A Model for Objective Sport Skills Tests in Schools: Physical Educator's Concern

給體育工作者使用的運動技能測試模式

Dr. Lobo H. Louie, *Assistant Professor*
Department of Physical Education
Hong Kong Baptist University
雷雄德博士
香港浸會大學體育系助理教授

摘 要

運動技能測試是一種常用於測定學童運動能力的工具。一般的運動技能測試有傳球測驗、對牆擊球測試、射球測驗或擲遠測試等。由於學校體育工作者日常工作繁忙，及並非專項學術研究者，故本文的目的是綜合有關運動技能測試的文獻，配合學校體育的情況，建議出一個給學校體育工作者適用的運動技能測試模式。

ABSTRACT

Objective sport skill tests are common measurement tools for physical educators to evaluate their students' performance in school setting. Traditionally, students are being graded in accordance with their performance in the skill tests, such as throwing objects for distance, wall volley test, or dribbling through a test course. The present paper attempted to collect a variety of information on the objective sport skill testing from experienced scholars. Since most school physical educators are not researchers, it is necessary to provide them with user-friendly techniques. After an extensive review of the related literature in the field of measurement in physical education and exercise science, a model of objective sport skills testing was developed which served as a guideline for the physical educators to follow. It is hoped that such user-friendly procedures would enhance the teaching qualities in the school.

Introduction

Sport skills tests are one of the common measurement tools used by the physical educators, from primary schools to college levels. Physical educators utilize the sport skill tests such as basketball free throw shooting to evaluate students' ability in the sport. They may also use the wall volley test to examine the volleyball playing abilities. However, sport skills are those movement specific to the performance in the particular sport (Hensley & East, 1989). Thus, the sport skills tests must require an environment similar to the game situation and standardize the procedures for test administration (Baumgartner & Jackson, 1995).

The development of psychomotor skills is extremely important during school age period. It is necessary for the physical educators to develop proper measurement tools to allow precise evaluation. The achievement of psychomotor skills can be measured by the following common methods (Baumgartner & Jackson, 1995): (a) skill tests, which are an objective and often-used way of evaluating a variety of psychomotor objectives; (b) rating scales, which are a subjective but systematic method for evaluate those skills that do not lend themselves to objective evaluation; (c) through performance, which can also provide an objective score for skill evaluation.

Functions of Sport Skills Tests

The most common use of sport skills measurement is to determine an individual's progress or level of achievement in a particular sport. Besides, the functions of sport skills tests are as follows: (a) classification - a skill test can be administered early in the instructional process of a sport to classify all participants into various learning groups for effective teaching; (b) diagnosis - determining the strengths and weaknesses of the students can help in the planning of unit objectives and identify students who may need special attention; (c) motivation - a skill test can motivate participants to improve their abilities in a sport when it is used correctly; (d) practice - while practising the test item, the students are actually practising the skills of the sport; (e) programme accountability - test scores and other information can be used to demonstrate to administration, parents and the public, the objectives and values of physical education (Miller, 1988); (f) prediction - when a large data set is obtained, student's ability can be predicted in accordance with the regression technique; (g) comparative evaluation

(Baumgartner & Jackson, 1995); and (h) teaching aid to supplement instruction (Johnson & Nelson, 1986).

Categories of Sport Skills Test

There are four common types of technique to evaluate one's sport skill ability, namely (a) accuracy test; (b) wall volley tests; (c) total bodily movement tests; and (d) throws, kicks, or strokes for power or distance (Baumgartner & Jackson, 1995).

Accuracy tests

Accuracy tests involve throwing, striking, or kicking an object toward a target for accuracy. Common examples are badminton serve tests, basketball free throws tests, or placement tests in tennis and table-tennis. Its purpose is mainly to encourage the examinee to obtain maximal scores by putting the objects into the target. The advantages are rather easy to handle and good for mass testing. However, physical educators should not over-emphasize on the score only since a complete sport skill depends upon many factors such as body movement, eye-hand coordination, and manipulative skills. Many students may practice the skill in a wrong way only because of gaining higher scores. On the other hand, the accuracy test sometimes may not allow sufficient discrimination among skill levels if the scoring system is not properly designed.

