DOCUMENT RESUME

ED 103 411 SP 009 018

| | |
|---|---|
| AUTHOR | Jackson, Andrew S. |
| TITLE | An Evaluation of the AAHPER Youth Fitness Test. |
| PUB DATE | Mar 75 |
| NOTE | 10p.; Paper presented at the Evaluation Section of the Annual Meeting of the American Alliance for Health, Physical Education, and Recreation (Atlantic City, New Jersey, March 1975) |
| EDRS PRICE | MF-$0.76 HC-$1.58 PLUS POSTAGE |
| DESCRIPTORS | *Athletic Programs; Elementary Education; Exercise (Physiology); Motor Development; *Muscular Strength; *Physical Education; *Physical Fitness; *Psychomotor Skills; Running; Secondary Education; Student Ability; Teaching |
| IDENTIFIERS | *AAHPER Youth Fitness Test |

ABSTRACT

From its initial development, the AAHPER Youth Fitness Test has been criticized for (a) not measuring only physical fitness components; (b) forcing performances that may be injurious to students; and (c) not accurately measuring aerobic endurance, a major goal of the tests. The focus of this study is to approach these criticisms and through discussion estimate how well the AAHPER Youth Fitness Test measures physical fitness. Test validity is examined by determining what traits or factors are measured by the battery, and by confronting the battery's ability to measure a known and valued physical fitness variable, maximal oxygen uptake. Figures attached include a model designed to define the components of the motor performance domain, examples from the Texas Physical Fitness Motor Ability Test, a Factor Analysis of Running Tests, and correlations between maximum oxygen uptake and AAHPER test items. A list of references is included. (JS)

# AN EVALUATION OF THE AAHPER YOUTH FITNESS TEST*

Andrew S. Jackson
Department of HPE
University of Houston
Houston, TX

It was E.L. Thorndike who said, "If a thing exists, it exists in some amount. If it exists in some amount, it can be measured." The "thing" of interest today is physical fitness. The task is to estimate how well the AAHPER Youth Fitness test does measure physical fitness.

From its initial development in 1957, the AAHPER test has been criticized by teachers, students, kinesiologists, exercise physiologists, measurement specialists, and many other definable groups. Yet, the battery has survived and millions of Americans have been tested. The test items can be objectively scored and several researchers have reported that the test items are reliable. A reliable test in this context, means that individual differences of something can be measured with a defined degree of error. Thus, the important question becomes: What are the "things" being objectively and reliably measured by the AAHPER battery? This is a question of validity and will be the focus of this paper.

The validity of the AAHPER battery will be examined from two different, but related perspectives. First, the construct validity of the battery will be examined. Construct validity will help determine the "things," traits, factors or constructs measured by the battery. This is a basic question of definition. Once the factors of fitness have been defined, concurrent validity can be used to estimate the battery's ability to measure a known, accepted, and valued physical fitness variable, maximal oxygen uptake. Concurrent validity questions the efficiency of the battery to measure this accepted "thing."

A logical approach was used to develop the AAHPER battery. This approach involved the definition of fitness components and selection of tests to measure the defined components. Both the components and test selection was based on the logical opinions of several judges. The six items of the initial AAHPER test supposedly measured strength, endurance, and proficiency in running, jumping, and throwing. Two questions to ask are: 1) Are these valid physical fitness components? 2) Are the tests correlated with or valid measures of the defined component? This is what construct validity is all about.

Clarke (1967) published a model (Figure 1)[1] designed to define the components of the motor performance domain. This categorization system, most likely, represented the thinking of the individuals that drafted the AAHPER battery. As I hope you can see on the screen, Clarke has defined general motor ability, motor fitness, and physical

---

*Presented at the Evaluation Section, AAHPER National Convention, Atlantic City, March, 1975.

[1]See Clarke, p 202.

fitness in terms of these logically deduced components. A common criticism leveled at the AAHPER battery is that it does focus on just physical fitness components. For example, it was rumored that softball throw for distance was included because the battery was "to fitness oriented at the expense of skill." This reflects interesting test construction logic.

