8-1-1969

An Investigation of the Reliability and Validity of the Grand Forks Fitness Test as Administered to Forty-Two Central High School Boys

Julian M. Gulsvig

AN INVESTIGATION OF THE RELIABILITY AND VALIDITY OF THE GRAND FORKS FITNESS TEST AS ADMINISTERED TO FORTY-TWO CENTRAL HIGH SCHOOL BOYS

by

Julian M. Gulsvig

B.S. in Physical Education, Concordia College 1963

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

August
1969

This thesis submitted by Julian M. Gulsvig 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. Bolonchuk
Chairman

W. C. Koenig

Marjorie P. Behringer

A. William Johnson
Dean of the Graduate School

Permission

Title AN INVESTIGATION OF THE RELIABILITY AND VALIDITY OF THE GRAND FORKS FITNESS TEST AS ADMINISTERED TO FORTY-TWO CENTRAL HIGH SCHOOL BOYS

Department Physical Education

Degree Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairman of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature [illegible]

Date [illegible]

ACKNOWLEDGMENTS

The writer wishes to extend his sincere gratitude and appreciation to Mr. William Bolonchuk and Mr. Walter Koenig for the many hours spent in the helpful guidance and assistance in preparation and completion of this study; to Mr. Richard Johnson and other staff members in the Computer Center at the University of North Dakota for their work on the statistical analysis; to Mr. Robert Leach for his work on the illustrations; to Mr. Arnold Bakke for his full cooperation in completing the study; and to the staff members and students of Central High School who were involved in the study.

A special extension of gratitude should be given to my wife and son for their patience and helpful encouragement during the completion of the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABSTRACT

The lack of fitness exhibited by American men tested for induction into the military service prompted the development of fitness tests to assay the physical fitness of our nation's youth. The Grand Forks Test was designed to determine the level of fitness of the youth in the Grand Forks Public Schools. This study was formulated to determine the reliability and validity of the Grand Forks Fitness Test.

A stratified random sample of forty-two boys from Central High School completed the testing phase of this study. Each subject was tested on three consecutive mornings. The Grand Forks Test was administered the first two mornings and a test-retest analysis by the Pearson Product Moment correlation technique was used to determine whether the difference between the means was significant. In order to validate the Grand Forks Test, a criterion test was administered the third day and the data compared to the scores recorded from the first test session. The criterion test was designed on the basis of a kinesiological analysis of the test items and as a result of Fleishman's criterion measure for each test item.

The recorded data were programmed into an IBM computer and the correlation coefficients were calculated. The significance of these coefficients was tested by comparing them to correlation values published in Garrett. The correlation coefficients indicated that all items in the Grand Forks Test were reliable. With the exception of the shuttle run, all of the test items were found to be valid.

CHAPTER I

THE SCOPE OF THE PROBLEM

When American youth were frequently turned down for military service during World War I, World War II and the Korean War because they were not physically fit, this country was made keenly aware of the necessity for a nationwide fitness program which would reach the majority of our young people. Studies completed from World War II to the early 1950's revealed disturbing deficiencies in the fitness of American youth. As a result of these studies, in 1956, our late President Dwight D. Eisenhower created the President's Council on Youth Fitness. The major concern of this council was to promote the totality of fitness to all youth serving agencies and therefore a Youth Fitness Test was developed to achieve this objective.[1]

Shortly after the formation of the council, the American Association of Health, Physical Education and Recreation, hereafter referred to as AAHPER, established its Youth Fitness Project. Through this project, the AAHPER Youth Fitness Test was developed to evaluate the fitness of American youth in our public schools, colleges and recreation programs. The results of the test were very widely used in planning school

---

[1]American Association of Health, Physical Education and Recreation, "The President's Council on Youth Fitness," JOHPER, XXXI (September, 1960), 35.

and community programs to improve physical fitness and to motivate the students to strive for a higher level of personal fitness.[1]

It is generally agreed that the physical vigor of our young people is one of America's important resources. Our late President John F. Kennedy stressed the importance of physical fitness by stating:

> If we waste and neglect this resource, if we allow it to dwindle and grow soft then we will destroy much of our ability to meet the great and vital challenges which confront our people. We will be unable to realize our full potential as a nation.[2]

Some schools, however, did not have the necessary facilities to adequately administer the AAHPER test as proposed, so they had to modify the original test battery to suit their situation. This was the case that presented itself in Grand Forks, North Dakota. Realizing the importance of improving the fitness of their youth, a selected committee of educators and physical educators in the school system devised a fitness test, using several AAHPER items and some substitute items, to assay the fitness of the youth enrolled in the city schools.

There has been a great deal of discussion among the physical educators of the city about the advisability of using the Grand Forks Test to measure this important resource. Some physical educators are convinced that it is not reliable or valid and some disagree with the administrative procedures. When this study was formulated, there were estimates given as to the reliability and validity of the test items in

---

[1]American Association of Health, Physical Education and Recreation, AAHPER Youth Fitness Test Manual (Washington, D.C.: AAHPER, revised edition, 1961), p. 1.

[2]John F. Kennedy, "The Soft American," Sports Illustrated, December 26, 1960, p. 16.

literature, but specific work on the Grand Forks Test battery had not been done. The results since then have warranted the determination of estimates of reliability and validity to find out if the test is accurate and if it is a suitable test of physical fitness. It is intended that this study will provide the local physical educators with preliminary evidence to show that the items of the test are or are not reliable and valid when administered according to the procedures outlined in the Grand Forks Test manual.[1] The students and physical educators will both profit by assaying the results of this study.

Review of Related Literature

Once the AAHPER fitness test was established with norms, studies were done abroad using those norms and the test to compare school children. A study done by Knuttgen in Denmark showed that Danish girls exceeded American girls in all seven categories and the Danish boys bettered the American boys in all categories but the softball throw.[2]

Other researchers also followed the same basic pattern of comparing American and European youth both before and after the development of the AAHPER fitness test. One highly significant study which actually led to the development of the AAHPER test was completed by Kraus and his associates. They developed tests of minimal muscular fitness to indicate the level of strength and flexibility for the major muscle groups in the body. Included in the test battery are

---

[1]Royal Goheen, et al., "Grand Forks City School Youth Fitness Test," Grand Forks, North Dakota, 1964, pp. 1-5. (Mimeographed.)

[2]Howard G. Knuttgen, "Comparison of Fitness of Danish and American School Children," Research Quarterly, XXXII (May, 1961), 191, 193.

six tests which measure the length of the back and hamstring muscles, the strength of the lower back, the strength of the upper back muscles, the strength of the abdominal muscles without the help of the psoas, the strength of the abdominal and psoas muscles working together and the strength of the psoas and the lower abdominal muscles. These tests were chosen because they represent the primary muscles used in the performance of the usual daily tasks. The tests were scored on a pass or fail basis and were proposed to indicate the minimum level of fitness that must be maintained for the individual to be able to function properly in the completion of his normal everyday routine. These tests became known as the Kraus-Weber tests.[1]

With these tests Kraus and Hirschland tested more than four thousand United States school children from both rural and urban communities and compared their achievement with more than three thousand European children tested. The results of this study showed that only 8.7 per cent of the European children failed one or more of the tests while 57.9 per cent of the American children failed. This, of course, only strengthened the American determination to promote physical fitness programs for our youth because our egos had been deflated and our national pride seriously damaged.[2]

The AAHPER Fitness Test satisfied an urgent need for some means of measuring the level of fitness in our youth. However, due to the

---

[1]H. Harrison Clarke, Application of Measurement to Health and Physical Education (2nd ed.; Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1951), pp. 214-17.

[2]Hans Kraus and Ruth P. Hirschland, "Minimum Muscular Fitness Tests in School Children," Research Quarterly, XXV (May, 1954), 178.

completion of additional research and the introduction of new philosophies, the suitability of the AAHPER Fitness Test has been questioned.

In a thorough evaluation of the AAHPER Test, Fleishman found that several of the tests, individually, came out very well, according to his criteria for statistical reliability. However, a fitness test battery cannot be evaluated in terms of the individual tests alone, but rather the key factors to consider would be the comprehensiveness and efficiency of the test as a whole. In view of these factors, 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 stamina in the 600-yeard run-walk. The trunk strength factor is measured imperfectly using the extended knee sit-up. Furthermore, the explosive strength factor is overemphasized in the AAHPER battery, since four of the seven tests fall into this category. Measures of static strength, extent flexibility, gross body equilibrium, gross body coordination and dynamic flexibility are not covered in the AAHPER battery. Therefore, this test battery, according to Fleishman, would not be an accurate or suitable measure of physical fitness.[1]

What exactly is physical fitness? Physical fitness refers to the "ideal mixture of bodily health plus the physical condition to perform everyday tasks effectively and to meet emergencies as they arise."[2]

---

[1]Edwin A. Fleishman, The Structure and Measurement of Physical Fitness (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964), pp. 155-56.