Wall Volley tests

Baumgartner and Jackson (1995) stated that wall volley tests require the students to repeatedly stroke, pass, throw, or kick an object at a wall over a specified period of time with the number of successful trials the unit of measurement, or for a specified number of successful trials with time as the unit of measurement. As indicated by Kirkendall, Gruber and Johnson (1987), the wall volley is certainly a good drill for beginners and may have merit in evaluating beginners' skills. But the use of wall volley as a measure of playing abilities may not reflect the player's actual playing skill. For example, volleying a volleyball back and forth above a restraining line on a wall for 30 seconds is not the way a volley is actually played during a game. This test item may have validly in grading wall volley ability, but its use in estimating volleyball playing ability is doubtful. Because a volleyball player should return many different types of shots in the real game situation, an effective volleyball skills test should measure a player's skill in returning these various types of shots. Furthermore, Louie and Lam (1996) have demonstrated that the use of wall volley techniques to evaluate badminton playing abilities is no longer valid. Comparative low validity and reliability coefficients were obtained by the researchers when these wall volley tests correlated with the criterion measures. Students should not be taught and trained too long while practising the badminton skills using the wall volley method. The wall volley techniques should only be introduced as motivation or for a warm-up drill in badminton.

Tests of Total Bodily Movement

These test items require the students to run a standardized test course using movements characteristic of the sport (Baumgartner & Jackson, 1995). Common examples are basketball control dribble test and shuttle run test (AAHPERD, 1984). The tests themselves are similar to the movement demanded by the particular sport, which can serve as skill practice purpose. Since the test courses are analogous to the real play, these test items often receive acceptable validity and reliability coefficients. When students are travelling around the test course, they are indeed practising the skill needed during the game.

Throws, kicks, or strokes for power or distance

It measures the student's ability to throw, kick, or strike an object forcefully. Softball throw for distance and badminton drive test for distance are usual examples of this type of skill test. Its purpose is to evaluate a subject's maximal ability to perform certain skill. These kind of test items are rather common for children, such as softball throwing and beam bag throwing.

Combination tests

Baumgartner and Jackson (1995) further mentioned that these items are a combination of several of the above four groupings, usually speed and accuracy. In basketball, speed spot shooting test which measures the skill in rapidly shooting from specified positions; and passing test which evaluates the skill in passing and recovering the ball while moving, are common cases of this type (AAHPERD, 1984).

Test Validity for Sport Skills Tests

Validity is defined as how well a test measures what it is intended to measure (Kirkendall et al., 1987). A good test should possess acceptable validity both in

norm-referenced or criterion- referenced basis (Safrit, 1990). The most common types of validity are content or logical validity, criterion-related validity, and construct validity. Logical validity is often used when a test obviously involves the skill or ability that is being evaluated (Johnson & Nelson, 1986). Safrit (1990) defined logical validity as the extent to which a test measures the most important components of skill necessary to perform a motor task adequately. There are generally four steps in assessing logical validity of a skill test: (a) review the test developer's statement on the purpose of the test and the components of skill the test is supposed to measure. These test components should be clearly stated in the description of the test and then set up a list of these components; (b) examine the test and list the components actually measured in the test; (c) compare the two lists to see whether they are measuring the same objectives; and (d) examine the educational importance of the test, such as appropriateness, procedures, or the importance of the components. Criterion-related validity is illustrated by comparing test scores with one or more external variables that are considered direct measures of the characteristic (known as the criterion). This procedure for stablishing the validity of a test is considered a traditional method. The rule of thumb is to generate the correlation between the new test and criterion measure. The resulting correlation coefficient will be an estimate of the validity of the particular test. When the correlation is found to be close to one, it indicates that the instrument measure is similar to the criterion measure and is regarded to be valid. However, the physical educators should select the criterion measure carefully because it will directly affect the validity. It is preferable if the tester can collect a variety of criteria and choose the most suitable one.

Construct validity is the degree to which a test measures an attribute or trait that cannot directly measured (Safrit, 1990) or the degree to which a test measures a hypothetical construct, usually established by relating the test results to some behaviour (Thomas & Nelson, 1990). Baumgartner and Jackson (1995) mentioned that construct validity is based on some scientific method. First there is a substantive hypothesis that a test or tests measure some abstract trait. Secondly, a theory is developed to explain both the construct and the tests that measure it. Lastly, various statistical procedures are applied to confirm or reject the hypothesis. Miller (1988) further explained that the construct validity can be estimated by comparing higher skilled individuals with lesser skilled individuals. For examples, when testing both varsity tennis team players and students from tennis skill classes, the varsity players will most likely score better than the students from the skill classes which demonstrates that the tests possess a certain degree of construct validity. This is also called "known group difference construct validity".