A more serious criticism, however, lies with the use of Clarke's model for the construction of valid motor performance test batteries. As an illustration, the research from industrial psychology and motor learning reflect a need to alter our view on one of our cherished "things", general motor ability.

Clarke's model has been useful for the logical definition of motor performance terms. The model is not, however, useful for developing test batteries because "pure" components such as strength, endurance, and soforth, have not been isolated through scientific research. This may be illustrated by example. Flexibility may be defined as the range of movement in a joint or sequence of joints. Any test that fits this definition may be categorized as a flexibility test; however, the intercorrelations among several flexibility tests is near zero which indicates that different types of flexibility exist. In fact, the results of a study published by Margaret Harris (1969) revealed that a single factor of flexibility does not exist and there may be 13 or more different types of flexibility. This indicates that if one wants to measure individual differences in flexibility, at least 13 different tests could be used.

What I am saying is that the AAHPER test evolved from a logical rather than scientific model. The essence of the scientific approach is to objectively test logic. I am suggesting the need for the construct validity model used by psychologists to define such domains as personality. The Cattell 16 personality factor questionnaire which is used by physical educators evolved through this scientific approach. As applied to defining the physical fitness domain, the approach requires the researcher to first hypothesize basic physical fitness factors. Next several tests suspected to measure each factor are administered to a defined sample of subjects. The data are statistically analyzed with factor analysis to confirm or reject ones logic which are the hypothesized factors. This approach is useful because the basic factors or constructs of the domain are scientifically identified, and the construct validity of tests can be established. In this context, construct validity is the correlation between a test item and an identified factor. By this method, it is possible to identify several tests that validly measure the same factor. Fleishman's book The Structure and Measurement of Physical Fitness provides an example of this approach applied to the motor performance domain.

Since Fleishman's study (1964) several physical educators, some of which are present today, have conducted factor analytic studies designed to define the motor performance domain. These recent factor analytic studies, with Fleishman's original study, provided the research base for the recently developed Texas Physical Fitness - Motor Ability. This test was developed by the Texas Governor's Commission on Physical Fitness (Baumgartner and Jackson 1975). The basic factors and test items appear on the screen (Figure 2). The test was developed in 1973, and since it was developed due to the dissatisfaction with the AAHPER Test, I would like to use the battery for comparative purposes.

A basic criticism of the AAHPER battery is the test does not measure just physical fitness components. For the Texas Test, the physical fitness and motor ability factors were separated. Running speed, agility, jumping proficiency, and throwing proficiency were not considered fitness factors. I feel this is important because a test battery is the principle measure by which the concept of physical fitness is communicated to our students. What we are saying is that if you can run fast or throw far you are physically fit. While at Indiana University I had the opportunity to have several world class swimmers in some of my classes. Many of the athelets were slow a-foot and could not throw a ball with skill, but they were physically fit. This is a basic issue concerning validity.

The physical fitness components of the Texas Test have been identified through scientific research. It is important to understand that a general factor is being measured rather than a "pure" trait such as strength or endurance. The listed tests are valid tests of this general factor; thus, these tests possess construct validity, or in other words, the tests are correlated with the general factor. Several different tests may be used to measure the same factor.

Let's turn our attention to the fitness factors measured by this battery. The factor muscular strength and endurance of the arms and shoulder girdle involves the ability to move or support one's body weight with the arms. Both strength and endurance are needed to execute these tests, but to varying degrees for different people. This factor is negatively correlated with body weight.

It is difficult to say that these tests measure either strength or endurance. Let me illustrate what I mean. The tests dips and bench press are used to evaluate performance in our body conditioning course at the University of Houston. For the bench press test, a 110 pound weight load is used for all subjects and the test is scored by the number of repetitions the student can execute to a set cadence. Note, both the bench press and dip tests are performed to exhaustion which is endurance according to Clarke's model. However, the correlation, using over 500 college men was low, only .32. This indicates that the tests were statistically measuring different factors. When the repetition bench test was correlated with the maximum weight that the student could lift, which would be defined as strength, the correlation was over .90.