[2]Fred V. Hein, "What is Physical Fitness?", National Education Association Journal, LI (February, 1962), 34.

Clifford E. Keeney raises a question that is often overlooked in discussions about physical fitness. That is, "What are you trying to be fit for?" To answer his question he proposes there are three levels of fitness. The first level is the subnormal level which includes the ill and handicapped individuals. The normal level found in educational practices and the armed services constitutes the second level and the third level is labeled the supernormal level where the highly trained athlete has been placed. In any fitness testing, the administrator must be absolutely sure in what level he is working because tests appropriate for one level would certainly be inadequate for the others.[1]

A high level of fitness is a need all people have in common regardless of individual occupations. To satisfy this need, Kaufman challenges physical educators who call their program a physical education program, to include physical fitness as a primary objective. Physical fitness should be the unique contribution of physical education to the general education of America's youth. Physical educators and their programs must accept the responsibility for the physical growth and development of young people as well as instill in them the importance of maintaining a high level of fitness as they enter the adult world and their chosen occupation.[2]

Inseparable from measuring physical fitness is the motivation or willingness possessed by individuals to extend themselves in

---

[1]Clifford E. Keeney, "Work Capacity," JOHPER, XXXI (September, 1960), 30.

[2]Raymond Kaufman, "Our Unique Contribution," JOHPER, XXXI (September, 1960), 27.

performing the test items.[1] Wireman, in studying this aspect, points to a knowledge of the results as facilitating an increase in physical fitness performance.[2] Johnson and Nelson in their study concluded that strength measures were greatly influenced by the level of motivation during testing.[3] Level of aspiration and team competition are the greatest motivating factors according to the results of a study by Strong. It is interesting to note that motivation improves the performance of boys more than girls. Strong concludes his study by stating: "The validity of the measures of physical fitness tests is dependent upon motivating conditions under which the tests are administered."[4]

Validity and reliability are defined by H. Harrison Clarke in his text concerning measurement in physical education. 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, which is the reason for the inclusion of the kinesiological analysis; and (2) the amount of relationship shown between the test and the criterion, which was shown by correlational measures in this study. When showing the relationship by correlational measures, 0.90 or above is desirable, however, 0.80 is significant.[5]

---

[1]Keeney, "Work Capacity," p. 30.

[2]Billy O. Wireman, "Comparison of Four Approaches to Increasing Physical Fitness," Research Quarterly, XXXI (December, 1960), 664.

[3]Barry L. Johnson and Jack K. Nelson, "Effect of Different Motivational Techniques During Training and Testing Upon Strength Performances," Research Quarterly, XXXVIII (December, 1967), 636.

[4]Clinton H. Strong, "Motivation Related to Performance of Physical Fitness Tests," Research Quarterly, XXXIV (December, 1963), 506.

[5]Clarke, Application of Measurement, pp. 30, 33.

In regard to reliability, Clarke states that there are three major factors that influence the reliability of a measure. They are: (1) the randomness of the sample, (2) the size of the sample, and (3) the variability of the distribution. If the relationship on a test-retest basis is to be shown by correlational measures, 0.70 to 0.79 is considered adequate for group measurement with higher figures more desirable.[1]

The test items in the Grand Forks City School Youth Fitness Test are the vertical jump, the pull-up, the squat thrust, the shuttle run, the standing broad jump and the sit-up.[2] The shuttle run measures explosive strength in general because of starting and stopping so often as well as measuring speed and agility to a degree. The vertical jump and standing broad jump are both measures of explosive strength of the legs. The explosive strength factor refers to the capacity to exert a force against any resistance resulting in a bursting forth of the body or the resistance. The squat thrust and pull-up, on the other hand, are designated as measures of dynamic strength which refers to the capacity to exert a force against any resistance which results in movement. The squat thrust measures dynamic leg strength while the pull-up is used as a measure of dynamic arm strength. The remaining test item, the sit-up, is shown to be a weak measure of abdominal strength.[3]

---

[1]Clarke, Application of Measurement, pp. 36, 446.

[2]Goheen, Grand Forks Fitness Test, pp. 1-5.

[3]Fleishman, The Structure and Measurement of Physical Fitness, pp. 155-56.

An analysis of the standing broad jump and vertical jump demonstrates the necessity of knowing what muscles are involved in an activity and the degree of difference in their action. These jumps measure the same criterion so they can be validated by comparing the obtained scores for one administration of the test.

Glencross reports a high statistical reliability for both the standing broad jump (0.96) and the vertical jump (0.93) as well as a correlation coefficient of (0.71) between the two jumps.[1] However, Clarke and Degutis point out in their research that even though these tests measure the same criterion high correlations are not to be expected because both tests require different skills to complete.[2] Most researchers fail to consider jumping ability as highly important and each test requires a different technique since the vertical jump emphasizes only the vertical direction while the standing broad jump stresses distance combining both vertical and horizontal directions.[3]

Eckert, in studying the angular velocities and joint actions in the two jumps, points out a significant difference between the two. She reports a much greater angular velocity and range of movement in the hip action during the standing broad jump than is found in the vertical jump where greater angular velocities are recorded in joint actions of the knee and ankle. These differences in joint actions

---

[1]D. J. Glencross, "The Nature of the Vertical Jump Test and the STanding Broad Jump," Research Quarterly, XXXVII (October, 1966), 358.

[2]H. Harrison Clarke and Ernest W. Degutis, "Relationships Between Standing Broad Jump and Various Maturational, Anthropometric and Strength Tests of 12-Year-Old Boys," Research Quarterly, XXV (March, 1964), 258.

[3]Glencross, "The Nature of the Vertical Jump Test and the Standing Broad Jump," p. 359.

clearly indicate differences in time-force coordination of the joint action in the two types of jumps.[1]

Differences do exist between the findings of researchers in their studies, as illustrated in the preceding paragraphs. Therefore, Clarke points out that it is essential for the researchers to analyze the action of the muscles and joints in each test item in order to verify the use of the criterion test battery. By doing this, the tester can at least be sure that the criterion measure utilizes the same muscles and can point to other factors as probable causes for the computed correlational value.[2]

The shuttle run was included in the Grand Forks Test to measure speed and agility.[3] The Illinois Agility Run devised by Cureton was chosen as the validating criterion for this measure because the factors of speed, agility and changing direction are included in the completion of this test. There has been a correlation of 0.80 established between the shuttle run and the 50-yard dash showing the shuttle run does measure speed.[4]

The full squat, sometimes called the squat reach test, provided an adequate measure of dynamic leg strength as determined by

---

[1]Helen M. Eckert, "Angular Velocity and Range of Motion in the Vertical and Standing Broad Jumps," Research Quarterly, XXIX (December, 1968), 939.

[2]Clarke, Application of Measurement, p. 30.

[3]Serge Gambucci, Head of Physical Education Department at Central High School, Grand Forks, North Dakota, interview, June 21, 1968.

[4]Fleishman, The Structure and Measurement of Physical Fitness, p. 129.

the Saskatoon Public School Board.[1] Hence this test was chosen in order to validate the squat thrust test. Some researchers consider the squat thrust a test of agility. However, McCloy and Young point out that the validity of the squat thrust as a measure of agility is much lower than that of tests of agility involving running.[2]

Sparks, in his research on the AAHPER Fitness Test, concluded that the pull-up test is a good indicator of arm and shoulder girdle strength.[3] McCraw suggests it makes very little difference whether the palm faces forward or backward while performing the test. The subjects showed no significant difference when tested both ways.[4] Nelson concluded in his research that the pull-up is a questionable fitness test item for measuring arm strength because there are so many subjects who cannot complete even one pull-up.[5] In accordance with Nelson's research, Soderberg has developed a modified pull-up using the Universal Gym which he validated in a thesis study.

---

[1]Drs. D. A. Bailey and W. A. R. Orban, Study Directors, The Saskatoon Physical Fitness Test Manual (Saskatoon: Saskatoon Public School Board, 1962), p. 5.

[2]Charles H. McCloy and Norman D. Young, Tests and Measurements in Health and Physical Education (New York: Appleton-Century-Crofts, Inc., 1954), p. 78.

[3]Raymond E. Sparks, "Modification of the AAHPER Youth Physical Fitness Test" (unpublished Master's thesis, Springfield College, 1965), p. 67.

