hypothesis is rejected at the .01 level.

Shot put

The shot put item had means of 31.4, 31.4, 31.8, and 37.6. The standard deviation obtained were 3.2, 3.0, 3.4, and 3.3. The test-retest coefficient calculated was 0.89. The validity coefficient was 0.77. Both of these values are significant at the .01 level.

600-yard run-walk

The means for this item were 113.0, 109.8, 114.2, and 18.6. Standard deviations obtained were 7.8, 6.1, 8.5, and 1.8. The 600-yard run-walk item was found to have a reliability coefficient of

0.55 and a validity coefficient of -0.55. The reason for a negative coefficient was because the 600-yard run-walk has as its objective the lowest possible time, while the twelve-minute run-walk measures the largest number of laps attainable in that time period. The coefficient was thus attained by comparing a minimum number with a maximum number. Both coefficients are significant at the .01 level, but the validity coefficient was narrowly larger than the hypothetical value.

These results indicate that the null hypothesis was rejected in only one of the seven items. This was the validity coefficient for the pull-up items. The standard deviations stated for the sit-up and 600-yard run-walk items indicate that a larger than normal dispersion about the mean is present.

TABLE I

THE MEANS, STANDARD DEVIATIONS, RELIABILITY COEFFICIENTS, AND SIGNIFICANCE FOR EACH OF THE SEVEN ITEMS OF THE MODIFIED AAHPER TEST

Item	Test No.	Mean	S.D.	Reliability Coefficient	Level of Significance
Sit-up	1	76.04	30.0	0.74	.01
	2	78.4	33.0		
Pull-up	1	8.8	3.7	0.91	.01
	2	8.7	3.9		
Standing Broad Jump	1	7.5	0.7	0.90	.01
	2	7.5	0.7		
Shuttle Run	1	9.1	0.5	0.74	.01
	2	9.1	0.5		
50-Yard Dash	1	6.7	0.4	0.59	.01
	2	6.6	0.3		
Shot Put	1	31.4	3.2	0.89	.01
	2	31.4	3.0		
600-Yard Run Walk	1	113.0	7.8	0.55	.01
	2	109.8	6.1		

Note: Significance was determined at the .01 level for 35 degrees of freedom.

The critical value of significance was 0.381 for N - 2 degrees of freedom.

TABLE II

THE MEANS, STANDARD DEVIATIONS, VALIDITY COEFFICIENTS, AND SIGNIFICANCE FOR EACH OF THE SEVEN ITEMS OF THE MODIFIED AAHPER TEST AND FOR EACH OF THE SEVEN CRITERION ITEMS

Item	Test No.	Mean	S.D.	Validity Coefficient	Significant
Sit-up	1	71.8	26.4	0.74	.01
Inclined Sit-up	3	35.7	24.9		
Pull-up	1	9.3	2.2	0.41	N.S.*
Designed Pull-up	3	17.1	3.7		
Standing Broad Jump	1	7.7	0.7	0.63	.01
Vertical Jump	3	1.8	0.2		
Shuttle Run	1	9.0	0.4	0.72	.01
Designed Shuttle Run	3	16.5	0.8		
50-yard Dash	1	6.7	0.3	0.79	.01
Designed 50-Yard Dash	3	5.8	0.3		
Shot Put	1	31.8	3.4	0.77	.01
Medicine Ball Throw	3	37.6	3.3		
600-Yard Run-Walk	1	114.2	8.5	-0.55	.01
12 Minute Run-Walk	3	18.6	1.8		

Note: Significance was determined at the .01 level for 17 degrees of freedom.

The critical value of significance was 0.528 for N - 2 degrees of freedom.

* N.S. means the value was not significant.

CHAPTER IV

DISCUSSION

Reliability

All reliability coefficients were found to be significant at the .01 level, however some were much higher than others. The reliability coefficients for the pull-up, standing broad jump, and shot put were 0.91, 0.90, and 0.89 respectively. These coefficients indicate acceptability by many standards. Clarke[24] suggests that a physical test should have a minimum reliability coefficient of approximately 0.90. The pull-up item was a measure of dynamic strength. The standing broad jump and shot put items were measures of explosive strength. These facts indicate that the items directly measuring explosive and dynamic strength tend to have high correlation coefficients on a test-retest basis.