It is important to understand that tests such as dips, or chins involve moving one's body mass; thus, the low correlation between dips and bench press was partly due to indiviudal differences in body weight. For the sample studied, the correlation between body weight and dips was a negative .46 while the correlation between bench press and body weight was a positive .39. In essence, body weight suppressed the correlation between dips and bench press. When body weight was held constant, the partial correlation between these two tests was significantly higher, .61. What I am saying is that the test dips obviously involves both strength and endurance, but it is used to move one's body weight.

Another criticism of the AAHPER Battery was that the items straight leg situp and softball throw for distance may be injurious to a student. The softball throw was excluded from the Texas test and the bent-knee situp was used in place of the straight leg situp. Numerous articles have been written on the potential hazards of these items and it is not my purpose to review these. Rather, I would like to continue with my examination of the statistical validity of the tests.

Using electromyography, kinesiologists have reported that different muscle groups are used to execute the straight leg and bent-knee situps. Kinesiologically this is true, but statistically the tests measure the same basic factor which is the ability to move the body mass with the muscles in the abdominal region. What I am saying is that these performances are correlated. Thus, if the straight leg situp is potentially harmful, why not use bent-knee situp?

Fitness tests not only are used to evaluate student performance, they communicate to the students the types of things deemed important. At this point, I think it is important to ask: Should our fitness programs be designed to develop these first two fitness components? Since both involve the student's ability to move his body weight, performance may be improved two basic ways. First, if body mass is held constant, higher levels of strength and endurance will yield higher scores. Second, if strength and endurance are held constant, shedding body weight will yield higher performance. The reason many students do poorly on these tests may not be due to low strength, but that students are too fat. I have heard many teachers say that the test chins is unfair because it penalizes the "heavy-set boy." The heavy-set youngster, in reality, tends to be an obese youngster, and the health hazards associated with obesity are well known.

In my opinion, the most serious criticism of the AAHPER Battery is that the 600 yard run is to short, and thus, does not measure aerobic endurance. This obviously reflects the aerobics concept advanced by Dr. Kenneth Cooper, who I might add, is a member of the Texas Governor's Commission. Numerous studies designed to examine the validity of distance run tests have been recently published. I would like to turn my attention to these studies.

The majority of these studies have examined the concurrent validity of distance run tests. The strategy has been to examine the correlation between maximal oxygen uptake and distance run performance. I will discuss these studies in a minute, but first I would like to examine the issue that distance runs should be longer than 600 yards.

In a study (Disch, 1975) to be published in the Research Quarterly, we examined the construct validity of distance run tests. The factor analytic findings of this study are presented in Table 1. These ten running tests were administered to 60 college males and the analysis revealed that only two basic factors were needed to measure individual differences in running performance. For those of you who are not familiar with factor analysis, the values in columns F-1 and F-2 represent the correlation between the test and general factor. The values listed in the column $h^2$ may be interpreted as lower bound reliability estimates. The correlations underlined are significantly different from zero. The basic factors measured by these ten running tests are individual differences in running speed and the qualities needed to run long distances. The "purest" measures of the distance run factor are 1.25 miles or longer while shorter runs measure both speed and distance run endurance. The 600 yard run was not part of this study, but one could hypothesize that the test would be substantially correlated with the speed factor. This tends to confirm the criticism that distance runs longer than 600 yards are needed.

These findings add credence to Cooper's recommendation of using the 1.5 mile or the 12-minute run for distance; however, these findings suggest that other distances may be used to measure this same factor. We have identified these same factors with samples of elementary school children. This latter study (Jackson, 1975) is being reported at a research session of this convention. These two studies provided the research base for including the multiple distances of the Texas test.

Thus far, my concern has been to provide a statistical definition of physical fitness. This approach is useful for identifying basic factors by which individual differences may be reliably measured. Now lets more fully examine what these factors are measuring. Maximum oxygen uptake is considered by exercise physiologists to be a valid criterion of physical fitness. The concurrent validity of a test is estimated by correlating the test with maximum oxygen uptake.

Metz and Alexander (1970) examined the concurrent validity of AAHPER test items and these correlations are reported in Table 2. For both age groups, the tests pullups, shuttle run, standing broad jump, and 50 yard dash were significantly correlated with maximal oxygen uptake. For all tests, superior motor performance was associated with higher aerobic capacities. On the surface this is somewhat difficult to understand because maximal oxygen uptake supposedly measures an individual's ability to continue exhausting work.