[4]Lynn W. McCraw, "Effects of Variation of Forearm Position in Elbow Flexion," Research Quarterly, XXXIV (December, 1964), 509.

[5]Dale O. Nelson, "Focus on Two Fitness Exercises," JOHPER, XXXV (May, 1964), 23.

Therefore, on the basis of Soderberg's validation, the modified pull-up was used in the criterion test battery for this study.[1]

The sit-up test should be performed with the knees flexed. This type of exercise is referred to as the bent knee sit-up. This position makes it possible for the abdominal muscles to do more work in raising the trunk and the hip flexors less. However, the action of the hip flexors can never be eliminated completely. In the knees extended position, weak abdominal muscles can go undetected. Research by Kendall indicates that by doing the bent knee sit-up without having the feet held down also tends to increase the action of the abdominal muscles while minimizing any low back strain that commonly occurs when the knees are extended. The bent knee sit-up was included in the criterion test battery because it was determined in Kendall's research to be a much better test of abdominal strength than the extended knee sit-up.[2]

The qualities that make a good fitness test are very hard to define. Kendall's article suggests certain criteria to be satisfied in a good fitness test. There should be no zero scores, there 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

---

[1]Paul Soderberg, "The Reliability and Validity of the Modified AAHPER Fitness Test as Applied to Freshman College Males at the University of North Dakota" (unpublished Master's thesis, University of North Dakota, 1969), p. 50.

[2]Florence P. Kendall, "A Criticism of Current Tests and Exercises for Physical Fitness," Journal of the American Physical Therapy Association, XXXV (March, 1956), pp. 189, 193-5.

meaningful, be of suitable difficulty and the factor of safety should be given thoughtful consideration.[1]

Summary

The lack of fitness exhibited by our young men tested for entry into the military prompted the development of tests to assay the fitness of American youth. The Grand Forks Fitness Test, following closely the design of the AAHPER Fitness Test, was devised to test the level of fitness of the youth in the Grand Forks Public Schools.

In order to appraise the accuracy and suitability of the Grand Forks Test, a determination of the reliability and validity of the test items is necessary. A criterion test battery was designed to determine how well the Grand Forks Test measures the qualities it was intended to measure. The items composing the criterion test battery were determined by a review of related literature.

The research completed in this study was valuable to both the students and the administrators involved with the Grand Forks Fitness Test. The results are intended for use as preliminary evidence to indicate if this test was reliable and valid when administered as outlined in the test manual.

---

[1]Kendall, "A Criticism of Current Tests and Exercises," pp. 188-9.

CHAPTER II

METHODOLOGY

The authors of the Grand Forks Fitness Test have described it as a motor performance and/or motor fitness test. They proposed that their test was specifically a measure of power, flexibility, speed, agility and balance. To the authors of the test, the term physical fitness meant "a capacity for sustained physical activity." The test was devised to measure the physical fitness levels of the youngsters in the Grand Forks School System.[1]

It was intended, in the early stages of the program, to keep a cumulative record of the performance of each child from the time he first took the test until its final application so that physical educators would be able to show the student his progress and also be able to send a progress record home to his parents. This concept, however, became too time consuming and the practice was dropped. Another factor which influenced the decision to discontinue the policy was in the method of scoring. The parents failed to understand how the tests were scored and therefore the progress record had little value for them.

The inconsistency in testing scoring and the lack of agreement among local physical educators concerning the use of the test warranted

---

[1]Goheen, Grand Forks Fitness Test, p. 1.

an appraisal of its value. The physical educators who continually use the test must be convinced that the Grand Forks Test actually measures what it purports to measure or the results have no practical value to the physical educators or their students. The appraisal was accomplished by computing reliability and validity coefficients after testing and an analysis of the data had been completed.

In order to statistically compute reliability and validity coefficients, subjects were chosen from the population for which the test was intended. The subjects were all male students at Central High School at the time of their testing. It had been previously decided to select a stratified random sample of 60 students from a finite population of 450 male students. However, only 42 of the original 60 completed the testing. Lack of motivation, sore muscles and oversleeping were responsible for the failure of the 18 to miss one or more of the test sessions and therefore their test scores were not included.

In choosing the 60, the names of all the male students were placed in boxes by grade level. Then names were drawn from the boxes and recorded in order until a total of 50 students from each class had been chosen. Fifty from each class were chosen because there would undoubtedly be some students who could not participate for a variety of reasons. Personal contact was then made with each name on the list, starting with number one, until 20 sophomores, 20 juniors and 20 seniors had committed themselves to the testing process. The reason for selecting 20 from each stratum is outlined in Table 1. Some students were eliminated immediately because of conflicting schedules, physical limitations which limited participation and because of which they were excused by approval of the doctor, and a lack of desire to participate. The desired number of subjects was acquired after contact had been made with 29 sophomores, 40 juniors and 36 seniors. Each of the subjects, as he was contacted, was made fully aware of his role in the testing and of the value of the study to himself as well as to the physical educators and his fellow students.

TABLE 1

FINITE POPULATION AND SAMPLE SUBJECTS BY CLASS FOR GRAND FORKS CENTRAL MALES PARTICIPATING IN THE STUDY

Sophomores			Juniors			Seniors		
N	%	n	N	%	n	N	%	n
152	33.77	20	151	33.55	20	147	33.11	20

The single group design was chosen to complete the test analysis because it allows for a minimum of time between tests and thus the results would not be biased in a positive direction. In using this design, each subject was his own control and therefore, no outside controls were necessary. With this design, paired data from two different test sessions were used to statistically compute correlation coefficients and significance levels for determining reliability and validity.

One complete fitness test was administered to all subjects on each of three consecutive mornings. The subjects reported to Central High School and were dressed and ready for testing in the gymnasium at 7:15 a.m. each of the three mornings. The first two mornings were used for a test-retest of the Grand Forks Test. The third morning was used for administering the criterion test. The scores obtained on the criterion test were compared to those recorded the first morning for the purpose of determining validity.

The tests were administered by the investigator and five test assistants. All of the assistants had training in physical education and three of the five were physical educators in the city who are involved in using the Grand Forks Test each year. In a general meeting with all the assistants, the writer gave each an instruction sheet explaining the administrative procedure for the specific test they were to administer. All of the testers were present to administer their respective test each day.

In order not to inject an uncontrollable variable, no attempt was made to motivate the subjects during the test periods. Each of the testers could have a different method of motivation and the subjects were allowed to follow a normal daily routine without any schedule established for them.

The Grand Forks Test was administered exactly as it is outlined in the test manual.[1]

---

[1]Goheen, Grand Forks Fitness Test, pp. 2-5.

Extended Knee Sit-up

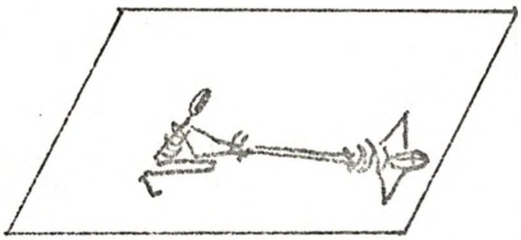


Figure 1.--Starting position.

Figure 2.--Trunk raised and elbow touching the opposite knee.

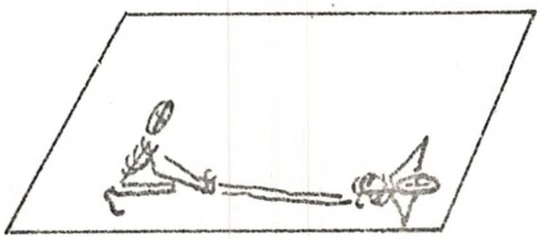


Figure 3.--Finishing position after one complete sit-up.

Pull-up

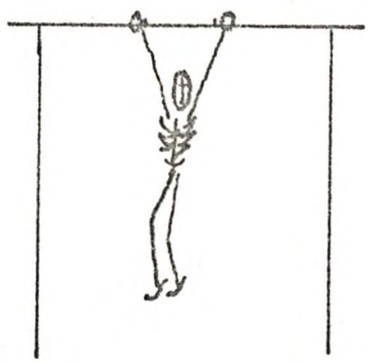


Figure 4.--Starting position.

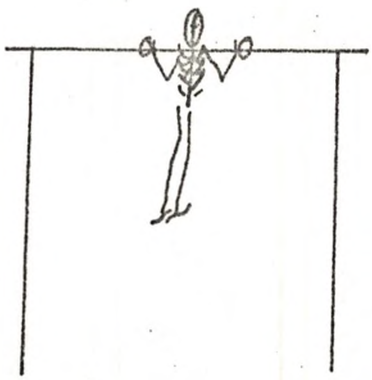


Figure 5.--Elbow flexed position.