The use of multiple regression statistical analysis can be applicable to estimate the construct validity. More commonly, factor analysis is another recent technique for estimating construct validity in measuring psychomotor skills. Examples utilising this methodology includes studies from Hopkins (1976), Hensley (1979), and Louie (1990). With the aid of the user-friendly SPSS windows software, physical educators may try to establish a pool of tests within the particular sport. Then entering the data into the computer and see whether the tests can formulate some kind of 'factor' which can represent the general playing ability of the sport. If so, the physical educators can design the grading method according to the factor structure of the sport. There are two kinds of factor analysis including exploratory and confirmatory factor analysis (see statistics books for details).

Baumgartner and Jackson (1995) outlined the factors that will affect the validity of a test. Firstly, if the selected criterion measure is inappropriate, the validity will be influenced. Secondly, the characteristics of the individuals tested also play a part in determining validity. A test developed for primary school children may not be valid for the college students. Thirdly, the validity coefficient is directly related to the reliability of both the test and the criterion measure. Fourthly, a lack of objectivity, which is the agreement of two or more competent judges or scorers about the value of a measurement, reduces the validity of a test. Fifthly, the more measures the tester obtains for an individual, the more valid an indication of the individual's true ability. The more test trials or lengthening the test will enhance its validity. Lastly, the size of the validity coefficient may also play into consideration. Since many validity coefficients are determined by statistical analysis, it depends heavily on the statistical control of the experiment. Furthermore, it is suggested that physical educators should check the validity when the test items are used for the first time or being modified. If poor administrative techniques or unclear instructions are given to the subjects, or the students being evaluated under different environmental conditions, the validity will also be affected (Miller, 1988).

Test Reliability for Objective Sport Skill Tests

Reliability pertains to the consistency or repeatability of a measure (Thomas & Nelson, 1990). The 'obtained score' of an individual is equal to the sum of the 'true score' plus the 'error score' where the true score represents the level of performance that is truly indicative of the individual's ability, and the error score represents the part of the individual's obtained score that is due to a factor or factors other than the true performance. If a test is perfectly reliable, the obtained score is equal to the true score. When a test is measuring the true scores, a person taking the test more than once will score the same each time. Specifically, the total variance in the obtained scores is found, then the amount of variance due to true performance and the amount due to error performance are established. Reliability is then the percentage of total obtained variance accounted for by the true score variance (Kirkendall et al., 1987).

Reliability theory assumes that any measurement on a continuous scale contains an inherent component of error, the measurement error. Any one or more of the following factors can be a source of measurement error: (a) lack of agreement among scorers (objectivity); (b) lack of consistent performance by the individual tested; (c) failure of an instrument to measure consistently; and (d) failure of the tester to follow standardized testing procedures (Baumgartner & Jackson, 1995). It is rather common to use test-retest technique to obtain the 'stability reliability coefficient' in sport skill tests. Physical educators should always remember that conditions for test 1 and 2 must be the same in order to generate high reliability value of the test item. If the test trials are multiple, physical educators can use the analysis of variance (ANOVA) approach to estimate the reliability coefficient (see Baumgartner & Jackson, 1985 for details). This estimation technique is the current acceptable method of determining the reliability of psychomotor tests with a variety of trials scored over two or more days.

In general, physical educators should comply with the following guidelines, recommended by Baumgartner (1989): (a) the internal consistency reliability coefficients for a group will be higher than its stability coefficient because there are fewer potential sources of change in a person's score from trial to trial than from day to day; (b) the magnitude of the reliability coefficient is affected by the homogeneity or heterogeneity of the data. The larger the range of the test scores, the larger the reliability coefficient; (c) a minimum sample size of 30 to 50 is recommended in estimating the reliability of a particular sport skill test; (d) the test administrator should pay attention to age, gender, and experience level of the subjects, which may all contribute as factors affecting the reliability coefficient; (e) the physical educators should be aware of the students' readiness to participate in the testing, such as psychological concern; and (f) the characteristics of the test administrators and the raters may also be a concern.

Summary

In order to accurately measure the sport skill abilities of the students, physical educators should first gather information about the test item and look for the methodology to obtain acceptable reliability and validity. According to the guidelines presented above, it is hoped that physical educators can develop their own valid and reliable test item for grading and other purposes. Resources include test and measurement textbook, research articles and notes, experienced teachers and coaches in the field, etc. The commonly used statistics for validity and reliability can be found in many measurement books, such as Berg and Latin (1994) and Kirkendall et al.(1987). Additionally, Baumgartner and Jackson (1995) eventually pointed out the following procedures for evaluating skill achievement:

(a) define what is to be measured;
(b) select a measuring instrument;
(c) pretest the instrument;
(d) revise the test and testing procedures;
(e) administer the instrument;
(f) evaluate the administered test including reliability, validity and feasibility;
(g) revise the final test battery;
(h) develop standards.

