CONSTRUCTION AND VALIDATION

OF A

VOLLEYBALL PROFICIENCY TEST:

COGNITIVE AND PSYCHOMOTOR DOMAINS

By

DONNA ANNE BAYDOCK

B.P.E., The University of Manitoba, 1977

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF PHYSICAL EDUCATION

In

THE FACULTY OF GRADUATE STUDIES

SCHOOL OF PHYSICAL EDUCATION

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1985

90

ABSTRACT

The purpose of this study was to construct and validate an assessment tool that could be used to determine the level of cognitive and psychomotor proficiency possessed at the introductory level of volleyball.

The proposed test was administered to 24 males and 24 females evenly stratified into three skill levels: elite, instructed and novice.

Analysis of variance was used to determine construct validity while the Pearson Product Moment Correlation, kappa coefficient and Generalizability coefficient were all used to determine reliability of various components of the test. Correlation between test components was investigated as was the relationship between achievement of mastery and skill level as demonstrated by the Chi Square statistic.

Data analysis led to the conclusion that all test components were valid and reliable measures of introductory level volleyball skill with some caution being advised in the interpretation of the kappa coefficient. Test components were related but not redundant and nine of the 11 test components showed a significant relationship between achievement of mastery and skill level.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the people who have helped in the realization of this thesis. Special thanks are due to:

Dr. Gary Sinclair, my advisor and mentor, for constant encouragement, enthusiasm and availability.

Dr. Robert Schutz and Mr. Jim Bjerring, members of my committee, for their time and interest in my paper.

Patty Schlafen and Sari Fleming, my two observers and close friends, who gave their time and encouragement freely.

Carl Schwartz, my statistical advisor, who helped me to see the light at the end of the tunnel and took the time to make sure I got there.

Dr. Wendy Dahlgren for her support and advice.

Mrs. Joyce Fromson, my employer and friend, who made sure I had the opportunity to finish this thesis.

My close friends and colleagues who believed in my ability and determination and gently pushed me to achieve my goal.

Sheryle Bergmann, my typist, who worked beyond the call of duty with almost the same fervor I did.

CHAPTER I

INTRODUCTION

> Proficiency testing in physical education, an historical concern, has taken on new significance as educational institutions move to performance-based or competency-based programs. Accountability demanded by colleges, universities and secondary schools has created the need for a reassessment of materials on proficiency testing and a framework for utilization of materials appropriate to our discipline. (McGee and Drews, 1974)

It is often the case that physical education majors have received specialized instruction in several activity areas before they reach university and it is this early exposure that prompts further involvement in the field of physical education. Unfortunately, many universities insist that students participate in activity courses regardless of previously attained competencies and in this manner actually limit a student's formal education. The increasing emphasis on quality of education combined with crowded facilities and limited budgets have prompted the development of proficiency tests (McGee and Drews, 1974). Although most educators agree with the concept of testing for proficiency at the student's request, a drawback has been the construction of valid tools of evaluation.

Whereas cognitive proficiency tests for academic subjects are relatively simple to construct and administer objectively, the same is not true for sport skill tests. Students completing an introductory level physical activity course are expected to demonstrate both cognitive and psychomotor skill in a given sport. Traditionally, evaluations have consisted of a final written exam and subjective ratings by an instructor who has seen the students in class over a period of months. A proficiency test, however, would consist of only one testing session thus a purely subjective rating would lend itself to tester bias and be a very unreliable method of evaluation. The sport of volleyball is offered extensively as a credit course in Physical Education degree programs throughout North America. Consequently, it is appropriate to select volleyball for the development of a valid and reliable proficiency test.

STATEMENT OF THE PROBLEM

The purpose of the study is to construct and validate a measurement tool that can be administered to assess the level of cognitive and psychomotor proficiency possessed at the introductory level of volleyball.

RATIONALE FOR THE STUDY

To be considered proficient at a sport or physical activity one must demonstrate competence in both the cognitive and motor domains of behavior as related to a particular activity. The cognitive domain assessment should cover such sub-domains as skill techniques, strategy, principles of movement and to a lesser degree, rules, equipment and safety. To effectively test the motor domain it is necessary to use both objective skill tests to evaluate the product of performance and subjective performance ratings to evaluate the process of performance.

Initial attempts at evaluating volleyball playing ability included repeated wall volley tests and service accuracy tests. Generally, reliability coefficients were high but validity coefficients were questionable. In most cases the criterion used for validity was a subjective rating of performance in a game situation. Logically, it seems difficult to infer volleyball playing ability from the limited information available in an individual skill test so it is not suprising that validity was marginal.

There have been more recent attempts at making individual skill tests more game-like (Chun, 1969) and at combining a number of skill tests (AAHPER, 1967; Williams and Fawcett, 1975) in order to better predict volleyball playing ability. Although some advancements have been made in the construction of skill performance tests there has been no attempt to combine this information with volleyball knowledge tests.