Pull-up continued

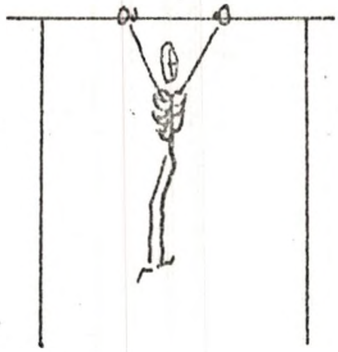


Figure 6.--Return to start.

Shuttle Run

Figure 7.--Diagram of the shuttle run.

Squat Thrust

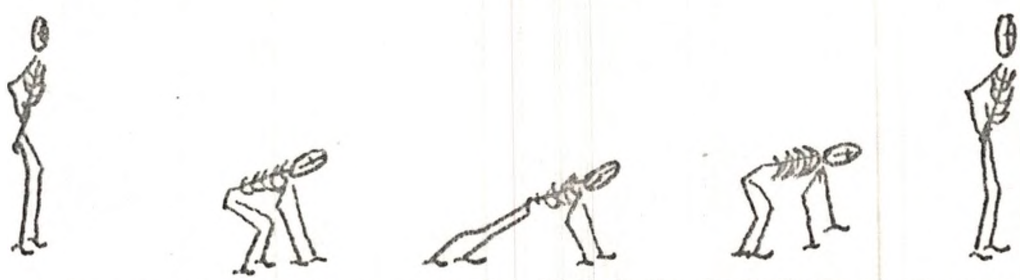


Figure 8.--Squat thrust showing the five body positions necessary to complete the exercise once.

Standing Broad Jump

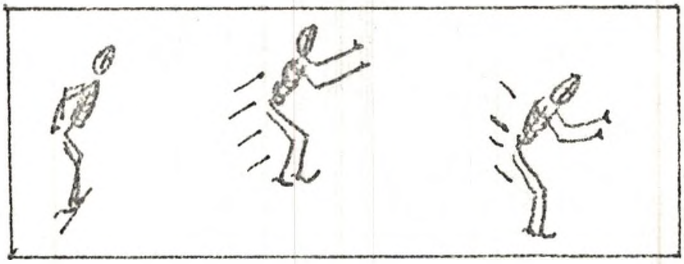


Figure 9.--Diagram of the standing broad jump showing the sequence used to complete one jump.

Vertical Jump

Figure 10.--Preliminary marking position.

Figure 11.--Starting position.

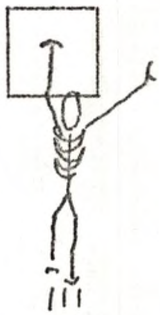


Figure 12.--Extended position.

The criterion test items were administered as described in the various manuals in which they were published. These tests were assumed to be valid measures of their respective criterion because they have been published for use by the entire population.

Modified Pull-up

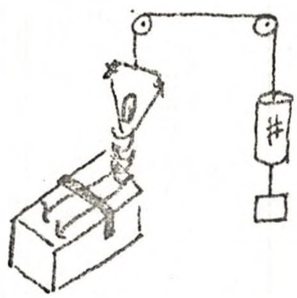


Figure 13.--Starting position.

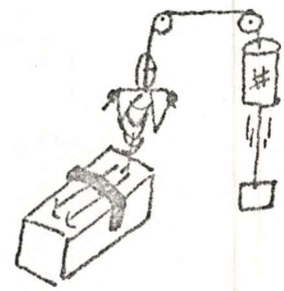


Figure 14.--Working against resistance on the machine.

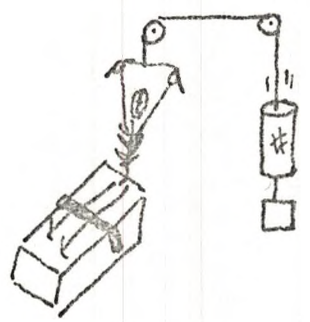


Figure 15.--Return to start.

The subject's weight was recorded on the score card before the testing session began the third morning. Each subject reported to the weight room because the Universal Gym was essential to complete this phase of the test. The subject then strapped himself down to a specially designed bench so his legs would be in a horizontal position and he would not move upward as he pulled down on the resistance. The tester adjusted the machine so that the subject was required to pull down a total resistance equal to three-fourths of his body weight.

Using the overhand grasp, the subjects lowered the bar until it passed below the level of the chin and then the bar returned to starting position by gravity. The subjects repeated the process as many times as possible and the total number correctly done was recorded on the score card.[1]

Full Squat

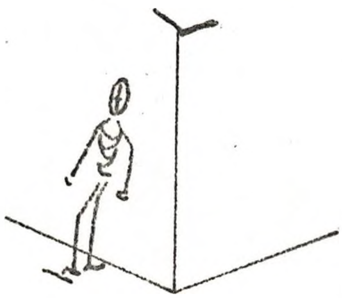


Figure 16.--Determining where to mark the floor.

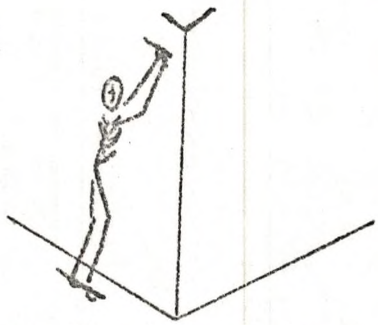


Figure 17.--Determining where to mark the wall and starting position.

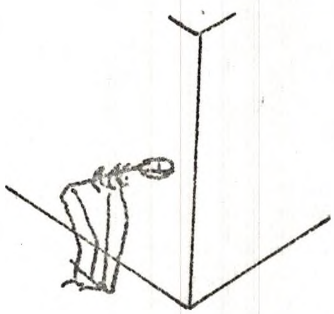


Figure 18.--Squat position.

[1]Soderberg, "The Reliability and Validity of the Modified AAHPER Fitness Test," p. 22.

Each subject lined up next to the wall of the gymnasium with his back and heels against the wall and then walked two foot lengths from the wall. A line was then drawn on the floor behind which the subject had to remain when performing the test. This is illustrated in Figure 16. The subject then faced the wall and keeping his feet behind the drawn line, placed his palms against the wall slightly above shoulder height. A line was drawn on the wall beneath the heels of the subject's hands parallel to the floor. The subject had to place his palms above this line when returning to the upright position after each squat. This position is illustrated in Figure 17, page 22.[1]

In performing the squat, the subject bent his knees and keeping the back straight, placed the palms of his hands against the floor between his feet. When squatting, the buttocks should be kept down and the head held erect with the eyes focused on the wall. The squat position is illustrated in Figure 18, page 22. The subject then stood and placed his hands against the wall assuming the starting position. Only the number of full squats done correctly in a period of one minute was recorded.[2]

Illinois Agility Run

The subject assumed a prone position with his hands at the sides of his chest to start the run. On a verbal signal from the tester, the subject rose quickly and ran through the course diagrammed in Figure 19,

---

[1]Bailey and Orban, The Saskatoon Physical Fitness Test Manual, pp. 5-7.

[2]Ibid.

page 24. The time required to run through the course was recorded on the score card to the nearest half-second.[1]

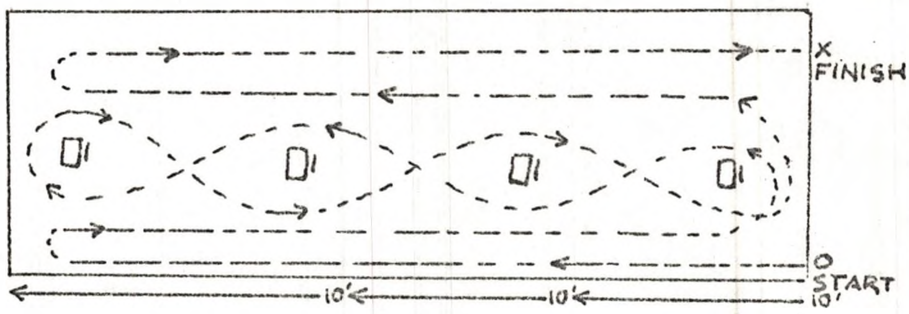


Figure 19.--Diagram of the path to follow in completing the Illinois Agility Run.[2]

Bent Knee Sit-up

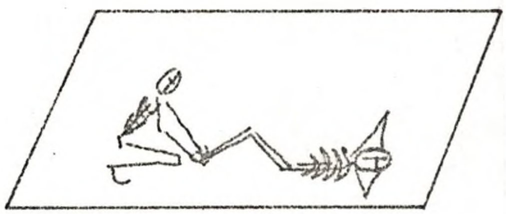


Figure 20.--Starting position.