A study conducted by Falls and Associates (1966) sheds light on the nature of these significant correlations. The Falls study used 87 adult males and these correlations are reported in Table 3. The first column represents the correlation between the motor performance tests and maximum oxygen uptake per kilogram of body weight. All these correlations were significantly larger than zero. Falls also measured lean body mass and then calculated maximum oxygen uptake per kilogram of lean body weight and these correlations are in column two. As you will note, when the component of body fat was removed from body weight, all the correlations were lowered. The sample that Falls studied ranged in age from 23 to 58 years. It has been established that lower levels of maximum oxygen uptake and motor performance are related to older age. It is my suspicion that the remaining significant correlations represented in column two are due to individual differences in age.

Broadly speaking, a test is a valid measure of anything that it is correlated with. According to Newton's second law of motion, force is equal to mass times acceleration. Thus, if one gets fatter, mass is increased but force remains constant and thereby, lowering acceleration. This was statistically confirmed in a recent doctoral study (Williams, 1974) at the University of Houston. With a sample of junior high school boys, performances on six speed, jumping, and agility test were significantly correlated with body fat. The correlations ranged from .36 to .47 and reflected a negative relationship -- high body fat was associated with poor motor performance.

One could not accurately predict percent body fat with correlations of this magnitude, but this supports the hypothesis that the significant correlations reported by Metz and Alexander were due to oxygen uptake was divided by body weight and not lean body mass. Thus, the fatter boy would have a lower oxygen uptake value because of the additional weight. Body fat would be a common course of variance for both tests and significant correlations would make statistical sense.

Numerous investigators have reported significant correlations between maximum oxygen uptake and running ability. Using the factor analytic findings previously reported as a guide, running tests were categorized into two groups: runs of one mile or shorter; and runs longer than one mile, including the 9 and 12-minute runs for distance. The range and median product-moment correlations are presented, on the next slide.[2] As you can see, the reported correlations for the longer runs are considerably higher. Of the 11 correlations for tests of 1 mile or shorter, only 5 were significantly different from zero. Of the 16 correlations for longer distances, 15 were significantly different from zero. The nonsignificant correlation was reported with a sample of 17 college cross country runners. With such a small, homogeneous sample, a nonsignificant correlation would be expected. Of interest, five correlations were higher than .80.

It is unrealistic to expect distance run tests to duplicate the maximal oxygen uptake measured in the laboratory. However, as previously mentioned, longer distance runs provide a factor by which individual differences may be reliably measured. The correlations summarized on the screen support the hypothesis that aerobic working capacity is being significantly measured to some extent. Undoubtedly, such things as motivation and experience effect a student's distance run performance.

Experimental studies offer evidence confirming the relationship between distance run performance and aerobic working capacity. Numerous training studies have been published with running as the independent variable and maximal oxygen uptake and body composition as dependent variables. The findings of these studies reveal that when the training program is of sufficient intensity and duration, running performance is improved, maximum oxygen uptake increased, and body fat lowered. Many feel that a correlation coefficient must be reported to establish validity. It is my opinion that these experimental studies offer the strongest evidence supporting distance run tests. Improved running performance produces desired physiological changes. What more does one want?

Presented on the screen are the items of the 1975 AAHPER revised battery. It is my candid opinion that this test did not evolve through the scientific process. In fact, I feel that the Texas Test and the new California State Test are responsible for the revision. To support this contention, the new bent-knee situp test and optional distance run tests were taken, by permission, from these state batteries. The norms for these tests which are published in the 1975 AAHPER manual, are norms developed with and for California and Texas children. However, there is no credit given to these states for their efforts. For all intent and purpose the norms are "passed-on" as National Norms. Professionally, I consider this act of omission to be unethica as "hell." Maybe this is my biggest criticism of the AAHPER st -- it has been politically rather than scientifically motivate.

## REFERENCES


Baumgartner, Ted A. and Jackson, Andrew S. Measurement for Evaluation in Physical Education. Boston: Houghton-Mifflin Company, 1975.