The sit-up item and the shuttle run had the next highest coefficents of reliability. The reliabilities were 0.74 for both items. The sit-up item is an efficiency test of the abdominal and hip flexor muscles. The shuttle run measures speed and change of direction. The coefficients stated for these items are not high enough for the items to be acceptable measures of fitness. It should be noted that the sit-up has a high standard deviation. This indicates that the scores

---

[24]Clarke, op. cit., p. 35.

were widely dispersed about the mean. The scores of the test-retest of the sit-up varied greatly for some subjects. This was attributed to the subjects desire and the manner that the muscles were used in the item.

The lowest reliability coefficients were attained from the two running items. The 50-yard dash, which judges speed, had a reliability of 0.59. The reliability of the 600-yard run-walk, a test of cardiovascular efficiency, was 0.55. The low coefficient in the 50-yard dash can be attributed partly to the error received in stop watch timing and the inability of the subjects to start consistently. The 600-yard run-walk coefficient was low mainly because of the motivational factors involved. The standard deviation of the endurance item was second highest of the items. This indicates some unnatural spread about the mean. Although the testers instructed the subjects to do their best in the 600-yard run-walk, many subjects were reluctant to put forth their maximum effort. These two items are not recommended test items for fitness because of their low reliabilities, but if more stringent administrative procedures were followed the reliability coefficients could be improved.

Validity

The validity of the Modified AAHPER Test was determined by comparing the results of the first administration of the Modified AAHPER Test with the results of the criterion test. The criterion test was selected to test validity because it judges the same criteria as the Modified AAHPER Test Items. The validity testing was primarily concerned with knowing and being able to define the components of fitness

and determining how to measure them. In administering the criterion test the test procedures of the corresponding items were kept the same. This insured the writer that the test items were much the same and that more consistent results would be attained.

Four of the seven items were in the average validity range. They were the 50-yard dash, the shot put, the sit-up, and the shuttle run. Their coefficients of validity were 0.79, 0.77, 0.74, and 0.72 respectively. The validity of the 50-yard dash was higher than its reliability. This can be accounted for because of higher consistency in a running start.

Although the sit-up item has a significant coefficient, the muscles used in the exercise are primarily the psoas group. As Kendall[25] states, "weakness of the abdominal muscles can go undetected in the sit-up." If the bent-knee sit-up were used in the abdominal muscles would be the primary movers.

The shuttle run item in the criterion test requires more change of direction than the AAHPER Item. The criterion item also eliminated skill in picking up the blocks.

The standing broad jump and the 600-yard run-walk have low coefficients and their criteria is questionable. The standing broad jump had 0.63 as its validity coefficient. Both of the test items measure explosive power of leg extensors, but the vertical jump involves marking with a piece of chalk. The vertical jump also involves vertical velocity while the standing broad jump involves angular velocity. These two factors account for some of the inconsistency. The validity

---

[25]Kendall, op. cit., p. 189.

of the 600-yard run-walk was -0.55. Since the twelve-minute run-walk was found by Doolittle and Bigbee[26] to be a better measure of cardiovascular efficiency than the 600-yard run-walk, this factor appears justifiable. The 600-yard run-walk involves too much speed and too short a distance to be a good measure of endurance. Both of these values are low enough to state that the items are not valid.

The validity coefficient of the pull-up was 0.41. This is far below the acceptable value for significance. Kendall[27] states that "the pull-up falls into the category of too difficult." The criterion pull-up was done with less weight and did not involve leg movement. The AAHPER pull-up involves more body movement than the designed item and thus uses more accessory muscles. The means of the pull-up in the test-retest was 8.8 and 8.7 respectively, while the criterion pull-up had a mean of 17.1. These values show that the criterion item was easier and would be a better item to use in age groups where some subjects could not complete one repetition.

---

[26]Doolittle and Bigbee, op. cit., pp. 491-95.

[27]Kendall, op. cit., p. 189.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The purpose of this study was two-fold. The first phase of the problem was to determine the reliability of the Modified AAHPER Fitness Test. The second phase of the study was to determine the validity of this test. The Pearson-Product Moment Formula was used to establish the correlation coefficients. The "t" test was used to determine the significance of the correlation coefficients at the .01 level.

The reliability was established by selecting a random sample of thirty-seven male college freshmen from physical education classes at the University of North Dakota and administering the Modified AAHPER Fitness Test to them twice. The two test dates were kept close together so that the subjects' physical condition did not change.