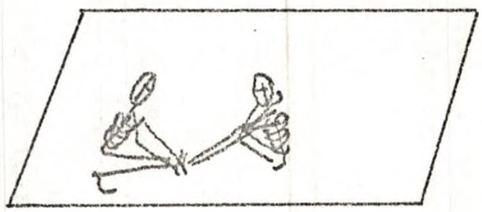


Figure 21.--Trunk raised and elbows touching the knees.

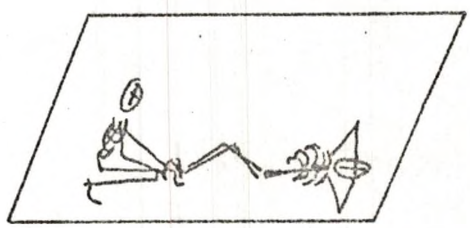


Figure 22.--Return to start.

---

[1]Thomas Kirk Cureton, Physical Fitness Workbook (St. Louis: The C. V. Mosby Company, 1947), p. 23-4.

[2]Cureton, Physical Fitness Workbook, pp. 23, 24.

The subject lay on his back on the mat with his hands behind his head. He then flexed his knees until the soles of his feet were placed flat on the mat and a partner held them down during the exercise. When the starting signal was given, the subject raised his upper body from the mat and touched his elbows to his knees, right elbow to right knee and left elbow to left knee. After touching the knees, the subject lowered himself to the mat, leaving his knees flexed, and touched his head to the mat again before attempting the next sit-up. The number of sit-ups done in two minutes was recorded.[1]

To administer the Grand Forks Fitness Test, six stations were set up around the gym with one tester at each station. When each group had completed the test item at their station, they proceeded to the next station in the sequence. Each group followed the same pattern both days. A similar pattern was also established for the administration of the criterion test, but only four stations were used since the scores derived from the first day's test would be used to validate the standing broad jump and vertical jump.

The data collected on each subject were recorded on individual score cards. Samples of these score cards are shown in Appendix A, pages 46-48. The data were then transferred to tabulation sheets designed for the whole group and the proper statistical procedure was applied.

---

[1]Kendall, "A Criticism of Current Tests and Exercises," pp. 187-97.

CHAPTER III

KINESIOLOGICAL ANALYSIS OF FITNESS TEST ITEMS

The development of the criterion test for measuring validity was dependent on a kinesiological analysis of the items in the Grand Forks Test. This analysis showed that certain muscles were primarily responsible for the movement of the body so that a test measuring the same criterion could be included in the criterion test. As previously stated, Fleishman proposed criterion items with which items of the Grand Forks Fitness Test could be validated.[1]

The analysis specifically points out the major muscles responsible for the performance of the exercise. These muscles, designated as being the major factor involved in a specified joint action, are termed the prime movers. Other muscles, which are used to maintain a definite fixed position of the skeleton during muscular contraction, are termed stabilizers.[2]

Muscles are body organs specifically designed to contract and cause body movement in response to environmental changes. The term contraction simply refers to the development of tension without a muscle.

---

[1]Fleishman, The Structure and Measurement of Physical Fitness, pp. 155-56.

[2]Philip J. Rasch and Roger K. Burke, Kinesiology and Applied Anatomy (3rd ed.; Philadelphia: Lea and Febiger, 1967), p. 47.

There are two major types of contraction involved in any joint action; concentric and eccentric contraction. Concentric contraction occurs when a muscle develops enough tension to visibly shorten and move a body part in spite of a given resistance. On the other hand, during an eccentric contraction, the muscle fibers are contracted but the resistance is great enough to cause the muscle to lengthen. During eccentric contraction, the muscle actually attempts to withstand the total resistance. The identification of these two types of contraction is essential in analyzing any exercise.[1]

In analyzing the squat thrust and full squat, it was determined that the hip, knee and ankle joints were the primary regions of the skeleton affected. To start the exercise, the subject moves with gravity from the anatomical position to the squat position. In the anatomical position, the subject stands erect with the soles of his feet flat on the floor, the arms hanging loosely at the sides and the palms facing forward. Because of gravitational force, the body has a tendency to collapse when in the erect position. Extensor muscles inhibit this tendency. When performing a movement in which the body goes with gravity, the extensor muscle groups contract eccentrically to control this movement.

Flexion of the thigh at the hip joint in performing the squat movement is primarily accomplished by an eccentric contraction of the gluteus maximus and upper portion of the hamstring muscle group. Knee flexion, with gravity, is completed by an eccentric contraction of the quadriceps muscle group. In the ankle joint, dorsal flexion of the

---

[1]Rasch and Burke, Kinesiology and Applied Anatomy, p. 47.

foot is the primary responsibility of the gastrocnemius, soleus, tibialis posterior and the peroneus longus and brevis. These muscles all contract eccentrically.[1]

The second phase of this exercise involves moving from the squat position to the anatomical position in terms of the squat thrust and an upright position with the hands against the wall in the full squat. All of the muscular contractions are of a concentric nature for the prime movers. The gluteus maximus and hamstrings are responsible for extending the thigh at the hip joint. The leg is extended at the knee primarily by the quadriceps muscle group with the gastrocnemius providing assistance. The foot at the ankle joint is plantar flexed as the subject rises due to the contraction of the gastrocnemius, soleus, peroneus longus and brevis and the tibialis posterior.[2]

An analysis of the standing broad jump and vertical jump indicates that the joint actions and muscles primarily responsible for these actions are the same as those described in the second phase of the squat thrust in the preceding paragraph. Eckert, indicated that a definite difference in joint action and angular velocity existed in the completion of the two jumps because the vertical jump emphasizes vertical distance while the standing broad jump stresses both vertical and horizontal distance.[3] Although the same muscles are responsible

---

[1] Clem W. Thompson, Kranz Manual of Kinesiology (4th ed.; St. Louis: The C. V. Mosby Company, 1961), p. 115.

[2] *Ibid.*, p. 116.

[3] Eckert, "Angular Velocity and Range of Motion," p. 939.

for the completion of both jumps, a degree of difference exists concerning the manner in which they are used.

The movement of the arms is also very significant in the completion of each exercise. The shoulder joint is flexed primarily by the concentric contractions of the pectoralis major and the deltoids. The degree of flexion is much greater in the completion of the vertical jump than in the standing broad jump.[1]

The pull-up and modified pull-up exercises involve flexion of the forearm at the elbow, scapular stabilization and extension of the arm at the shoulder. Concentric contractions of the biceps brachii, brachialis and brachioradialis are primarily responsible for the flexion of the forearm at the elbow. The scapula is stabilized mainly due to the action of the rhomboids and trapezius. Because of their action, these muscles are termed stabilizers or fixators when used in the completion of this exercise. Finally, the extension of the arm at the shoulder occurs because of the concentric contractions of the latissmus dorsi, teres major and posterior deltoid muscles.[2]

In completing the sit-up exercise, the abdominal muscles and the hip flexors are utilized. During the extended knee sit-up, as the subject rises, concentric contractions of the rectus abdominis, external oblique and internal oblique muscles are responsible for trunk flexion. Hip flexion is accomplished by the concentric contractions of the iliopsoas, rectus femoris and pectineus assisted significantly

---

[1]Rasch and Burke, Kinesiology and Applied Anatomy, p. 181.

[2]Ibid., pp. 164, 181, 196.

by the sartorius. As the subject returns to the reclining position, eccentric contractions of the same muscles are primarily responsible.[1]

When completing the bent knee sit-up, the same muscles and joints are involved. However, the action of the hip flexors is minimized when the leg is flexed at the knee and the thigh is flexed at the hip. Kendall pointed out that the action of the hip flexors can never be completely removed, but the bent knee position reduces their effect in the completion of the exercise.[2] The abdominal muscles are required to do more work in raising and lowering the trunk when the heels are moved closer to the buttocks and when the subject's feet are not held down by a partner.[3]

The shuttle run and Illinois Agility Run primarily involve the legs because of the running. Running is merely a form of locomotion where gravity is a prime factor. The center of gravity of the body moves outside the base of support and the movements of the lower extremities are used merely as a means of catching the body as it moves forward. The muscles involved in extension and flexion at the hip, extension and flexion at the knee and dorsal and plantar flexion of the ankle, already identified previously, are the prime movers in the running exercise. The muscles involved in plantar flexion of the ankle are especially important in running.[4]

---

[1]Rasch and Burke, Kinesiology and Applied Anatomy, pp. 253, 293.