Based on the review of the related literature, the present author summarized the information and produce a simplified and user-friendly model on objective sport skill testing for the school physical educators to follow, as shown in Figure 1.

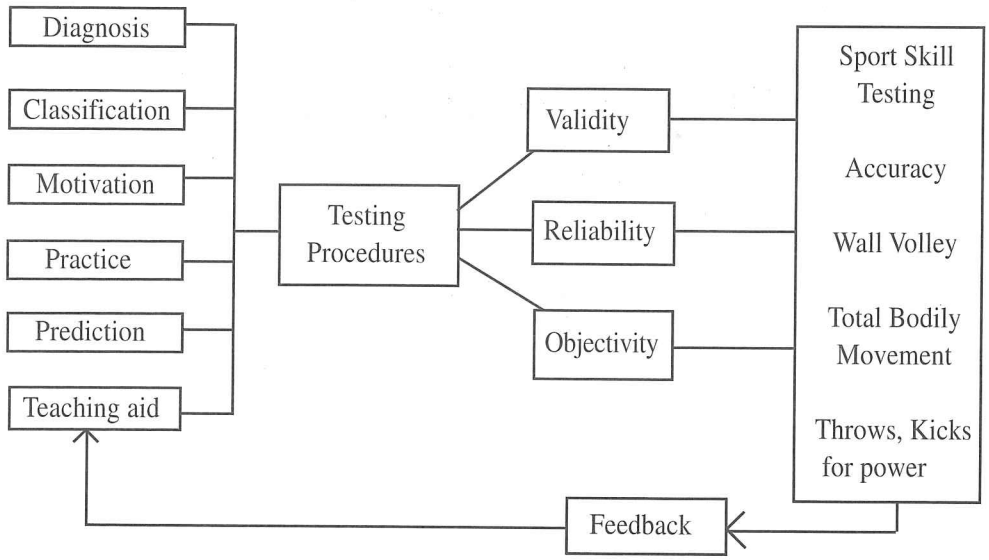

Figure 1. Objective Sport Skill Testing Model for Physical Educators

Reference

American Alliance for Health, Physical Education, Recreation & Dance. (1984). Skills test manual. Reston, VA: Author.

Barrow, H. M., McGee, R., & Tritschler, K. A. (1989). Practical measurement in physical education and sport (4th ed.). Philadephia, PA: Lea & Febiger.

Baumgartner, T. A. (1989). Norm-referenced measurement reliability. In M.J. Safrit & T. M. Woods (Eds.), Measurement concepts in physical education and exercise science (pp.45-70). Champaign, IL: Human Kinetics.

Baumgartner, T. A., & Jackson, A. S. (1995). Measurement for evaluation in physical education and exercise (5th ed.). Bubuque, IA: Wm. C. Brown.

Berg, K. E.,& Latin, R. W. (1994). Essentials of modern research methods in health, physical education, and recreation. Englewood Cliffs, NJ: Prentice Hall.

Hensley, L. D. (1979). A factor analysis of selected tennis skill tests. Unpublished doctoral dissertation, University of Georgia, Athens, GA.

Hensley, L. D., & East, W. B. (1989). Testing and grading in the psychomotor domain. In M. J. Safrit & T. M. Wood (Eds.) Measurement concepts in physical education and exercise science (pp. 297-321). Champaign, IL: Human Kinetics.

Hopkins, D. R. (1976). A factor analysis of selected basketball skill tests. Unpublished doctoral dissertation, Indianna University, Bloomington, IN.

Johnson, B. L., & Nelson, J. K. (1986). Practical measurement for evaluation in physical education (4th ed.). Edina, MN: Burgess.

Kirkendall, D. R., Gruber, J.J., & Johnson, R. E. (1987). Measurement and evaluation for physical educator (2nd ed.). Champaign, IL: Human Kinetics.

Louie, L. H. (1990). A factor analysis of selected badminton skill tests. Unpublished doctoral dissertation, Springfield College, MA.

Louie, L. H., & Lam, B. (1996). The reliability and validity of wall volley badminton skill tests. Applied Research in Athletics and Coaching 1996. Boston, MA: American.

Miller, D. K. (1988). Measurement by the physical educator: why and how. Indianapolis, IN: Benchmark.

Safrit, M. J. (1990). Introduction to measurement in physical education and exercise (2nd ed.). St Louis, MO: Times Mirror & Mosby.

Thomas, J. R., & Nelson, J. K. (1990). Research methods in physical activity (2nd ed.). Champaign, IL: Human Kinetics.