Validity was established by selecting a fitness test that tested the same criteria as the Modified AAHPER Test Items. Nineteen subjects that participated in the first two Modified AAHPER Tests also participated in the criterion test. The results obtained in the criterion test were correlated with the results that these subjects obtained in the first administration of the Modified AAHPER Test.

The correlation coefficients obtained for reliability were 0.74 for the sit-up, 0.91 for the pull-up, 0.90 for the standing broad jump, 0.74 for the shuttle run, 0.59 for the 50-yard dash, 0.89 for the shot put, and 0.55 for the 600-yard run-walk. All of these values were significant at the .01 level. The correlation coefficients for validity were 0.74 for the sit-up, 0.41 for the pull-up, 0.63 for the standing broad jump, 0.72 for the shuttle run, 0.79 for the 50-yard dash, 0.72 for the shot put, and -0.55 for the 600-yard run-walk. The pull-up item was not significant at the .01 level. All other items were found significant.

Conclusions

Within the limitations and assumptions of this study the following conclusions appear to be justified:

1. The sit-up, shuttle run, 50-yard dash, and the 600-yard run-walk have correlation coefficient below 0.89.

2. The pull-up, standing broad jump, and shot put items have reliability coefficients equal to or greater than 0.89.

3. All the reliability coefficients were significant at the .01 level.

4. The correlation coefficient for the validity of the pull-up item was not significant.

5. All other validity coefficients were significant. However, the correlation coefficient for the 600-yard run-walk and the standing broad jump were lower than the generally accepted standard.

Recommendations

The following recommendations are suggested as a result of this study:

1. The 600-yard run-walk was found to be a weak measure of cardiovascular efficiency and should be replaced by a more valid and reliable measure. Based on recent research, the twelve-minute run-walk has been shown to be more valid and reliable.

2. The sit-up item is not a highly reliable item. The inconsistency of results indicate that this item as it is employed in the AAHPER Fitness Test, measures the strength of the abdominal muscles in an inaccurate manner, and weakness of these muscles can go undetected. The bent knee sit-up would correct this problem, if used in the fitness test.

3. The pull-up item was found to be highly reliable, but the coefficient for validity was not significant at the .01 level. More research is needed on the validity of this item.

4. The shuttle run would be a more acceptable item if three trials were used. This was confirmed by Montoye, Cunningham, Marmis and Kozar in the related literature.

5. The 50-yard dash is not a satisfactory item for testing speed. The standing start and error in stop watch timing lower the reliability of the item.

6. Future test to validate measures of this kind should measure more closely the same criteria.

APPENDICES

AAHPER TEST SCORES

Name _____
Last First Middle

High School _____
Name of School City State

Birth Date _____
Day Month Year

Class (circle) Fr. So. Jr. Sr. Wt._____

Class	Raw Score #1	Raw Score #2	Raw Score #3
1. Sit-ups			
2. Pull Ups			
3. Standing Broad Jump			
4. Shuttle Run			
5. 50-Yard Dash			
6. Shot Put			
7. 600-Yard Run-walk			