[2]Kendall, "A Criticism of Current Tests and Exercises," p. 189.

[3]Rasch and Burke, Kinesiology and Applied Anatomy, p. 249.

[4]Thompson, Kranz Manual of Kinesiology, p. 116.

The action of the shoulder joint should not be overlooked in analyzing these exercises. Because of the effects of angular reaction, a forward movement of the leg on one side of the body will result in a backward movement of the same hip. This can be overcome by a forward movement of the arm on the opposite side. Hence, a vigorous forward angular movement of an arm or leg on one side is normally accompanied by a corresponding forward angular movement of the other limb on the opposite side. This is usually called the principle of opposition.

Arm flexion and hyperextension must be vigorous because of the principle of opposition. Flexion of the shoulder joint is accomplished by concentric contractions of the anterior deltoid and pectoralis major. Hyperextension of the joint involved concentric contractions of the posterior deltoid, latissmus dorsi and the teres major. These two joint actions are essential in promoting body balance when running. From the kinesiological point of view, the individuals that lack kinesthetic sense and a controlled balance of the body will be at a distinct disadvantage in performing the shuttle run and Illinois Agility Run.[1]

---

[1]Rasch and Burke, Kinesiology and Applied Anatomy, p. 181.

CHAPTER IV

RESULTS AND DISCUSSION

The data collected in this study were continuous numerical data. The number of subdivisions was limited according to performance. The smallest recorded unit was in tenths. This type of data was also subject to numerical grouping so the data recorded on the score cards were in numerical form. The observed data from the chosen sample were used as a basis for generalizing to the larger untested population. The sample data indicated conditions that existed in the population and the statistics inferred factors relating to the advisability of using the Grand Forks Test to measure fitness.[1]

The Pearson Product Moment correlation procedure was used to analyze the data. Correlation coefficients were computed on the test-retest data as well as on the data recorded from test one and the criterion test. The r value for the test-retest data indicated the degree of reliability for the Grand Forks Test while the r value for test one-criterion test indicated the degree of validity.

The computational procedures for reliability and validity were completed at the Data Processing Center of the University of North Dakota. The data were supplied to an IBM 360/30 computer and run on a

---

[1]M. Gladys Scott, ed. Research Methods in Health, Physical Education and Recreation (2nd ed.; Washington, D.C.: AAHPER), p. 280.

standard program, (Code number S-10307), for their functions. The significance of the r values for reliability and validity was determined from the procedure published in Garrett which indicated the significance level for the calculated correlation value.[1] The .01 level of significance was used in accepting or rejecting the null hypothesis for 40 degrees of freedom.[2]

The null hypothesis of no correlation among the data was the hypothesis on which this study was based. The alternative hypothesis was that there was a correlation among the data.

The calculated correlation value needed for significance at the .01 level with 40 degrees of freedom was 0.393. Any r value above this level was significant and thus the null hypothesis was rejected and the alternative hypothesis accepted.[3]

The means and the reliability coefficients for the test-retest data are illustrated in Table 2, page 34. This table reveals that all of the test items in the Grand Forks Fitness Test are reliable because the calculated r value exceeds the value needed for significance. Thus the null hypothesis was rejected and the alternative hypothesis was accepted for all items of the test. It should be noted that the correlation coefficient for the shuttle run was very low.

---

[1]Henry E. Garrett, Statistics in Psychology and Education (New York: Longmans, Green and Company, 1959), p. 201.

[2]Allen L. Edwards, Statistical Methods (2nd ed.; Chicago: Holt, Rinehart and Winston, Inc., 1967), pp. 211, 266.

[3]Garrett, Statistics in Psychology and Education, pp. 200, 201.

TABLE 2

MEANS, RELIABILITY COEFFICIENTS AND SIGNIFICANCE LEVELS FOR DETERMINING THE RELIABILITY OF THE GRAND FORKS FITNESS TEST

Test Item	Mean Test 1	Mean Test 2	r	Value needed for significance at the .01 level
Extended knee sit-up	54.1	57.2	.70	0.393
Shuttle Run	9.8	9.6	.42	0.393
Standing Broad Jump	81.8	83.2	.85	0.393
Squat Thrust	20.8	20.5	.74	0.393
Pull-up	7.0	6.8	.93	0.393
Vertical Jump	19.7	19.8	.89	0.393

The means and validity coefficients for test one-criterion test data are shown in Table 3, page 35. Since the calculated value was greater than the table value, the null hypothesis was rejected for all items except the shuttle run. In the shuttle run, the correlation value of 0.37 was less than the value needed for significance, therefore, the null hypothesis was accepted.

TABLE 3

MEANS, VALIDITY COEFFICIENTS AND THE SIGNIFICANCE FOR EACH TEST ITEM USED IN DETERMINING VALIDITY

Test Item	Means	r	Value needed for significance at the .01 level
Extended Knee Sit-up	54.1	.64	0.393
Bent Knee Sit-up	65.8		
Shuttle Run	9.8	.37	0.393
Illinois Agility Run	16.8		
Standing Broad Jump	81.8	.84	0.393
Vertical Jump	19.7		
Squat Thrust	20.8	.45	0.393
Full Squat	51.6		
Pull-up	7.0	.69	0.393
Modified Pull-up	14		

Table 4, page 36, graphically presents the means from the test-retest data. It also indicates any changes and the direction of those changes that have occurred in the value of the mean.