The AAHPER volleyball skill test (1967) is one of the most inclusive and considers four measures: volleying, serving, passing and set-ups. However, there are still severe limitations to this instrument as a test of proficiency for volleyball playing ability;

1) the volleying and set-up test both measure the skill of overhead passing. The set-up test is more relevant because of its game-like situation thus rendering the wall volleying test redundant.
2) the skill of spiking is missing from the evaluation and it is

an essential component of the game of modern volleyball.

3) there is no evaluation of technique accompanying the accuracy tests so players may adopt any movement that achieves the goal without penalty for poor technique.

4) there is no cognitive component to the test to measure knowledge and performance analysis ability.

Difficulties in objectively and validly testing sport skills have left a void in the literature with respect to proficiency testing in physical education and especially volleyball. The lack of a relevant test to measure volleyball proficiency has prompted the development of the proposed test to selectively assess both cognitive and psychomotor skill at the introductory proficiency level.

DELIMITATION

The proficiency test is designed to discriminate between non-instructed players and those instructed at the introductory level. Since only introductory level skills are being evaluated there may be a ceiling effect that does not allow for differentiation between varying levels of elite players, i.e., varsity players and national team members.

ASSUMPTIONS

1) An introductory level volleyball course is concerned with teaching proper technique and comprehension of skills along with the strategies and rules of the sport --- teaching methodology may be incidentally learned but not evaluated.

2) As a preparation for coaching and/or teaching, the emphasis in an introductory course is equally weighted between the process and the product of skill acquisition and will be evaluated accordingly.

3) The personnel utilizing this assessment tool will have a thorough knowledge of modern volleyball. It is expected that the evaluators would possess, at least, Level I National Coaching Certification Program Volleyball Conductor certification or its equivalent.

CHAPTER II

REVIEW OF LITERATURE

The need for proficiency testing in physical education has been identified. The extensive popularity of volleyball makes the development of a volleyball proficiency test a practical and highly useful contribution to the field. In order to construct an effective measure it is necessary to review the theory behind proficiency testing, examine current test construction procedures and review existing tests of volleyball skill and knowledge.

The theoretical basis for the construction of proficiency tests stems from the philosophy of mastery learning. It simply states that most students can and will learn what they are taught if appropriate instructional methods are utilized, and the appropriate time is allowed. This philosophy has been the premise behind tutoring for a few thousand years but group-based mastery learning strategies are relatively new to the field of education, being introduced in the late 1960's (Bloom, 1968). One of the major reasons for the lack of use of mastery learning strategies was the adoption of the statistically valid normal curve as a seemingly necessary tool in grading student performance by assigning values over a range from A to F (90%-40%). Administrators are often critical of teachers for being either too lenient or too demanding if students' marks do not span the range of the normal curve. Unfortunately, this attitude in the school system is often counterproductive to educational goals. Teachers may begin to teach with the expectation that only very few students will master the material while students will come to believe that they are only capable of achieving a certain level of mastery, e.g., 60%-70%, and will not be motivated to work any harder.

In 1963, Carroll introduced the idea that a student's aptitude for a particular subject did not necessarily predict the level of achievement in that subject, but rather influenced the rate of learning. A student with a high aptitude for a subject would learn it quickly while a student with a low aptitude would learn it more slowly. The degree of learning would depend on the time the student spent on learning relative to the time required. Carroll identified a student's perseverance in studying and his actual opportunity to learn (class time) as key factors in the time spent on learning. On the other hand, the time needed was determined by a student's aptitude, the quality of instruction and his ability to understand the instruction.

It follows that if aptitude corresponds to the rate of learning rather than the actual level of achievement, it should be possible to set performance levels that all students can master at their own speed. Bloom (1968) exemplified this logic by stating that if students were normally distributed on aptitude for some subject and they were given equal opportunity to learn and equal quality of instruction then achievement levels would be highly correlated to aptitude and show normal distribution. However, if differential opportunity to learn and differential quality of instruction were available for those who needed it most, the majority of students could be expected to attain mastery providing, of course, that the criterion for mastery was appropriately set.

Mastery learning strategies have been tried in many parts of the world for a wide variety of subject areas across all levels of education (Block, 1979). They have been used in classrooms with a student-teacher ratio of 20 to 1, 30 to 1 and even 70 to 1 (Kim, 1971). In order to evaluate the effectiveness of these strategies student learning must be examined and this is most commonly done by measuring achievement. Often a mastery standard of 80% correct is set on a final examination and performances are compared between mastery and non-mastery students who studied the same subject. Available research generally indicates that two to three times as many mastery learning students have achieved the standard as have students learning by the usual lecture-recitation approach (Block, 1974). Kim (1971) used thousands of seventh grade Korean students to study the possible impact of mastery learning across a variety of subject areas. He found that 75% of the mastery learning students compared to only 40% of the non-mastery students achieved the 80% correct critierion on the final exam. There is also a great reduction in the number of students receiving marks of C, D, and F. According to these findings the cognitive aspects of student learning are positively influenced by master learning techniques.