TABLE III

TEST-RETEST RESULTS OF THE MODIFIED AAHPER TEST

I.D. Number	Sit-up		Pull-up		Standing Broad Jump (tenth of a foot)		Shuttle Run (sec.)		50-Yard Dash (sec.)		Shot Put (tenth of a foot)		600-Yard Run-Walk (sec.)	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
1	47	72	12	10	7.4	7.4	9.2	9.2	6.6	6.7	27.0	30.0	119	114
2	100	110	10	11	9.3	9.4	8.2	8.0	5.9	5.7	36.4	34.5	108	110
3	65	49	10	8	7.8	7.8	9.2	9.7	7.2	7.0	29.7	27.8	126	118
4	50	50	7	6	6.8	6.3	9.4	9.8	7.0	6.7	30.2	28.1	114	121
5	76	60	9	10	8.4	8.7	9.1	8.9	7.0	6.5	33.0	33.8	128	106
6	80	85	8	7	7.0	6.9	8.6	8.7	6.7	6.6	31.8	29.7	107	104
7	75	84	9	7	8.3	7.9	8.5	8.9	6.2	6.3	27.0	28.2	108	117
8	36	60	7	8	7.9	8.3	8.9	8.7	6.6	7.1	36.7	34.0	105	104
9	60	55	7	4	8.2	7.9	9.1	9.4	6.8	6.4	31.0	30.8	108	109
10	80	72	9	10	6.7	7.2	8.9	8.6	6.7	6.2	32.0	31.2	128	113
11	85	100	8	8	8.0	7.9	8.1	8.7	6.4	6.4	33.6	33.7	111	105
12	50	56	6	7	6.8	6.9	8.8	8.7	6.6	6.2	31.0	30.4	116	114
13	59	56	9	9	7.9	7.9	9.1	9.0	6.3	6.8	38.8	38.8	109	108
14	131	101	15	16	8.2	7.7	9.2	9.0	6.6	6.5	33.4	33.3	106	110
15	40	40	9	10	6.7	6.8	9.0	9.0	7.1	6.7	32.4	33.7	110	112
16	60	117	9	9	7.3	7.4	9.6	9.8	6.7	7.0	25.6	25.5	126	109
17	89	55	8	8	7.9	7.8	9.5	9.2	7.3	7.4	30.5	29.2	107	108
18	131	131	13	16	7.4	7.7	9.3	9.5	6.5	6.1	34.0	34.6	106	107
19	50	55	11	11	8.6	8.7	8.8	8.7	6.4	6.1	31.0	32.3	128	112
20	70	70	4	6	7.2	6.8	9.4	9.3	6.9	6.8	28.2	29.4	106	102

TABLE III--Continued

I.D. Number	Sit-up		Pull-up		Standing Broad Jump (tenth of a foot)		Shuttle Run (sec.)		50-Yard Dash (sec.)		Shot Put (tenth of a foot)		600-Yard Run-Walk (sec.)	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
21	56	48	4	5	7.7	7.9	9.3	9.5	6.5	6.9	30.5	29.2	116	120
22	70	131	10	11	7.2	7.2	9.0	8.7	7.1	6.5	34.2	34.7	124	112
23	64	65	3	2	7.0	6.8	9.4	9.1	7.4	6.6	28.5	29.7	113	111
24	53	50	14	16	7.3	7.8	9.7	8.7	6.7	6.3	34.4	37.4	105	104
25	131	131	10	10	8.2	8.2	8.9	8.6	6.2	6.6	31.0	31.5	107	106
26	51	44	6	6	6.8	6.5	9.5	9.1	6.7	6.8	28.8	29.3	102	104
27	101	110	7	7	6.5	6.1	10.7	10.1	7.6	7.0	30.7	30.2	121	124
28	140	131	17	19	6.8	6.7	9.3	9.1	6.4	6.4	28.2	29.7	103	103
29	131	131	9	8	8.0	7.8	8.4	8.5	6.5	6.6	31.5	32.8	101	97
30	40	35	3	4	7.3	7.7	8.7	8.7	6.7	6.3	40.0	39.5	109	111
31	50	30	5	5	7.2	6.9	9.2	9.1	6.5	6.4	26.4	28.6	118	120
32	100	142	4	2	6.7	7.0	9.4	9.5	6.6	6.5	32.0	31.6	114	112
33	103	128	7	7	8.6	8.8	9.2	8.9	7.1	7.1	28.8	29.0	115	107
34	50	50	11	9	7.6	7.0	8.6	8.7	6.2	6.3	31.5	32.1	108	106
35	65	70	3	5	6.9	6.8	9.4	9.3	7.0	6.8	30.7	32.0	114	107
36	131	60	12	7	7.9	7.0	8.8	9.4	6.7	6.9	31.4	28.4	115	116
37	70	65	19	17	6.7	6.8	9.5	10.0	6.9	6.6	28.7	26.9	120	111