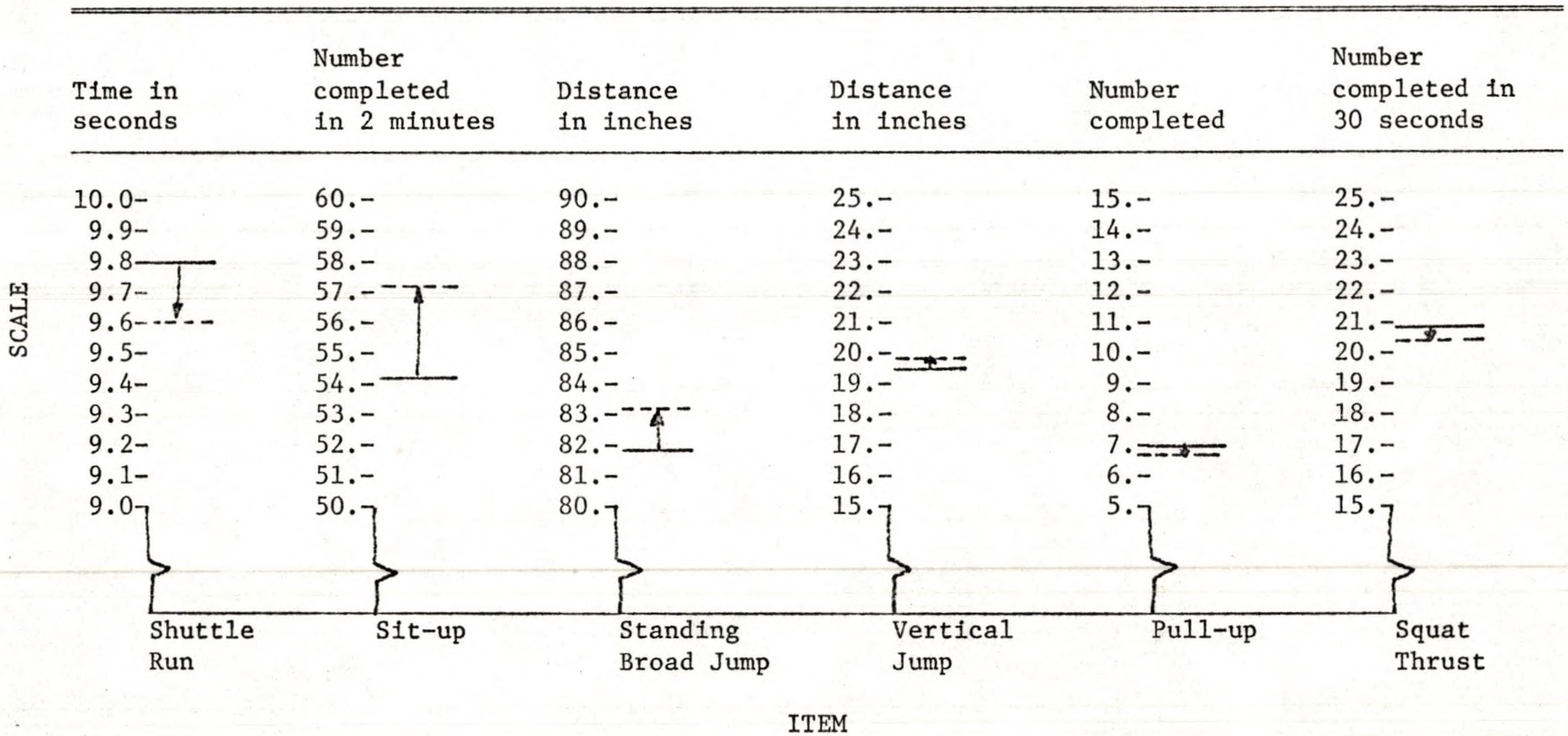
TABLE 4

GRAPH OF MEANS FROM TEST-RETEST DATA

KEY

TEST ONE MEAN _____

TEST TWO MEAN -----



The means obtained from the test-retest of the extended knee sit-up were 54.1 and 57.2 repetitions respectively. The statistical analysis revealed a correlation value of 0.70 for the two tests. Clarke has stated that this correlation value was considered to be rather low, however, it was designated as adequate for group measurement.[1] The reliability coefficient was above the level required for significance indicating the extended knee sit-up was a reliable test as administered.

A validity coefficient of 0.64 was obtained when comparing the mean of the extended knee sit-up from test one with the mean of the bent knee sit-up in the criterion test. The bent knee sit-up item of the criterion test had a mean of 65.8 repetitions. The statistical analysis indicated that the extended knee sit-up was a valid measure because the validity coefficient exceeded the value needed for significance and therefore the null hypothesis was rejected.

The bent knee position minimized the effect that the hip flexors had in raising the trunk. The reduction in activity of the hip flexors was accompanied by an increase in activity of the abdominal muscles. The action of the hip flexors can never be completely eliminated, but the bent knee position significantly reduced their role in raising the trunk. This evidence indicated that the bent knee sit-up was a much better test for measuring abdominal strength than the extended knee sit-up.[2,3]

---

[1] Clarke, Application of Measurement, p. 36.

[2] Kendall, "A Criticism of Current Tests and Exercises," p. 189.

[3] Rasch and Burke, Kinesiology and Applied Anatomy, p. 249.

The mean calculated from the data recorded in test one for the shuttle run was 9.8 seconds. The test two mean was 9.6 seconds. The value of these means was almost identical, yet the reliability coefficient was 0.42. This value was below the value of 0.70 which Clarke considered to be an adequate statistical measure for determining reliability.[1] Even though the reliability coefficient was low, it was still great enough to exceed the value needed for significance in this study. Hence, the shuttle run was considered to be a reliable test item as administered.

In determining validity, a coefficient was computed considering the test one mean of the shuttle run, which was presented above, and the mean from the Illinois Agility Run which was calculated to be 16.8 seconds. The computed validity coefficient was 0.37. This value was less than the value required for significance which meant that the shuttle run had to be considered invalid because the null hypothesis had to be accepted. This also led to the conclusion that these tests measured different criterion and because of this they may both be used in the same physical fitness test battery.

The reliability coefficient computed for the sdtanding broad jump was 0.85. The test one mean was 81.8 inches while the test two mean was 83.2 inches. This test was accepted as a reliable test since the reliability coefficient exceeded the value needed for significance and the null hypothesis was rejected.

The two means for the test-retest of the vertical jump were 19.7 inches for test one and 19.8 inches for the retest. The computed

---

[1] Clarke, <u>Application of Measurement</u>, p. 36.

value of the reliability coefficient, 0.89, was significantly higher than the required value. Therefore, this test was considered to be reliable.

The recorded scores from the first administration of the standing broad jump and vertical jump were used to statistically determine the validity of these tests. This was done because the kinesiological analysis indicated that the same muscles were involved in the performance of both tests. Fleishman also indicated that they both measure the same criterion.[1] The calculated validity coefficient of 0.84 was of sufficient magnitude to indicate that both of these tests measure the same criterion and therefore they should not be included in the same test battery.

The means of the two tests for determining the reliability of the pull-up were recorded at 7.0 repetitions for test one and 6.8 repetitions for the retest. The pull-up item was found to have a reliability coefficient of 0.93 which was the highest calculated for all the test items and thus the pull-up was deemed a reliable test.

Kendall concluded that this test was too difficult to be included in a fitness test battery because there were some subjects who cannot even complete one pull-up. She emphasized that any good fitness test would not have any zero scores recorded. There were three subjects who failed to complete one pull-up in the administration of the Grand Forks Test and therefore the test item was too difficult according to Kendall's specifications.[2]

---

[1]Fleishman, The Structure and Measurement of Physical Fitness, p. 155.

[2]Kendall, "A Criticism of Current Tests and Exercises," p. 189.

The mean calculated for the modified pull-up was 14 repetitions. When comparing the data from this test to the data from test one, a validity coefficient of 0.69 was computed. This value indicated that the pull-up was a valid test. However, the validity coefficient was less than adequate for group measurement in fitness testing by Clarke's standards.[1]

In the administration of the modified pull-up, there were no zero scores obtained which would satisfy that argument against using the test in a fitness test battery. However, the time needed and the cost of the equipment necessary to administer the test, made it impractical to include in a fitness test battery. Kendall again emphasized that a good fitness test should be easy to administer and be feasible in terms of the time and cost of equipment required for the administration.[2] These conditions were not satisfied in this study so the designed, modified pull-up could not be included as a basic part of any fitness test battery without revisions in the administrative procedure.

The test-retest means calculated for the squat thrust were 20.8 repetitions for test one and 20.5 repetitions for test two. This test item had a computed reliability coefficient of 0.74 which rendered it reliable. By the standards presented in Clarke, the value of 0.74 is categorized as only being adequate. Values higher than this are much more desirable.[3]

---

[1]Clarke, Application of Measurement, p. 36.

[2]Kendall, "A Criticism of Current Tests and Exercises," p. 189.

[3]Clarke, Application of Measurement, p. 36.

In determining validity, the mean of test one of the squat thrust item was compared to the mean of the full squat test which had a calculated value of 51.6 repetitions. The value of the validity coefficient (0.45) exceeded the value needed for significance (0.393). Thus, the null hypothesis was rejected and the squat thrust item was deemed valid.

It should be noted that even though the validity coefficient was significant, its value was very low. The squat thrust involves more body movement to complete one exercise than the full squat. The subject has to squat, vigorously extend his legs backward, return to squat position and finally raise himself to the anatomical position to complete one repetition of the exercise. When completing the full squat, the subject merely squats and returns to the starting position. There was obviously more agility and body coordination involved in performing the squat thrust when compared to the full squat which could account for the low coefficient.

The computed reliability coefficient for the shuttle run, (0.42), was an interesting statistic. The mean of test one was 9.8 seconds and the mean of test two was 9.6 seconds. One might think that means this close would support a substantially higher correlation value. However, the correlation coefficient seemed to refute the closeness of the means. A correlation like this would be calculated when the subject's score differently during the two days of testing. Some of the subjects scored better the second day and some had a higher score the first day. The calculated means were almost identical but because of the difference in scoring by the subjects, the correlation value was low.

The pull-up test exhibited a sharp contrast to the statistical analysis of the shuttle run. The mean from test one was 7 repetitions and the mean from test two was 6.8 repetitions. These means are almost identical and the calculated correlation value was 0.93. In this example, the correlation value substantiated the closeness of the means.

Clarke pointed out that if validity was to be computed by the correlation technique, the desirable relationship should be .9 or above. However, .8 would be considered a significant value. Only the validity coefficient for the standing broad jump and vertical jump satisfy this condition in this study. The remainder of the items have coefficients too low to be significantly acceptable by Clarke.[1]

Fleishman, in evaluating the AAHPER Fitness Test according to his criteria for statistical reliability, found that several of the test items, individually, came out very well. However, a fitness test battery cannot be evaluated in terms of the individual tests alone; rather, the key factors are the comprehensiveness and efficiency of the test as a battery.[2]

According to his criteria, the Grand Forks Fitness Test would only measure two factors well. These would be dynamic arm strength in the pull-up and explosive leg strength in the standing broad jump. The trunk strength factor was measured only imperfectly by the

---

[1]Clarke, Application of Measurements, p. 36.