Clarke, H. Harrison. Application of Measurement to Health and Physical Education. Englewood Cliffs: Prentice-Hall, Inc., 1967.

Disch, James, Jackson, Andrew S. and Frankiewicz, Ronald. "Construct Validation of Distance Run Tests," Research Quarterly, to be published.

Falls, Harold B., and others. "Estimation of Maximum Oxygen Uptake in Adults from AAHPER Youth Fitness Test Items," Research Quarterly 37:192-201, 1966.

Harris, Margaret L. "A Factor Analytic Study of Flexibility," Research Quarterly, 40:52-70, 1969.

Jackson, Andrew S. and Coleman, A. Eugene, "Validation of Distance Run Tests for Elementary School Children," Presented at Research Section, AAHPER Convention, 1975.

Metz, Kenneth F., and Alexander, John F. "An Investigation of the Relationship between Maximum Aerobic Work Capacity and Physical Fitness in Twelve-toFifteen-Year-Old Boys," Research Quarterly 41:75-81, 1970.

Williams, Robert S. The Relationship Between the Human Motor Performance Domains of Leg Strength and Speed of Body Movement. Unpublished doctoral dissertation, University of Houston, 1974.

Figure 2

TEXAS PHYSICAL FITNESS - MOTOR ABILITY TEST

I. Physical Fitness
   A. Muscular strength and endurance of the arms and shoulder girdle
      1. Chin-ups
      2. Dips
      3. Flexed arm hang (90 seconds)

   B. Muscular strength and endurance of the abdominal region
      1. Timed bent-leg situp (2 minutes)

   C. Cardiorespiratory endurance of distance running
      1. 12-minute run/walk for distance (Grades 7-12)
      2. 9-minute run/walk for distance (Grades 4-6)
      3. 1.5 mile run/walk for time (Grades 7-12)
      4. 1 mile run/walk for time (Grades 4-6)

II. Motor Ability
   A. Running speed
      1. Timed sprint of 50 yards
      2. 8-second dash for distance

   B. Running agility
      1. Shuttle run for distance
      2. Zig-Zag run

   C. Explosive Power
      1. Vertical jump
      2. Standing broad

TABLE 1 Factor Analysis of Running Tests (Disch, _et_._al_. 1975)

| Test | F-1 | F-2 | $h^2$ |
|---|---|---|---|
| 50 Yard Dash | .160 | .817 | .692 |
| 100 Yard Dash | .154 | .828 | .709 |
| .50 Miles | .720 | .546 | .816 |
| .75 Miles | .702 | .607 | .861 |
| 1.00 Miles | .814 | .431 | .848 |
| 1.25 Miles | .860 | .371 | .877 |
| 1.50 Miles | .724 | .287 | .607 |
| 1.75 Miles | .926 | .258 | .925 |
| 2.00 Miles | .900 | .086 | .818 |
| 12-Min Run | -.903 | -.056 | .819 |

TABLE 2 Product-Moment Correlations Between Maximum Oxygen Uptake (ml/kg/min) and AAHPER Test Items (Metz and Alexander 1970)

| Test Item | 12-13 Year-old-Boys N=30 | 14-15 Year-old-Boys N=30 |
|---|---|---|
| Pull-ups | .58** | .52** |
| Sit-ups | .24 | -.03 |
| Shuttle Run | -.52** | -.44* |
| Standing Broad Jump | .49* | .50** |
| 50-Yard Dash | -.65** | -.54** |
| Softball Throw | .42* | .28 |
| 600-Yard Run-Walk | -.66** | -.27 |

*p > .05
**p > .01

TABLE 3 Product-Moment Correlations Between AAHPER Test Items and Maximum Oxygen Uptake (Falls et.al. 1966)

| Test Item | Max $VO_2$ ml/min/kg Body Weight | Max $VO_2$ ml/min/kg Lean Body Weight |
|---|---|---|
| Pull-ups | .48** | .3 |
| Sit-ups | .40** | .32** |
| Standing Broad Jump | .47** | .15 |
| 50-Yard Dash | -.48** | -.24* |
| Shuttle Run | -.61** | -.45** |
| 600-Yard Run-Walk | -.64** | -.48** |

*p > .05
**p > .01