T1 The first administration of the Modified AAHPER Test

T2 The second administration of the Modified AAHPER Test

TABLE IV

RESULTS OF THE FIRST MODIFIED AAHPER TEST AND THE DESIGNED TEST

I.D. Number	Sit-up		Pull-up		Standing Broad Jump (tenth of a foot)		Shuttle Run (sec.)		50-Yard Dash (sec.)		Shot Put (tenth of a foot)		600-Yard Run-Walk (sec.)	
	T1	T3	T1	T3	T1	T3	T1	T3	T1	T3	T1	T3	T1	T3
1	47	31	12	23	7.4	1.6	9.2	16.9	6.6	6.1	27.0	34.7	119	18.00
2	100	50	10	19	9.3	2.4	8.2	15.4	5.9	5.2	36.4	41.7	108	20.00
3	65	32	10	13	7.8	1.9	9.2	16.0	7.2	5.9	29.7	32.2	126	19.00
4	50	30	7	13	6.8	1.6	9.4	16.3	7.0	6.1	30.2	33.5	114	18.50
5	76	48	9	14	8.4	1.7	9.1	16.5	7.0	5.9	33.0	40.2	128	18.50
6	80	30	8	18	7.0	1.8	8.6	16.1	6.7	5.6	31.8	37.7	107	19.50
7	75	30	9	17	8.3	1.8	8.5	15.7	6.2	5.7	27.0	38.0	108	19.50
8	36	18	7	24	7.9	2.2	8.9	16.1	6.6	6.0	36.7	40.6	105	22.75
9	60	18	7	10	8.2	1.7	9.1	16.5	6.8	6.0	31.0	37.3	108	18.50
10	80	35	9	17	6.7	1.7	8.9	15.6	6.7	5.7	32.0	37.8	128	17.00
11	85	30	8	18	8.0	1.8	8.1	15.1	6.4	5.2	33.6	41.2	111	19.50
12	50	20	6	16	6.8	1.5	8.8	16.1	6.6	5.7	31.0	38.5	116	16.75
13	59	28	9	15	7.9	1.5	9.1	16.7	6.3	5.2	38.8	44.1	109	18.50
14	131	57	15	24	8.2	1.9	9.2	16.9	6.6	5.7	33.4	41.5	106	17.25
15	40	20	9	19	6.7	1.6	9.0	17.2	7.1	6.1	32.4	33.7	110	17.25
16	60	30	9	17	7.3	1.3	9.6	17.6	6.7	5.8	25.6	32.7	126	18.50
17	89	15	8	17	7.9	1.7	9.5	17.1	7.3	6.5	30.5	36.2	107	21.75
18	131	131	13	18	7.4	1.7	9.3	18.4	6.5	5.8	34.0	38.7	106	19.25
19	50	25	11	13	8.6	1.9	8.8	17.3	6.4	5.4	31.0	34.5	128	14.00

T1 The first administration of the Modified AAHPER Test

T3 The administration of the Criterion Test

APPENDIX D

STATISTICAL PROCEDURE USED TO DETERMINE THE RELIABILITY AND VALIDITY CORRELATION COEFFICIENTS

PROCEDURE:

Step 1 - Collect data
Step 2 - List data on coding sheets
Step 3 - Punch data on data cards
Step 4 - Attain program for finding correlation coefficients from computer library
Step 5 - Change format statement and insert data into program deck
Step 6 - Run program through computer to attain results

DATA:

1. Data in inches was converted to decimal system.
2. When results were attained they were rounded off to the nearest hundredth, but are listed in Appendices C and D as they were computed.
3. The mean, standard deviation, and variance are stated in Appendices C and D, but were not used for evaluating the results of the Modified AAHPER Test.

PEARSON PRODUCT-MOMENT FORMULA:

$$r = \frac{\Sigma xy - \frac{(\Sigma x)\ (\Sigma y)}{N}}{\sqrt{(\Sigma x^2 - \frac{(\Sigma x)^2}{N})\ (\Sigma y^2 - \frac{(\Sigma y)^2}{N})}}$$

APPENDIX E

COMPUTER RESULTS FOR RELIABILITY

PROBLEM 1

N = 37

Variable	Mean	Standard Deviation	Variance
1. Sit-up	76.8	30.0	898.7
2. Pull-up	8.8	3.7	13.8
3. Standing Broad Jump	7.5	0.7	0.5
4. Shuttle Run	9.1	0.5	0.2
5. 50-Yard Dash	6.7	0.4	0.1
6. Shot Put	31.4	3.2	10.1
7. 600-Yard Run-Walk	113.0	7.8	60.3
8. Sit-up	78.4	33.0	1086.9
9. Pull-up	8.7	3.9	15.6
10. Standing Broad Jump	7.5	0.7	0.5
11. Shuttle Run	9.1	0.5	0.2
12. 50-Yard Dash	6.6	0.3	0.1
13. Shot Put	31.4	3.0	9.2
14. 600-Yard Run-Walk	109.8	6.1	36.8

LISTED RESULTS

COMPARED VARIABLES	CORRELATION COEFFICIENT
1 and 8	0.74
2 and 9	0.91
3 and 10	0.90
4 and 11	0.74
5 and 12	0.59
6 and 13	0.89
7 and 14	0.55