[2]Fleishman, The Structure and Measurement of Physical Fitness, p. 150.

extended knee sit-up. In addition to this, the explosive strength factor was overemphasized in the test because three of the six test items were designed for this purpose. Also, fitness tests should not have more than one test measuring any given fitness criterion in the same battery.[1]

The Grand Forks Fitness Test, as administered at Central High School, also failed to include any specific measures of static strength, flexibility, coordination, cardio-respiratory endurance and body balance. Therefore, physical fitness, as defined by Fleishman, would not be adequately measured in the Grand Forks Fitness Test.[2]

---

[1]Fleishman, The Structure and Measurement of Physical Fitness, p. 36.

[2]Ibid.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The lack of fitness exhibited by American men tested for induction into the military service prompted the development of fitness tests to assay the physical fitness of our nation's youth. The Grand Forks Test was designed to determine the level of fitness of the youth in the Grand Forks Public Schools. This study was formulated to determine the reliability and validity of the Grand Forks Fitness Test.

A stratified random sample of 42 boys from Central High School completed the testing phase of this study. Each subject was tested on three consecutive mornings. The Grand Forks Test was administered the first two mornings and a test-retest analysis by the Pearson Product Moment correlation technique was used to determine whether the difference between the means was significant. In order to validate the Grand Forks Test, a criterion test was administered the third day and the data compared to the scores recorded from the first test session. The criterion test was designed on the basis of a kinesiological analysis of the test items and as a result of Fleishman's criterion measure for each test item.

The recorded data were programmed into an IBM computer and the correlation coefficients were calculated. The significance of these coefficients was tested by comparing them to correlation values published in Garrett at the .01 level. The correlation coefficients indicated that all items in the Grand Forks Test were reliable. With the exception of the shuttle run, all of the items were found to be valid.

Conclusions

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

1. All the reliability coefficients were significant at the .01 level which indicated that the test items were reliable.

2. The validity coefficient for the shuttle run was lower than the value needed for significance and therefore the shuttle run was deemed invalid.

3. All other validity coefficients were significant at the .01 level which indicated that those test items were valid.

4. The reliability coefficients for the extended knee sit-up, (.70), shuttle run, (.42), and squat thrust, (.74), were lower than the generally accepted standard for group measurement as recommended by Clarke.

5. All the other reliability coefficients were above acceptable standards.

6. The standing broad jump--vertical jump validity coefficient was the only one above the accepted standard.

7. The validity coefficient for the standing broad jump--vertical jump indicated that these two tests were highly correlated and therefore they should not be in the same test battery.

8. The validity coefficient for the shuttle run--Illinois Agility Run indicated that these tests measure different criterion.

Recommendations

1. The bent knee sit-up should replace the extended knee sit-up as a measure of abdominal strength because of the reduction in the action of the hip flexors in the flexed knee position.

2. Additional research was indicated for the shuttle run to determine more accurately its exact criterion of measure.

3. The modified pull-up test used in this study was too costly and time consuming to apply to general fitness testing. Revisions should be made in the administrative procedure, if a Universal Gym is available for use, when considering the use of this test.

4. The overemphasis on leg measures in the Grand Forks Test warrant the replacement of some test items. The vertical jump and shuttle run should be replaced by tests that measure the criterion in fitness testing that are not found in this test to make it more complete.

5. The full squat test would be more desirable as a specific measure of dynamic leg strength than the squat thrust which is presently being used in the Grand Forks Test.

APPENDIX

Name________________________________ Class___________________
Weight_________________

Grand Forks Youth Fitness Test

Individual Score Card

Test Item	Test 1	Test 2
1. Vertical Jump (record height to nearest inch in best of 2 tries)	First / Second / Final	First / Second / Final
2. Pull-ups (Record number completed)		
3. Burpees (Record number per 30 seconds)		
4. Shuttle Run (Record time to the nearest 1/10 second)		
5. Standing Broad Jump (Record distance to the nearest inch in the best of 2 trials)	First / Second / Final	First / Second / Final
6. Sit-ups (Record number completed in 2 minutes		

Name______________________ Class___________
Weight___________

Test Battery for Validation

Individual Scord Card

Test Item	Test Score
1. Pull-ups (Record number completed	
2. Squat reach (Record number done in one minute)	
3. Illinois Agility Run (Record time to the nearest 1/10 of a second)	
4. Bent-knee sit-ups (Record number done in two minutes)	

REFERENCES

SELECTED BIBLIOGRAPHY

Books

American Association of Health, Physical Education and Recreation. AAHPER Youth Fitness Test Manual. Washington, D.C.: AAHPER, revised edition, 1961.

Bailey, D. A., and Orban, W. A. R., Dr., Study Directors. The Saskatchewan Physical Fitness Test Manual. Saskatoon: Saskatoon Public School Board, 1962.

Clarke, H. Harrison. Application of Measurement to Health and Physical Education. 3rd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1959.

Cureton, Thomas Kirk. Physical Fitness Workbook. St. Louis: The C. V. Mosby Company, 1947.

Edwards, Allen L. Statistical Methods. 2nd ed. Chicago: Holt, Rinehart and Winston, Inc., 1967.

Garrett, Henry E. Statistics in Psychology and Education. New York: Longmans, Green and Company, 1959.

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

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. 3rd ed. Philadelphia: Lea and Febiger, 1967.

Scott, M. Gladys, ed. Research Methods in Health, Physical Education and Recreation. 2nd ed. Washington, D.C.: AAHPER, 1959.

Thomas, Clem W. Kranz Manual of Kinesiology. 4th ed. St. Louis: The C. V. Mosby Company, 1961.

Journals and Magazines

American Association of Health, Physical Education and Recreation. "The President's Council on Youth Fitness." Journal of Health, Physical Education and Recreation, XXXI (September, 1960), 354.

Clarke, H. Harrison, and Degutis, Ernest W. "Relationships Between Standing Broad Jump and Various Maturational, Anthropometric and Strength Tests of 12-Year-Old Boys." Research Quarterly, XXXV (March, 1964), 258-64.

Eckert, Helen M. "Angular Velocity and Range of Motion in the Vertical and Standing Broad Jumps." Research Quarterly, XXXIX (December, 1968), 937-42.

Glencross, D. J. "The Nature of the Vertical Jump Test and the Standing Broad Jump." Research Quarterly, XXXVII (October, 1966), 353-59.

Hein, Fred V. "What is Physical Fitness?" National Education Association Journal, LI (February, 1962), 34.

Johnson, Barry L., and Nelson, Jack K. "Effect of Different Motivational Techniques During Training and Testing Upon Strength Performances." Research Quarterly, XXXVIII (December, 1967), 630-36.

Kaufman, Raymond. "Our Unique Contribution." Journal of Health, Physical Education and Recreation, XXXI (September, 1960), 26-27.

Keeney, Clifford E. "Work Capacity." Journal of Health, Physical Education and Recreation, XXXI (September, 1960), 29-30.

Kendall, Florence P. "A Criticism of Current Tests and Exercises for Physical Fitness." Journal of the American Physical Therapy Association, XXXV (March, 1965), 187-97.

Kennedy, John F. "The Soft American." Sports Illustrated, December 26, 1960, pp. 15-17.

Knuttgen, Howard G. "Comparison of Fitness of Danish and American School Children." Research Quarterly, XXXII (May, 1961), 190-96.

Kraus, Hans, and Hirschland, Ruth P. "Minimum Muscular Fitness Tests in School Children." Research Quarterly, XXV (May, 1954), 178-86.

McCraw, Lynn W. "Effects of Variation of Forearm Position in Elbow Flexion." Research Quarterly, XXXV (December, 1964), 504-09.

Nelson, Dale O. "Focus on Two Fitness Exercises." Journal of Health, Physical Education and Recreation, XXXV (May, 1964), 22-23.

Strong, Clinton H. "Motivation Related to Performance of Physical Fitness Tests." Research Quarterly, XXXIV (December, 1963), 497-507.

Wireman, Billy O. "Comparison of Four Approaches to Increasing Physical Fitness." Research Quarterly, XXXI (December, 1960), 658-666.

Unpublished Materials

Goheen, Royal, et al. "Grand Forks City School Youth Fitness Test," Grand Forks, North Dakota, 1964. (Mimeographed.)

Soderberg, Paul. "The Reliability and Validity of the Modified AAHPER Fitness Test as Applied to Freshman College Males at the University of North Dakota." Unpublished Master's thesis, University of North Dakota, 1969.

Sparks, Raymond E. "Modification of the AAHPER Youth Physical Fitness Test." Unpublished Master's thesis, Springfield College, 1965.

Interviews

Gambucci, Serge. Head of the Physical Education Department at Central High School, Grand Forks North Dakota, personal interview, June 21, 1968.