APPENDIX F

COMPUTER RESULTS FOR VALIDITY

PROBLEM 1

N = 19

Variable	Mean	Standard Deviation	Variance
1. Sit-up	71.8	26.4	697.9
2. Pull-up	9.3	2.2	4.7
3. Standing Broad Jump	7.7	0.7	0.5
4. Shuttle Run	9.0	0.4	0.2
5. 50-Yard Dash	6.7	0.3	0.1
6. Shot Put	31.8	3.3	10.6
7. 600-Yard Run-Walk	114.2	8.5	72.1
8. Inclined Sit-up	35.7	24.9	621.7
9. Designed Pull-up	17.1	3.7	13.5
10. Vertical Jump	1.8	0.2	0.1
11. Designed Shuttle Run	16.5	0.8	0.6
12. 50-Yard Run	5.8	0.3	0.1
13. Medicine Ball Put	37.6	3.3	11.1
14. 12-Minute Run-Walk	18.6	1.8	3.3

LISTED RESULTS

COMPARED VARIABLES	CORRELATION COEFFICIENT
1 and 8	0.74
2 and 9	0.41
3 and 10	0.63
4 and 11	0.72
5 and 12	0.79
6 and 13	0.77
7 and 14	0.55

BIBLIOGRAPHY

Books

American Association for Health, Physical Education, and Recreation. AAHPER Youth Fitness Test Manual. Washington, D.C.: AAHPER, Revised Edition, 1965.

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

Edwards, Allen L. Experimental Design in Psychological Research. New York: Holt, Rinehart, and Winston, 1960.

Fleishman, Edwin A. Examiner's Manual for the Basic Fitness Tests. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

Fleishman, Edwin A. The Structure and Measurement of Physical Fitness. Englewood Cliffs, N.J. Prentice-Hall, Inc., 1964.

Jordan, Payton and Spencer, Bud. Champions in the Making. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968.

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

McCloy, Charles H. and Young, Norma D. Tests and Measurements in Health and Physical Education. New York: Appleton-Century-Crofts, Inc., 1954.

Rasch, Philip J. and Burke, Roger K. Kinesiology and Applied Anatomy. Philadelphia: Lea and Febiger, 1967.

Stanley, Julian C. Measurement in Today's Schools. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964.

Articles and Periodicals

Doolittle, T. L., and Bigbee, Rollin. "The Twelve-Minute Run-Walk: A test for cardiorespiratory Fitness of Adolescent Boys," Research Quarterly, XXXIX (October, 1968), 491-95.

Esslinger, Arthur A. "Perspective on Testing," Journal of Health, Physical Education, and Recreation, XXI (September, 1960), 36-37.

Hunsicker, Paul, and Grey, George. "Studies in Human Strength," Research Quarterly, XXVIII (May, 1957), 109-22.

Kendall, Florence P. "A Criticism of Current Tests and Exercises for Physical Fitness," Journal of the American Physical Therapy Association, XXXXV (March, 1965), 188-89.

Larson, Leonard A. "A Factor and Validity Analysis of Strength Variables and Tests with a Test Combination of Chinning, Dipping, and Vertical Jump," Research Quarterly, XI (December, 1940), 82-96.

Unpublished Materials

Field, Arthur E. J. "A Reliability Analysis of the American Association for Health, Physical Education and Recreation Youth Fitness Test Items." Unpublished Master's thesis, University of British Columbia, 1965.

International Committee on Standardization of Physical Fitness Tests. Tentative Physical Fitness Measurement Standards, First Report on Physical Fitness Measurement Standards Prepared at the Sandefjord, Norway 1966 Meeting following Preliminary Meetings in 1964 and 1965 in Tokyo, Japan. The Editorial Committee consisted of Leonard A. Larson, William G. Reddan, John C. Pooley, and Efraim Ben-Ari.

Montoye, Henry J., et al. "Reliability of Multitrial Items of the AAHPER Youth Fitness Test." Abstracts of Research Papers 1968 (Published by AAHPER, 1968), p. 71.

Pollock, Michael L., Rothermel, Bradley L., and Cureton, Thomas K. "AAHPER Fitness Test Score Changes as a Result of an Eight-Week Sport and Physical Fitness Program at the University of Illinois." Abstracts of Research Papers 1967 (Published by AAHPER, 1967), p. 82.