Evidence has also shown positive affective outcomes for mastery learning students. It seems that students show more interest and more positive attitudes toward the subject matter being learned. They also demonstrate an increased confidence in their ability to learn (Block and Anderson, 1975). The student's performance is compared to a predetermined standard or criterion and the student can clearly see if mastery of the criterion has been attained. This method of interpreting test results is called criterion-referencing and it differs from the commonly used standardized achievement tests which report test performance in terms of an individual's relative position in the class or in a sample population. This type of standardized test is called a norm-referenced test and in order to reliably differentiate between students' performances, a good spread of scores is essential so that statistical measures can be computed. In mastery learning no comparisons are made with the rest of the class and since there are no limitations as to how many students can achieve mastery, there seems to be a more cooperative atmosphere among students. As Gronlund (1973) points out, a normal distribution of scores is neither expected nor desired. If the test items adequately evaluate the initial objectives and specific learning outcomes and all of the students know their material, then all of them can and will achieve mastery. This probably indicates a teaching job well done rather than a test which is too easy. The result is positive reinforcement for the learners which is a strong psychological motivator for continued effort.

In order to test for mastery it is essential to use criterion-referenced tests. The formulation of criterion-referenced tests should be directed toward obtaining measures of achievement that can be witnessed in terms of student performance on clearly defined educational tasks. Attainment of this goal requires a specific and delimited domain of learning tasks that are presented as instructional objectives and can clearly be defined in behavioral terms and listed as learning outcomes.

Gronlund (1973) suggests two different levels of learning and discusses the use of criterion-referenced tests with each level of learning. Most subject areas can be clearly defined and stated as behavioral objectives when basic skills are being taught. At the introductory level it is very realistic and necessary to set mastery as the performance standard so that this knowledge can act as the basis for further learning in the field. Gronlund calls this learning of minimal essentials the mastery level of learning and explains that criterion-referenced tests are easiest to design, construct, and interpret at this level. Once students have mastered the minimum essentials in a field of study they enter a developmental level of learning where each student is encouraged to strive for the maximum level of achievement and excellence of which they are capable rather than the mastery of some pre-determined criterion. Obviously the use of criterion-referenced tests is limited. The learning outcomes are complex, the domain of learning tasks is virtually unlimited, and learning is seldom sequential so instructional objectives are used more as goals to work toward rather than goals to be mastered. Norm-referenced tests must be used to evaluate students' progress at this level.

From the preceding discussion it can be seen that criterion-referenced tests are best utilized in mastery learning situations where instructional objectives and learning outcomes can be very clearly defined. Once these learning outcomes are stated, an appropriate standard of student performance must be established. This is where a criterion-referenced test can be rendered either effective or ineffective.

Shepard (1980) examined the controversy existing in the standard-setting literature and presented a number of alternatives. She delineated the uses of criterion-referenced tests and suggested various standard-setting methods for each. The proposed volleyball proficiency test resembles her description of "pupil certification". Shepard states that when constructing a criterion-referenced test for pupil certification it is important to consider both absolute judgements about performance and passing rates of previous students. Absolute judgements are based on experts' opinions of a minimally qualified individual. Following the Angoff (1971) method the judges review all the test items and assign a probability or subjective estimate of how likely it is that a just-barely-qualified person will answer correctly. The mastery or cut-off score is set as the sum of the probabilities for all the items in the test. Of course, this standard is based on subjective ratings so as Shepard (1980) points out it is critcial to refer to previous passing rates to assure the mastery level is not artificially too high or too low.

When test items are being selected for criterion-reference tests, it is important that educators consider the stages of learning and the appropriate prerequisite abilities. To identify prerequisite abilities it is necessary to have a method of classifying behavior that enables behavioral skills to be placed in some order, preferably hierarchically from lowest to highest or simplest to most complex. Since the goals of education are focused upon the growth and development of the total child, educators must be concerned with all three domains of behavior: cognitive, affective and psychomotor. As Harrow (1972) points out, it is difficult to identify behaviors that belong exclusively to one domain but in order to set meaningful learning outcomes the primary purpose for studying a behavior must be identified and classified into one of these domains. Each of the domains has been organized into a hierarchical classification scheme of educational objectives called a taxonomy.

Taxonomies for the cognitive (Bloom et al, 1956; Gagne, 1965) and affective (Krathwohl et al, 1964) domains were established earlier, and provided a common foundation upon which teachers and curriculum developers could organize learning experiences for children. Taxonomies have also provided for clarification of terminology in a field, systematic development of a learning theory and the exchange of evaluative tools and procedures among teachers and researchers.

The trend toward movement efficiency as an essential factor for optimum development in all learning domains sparked the construction of a taxonomy for the psychomotor domain. Since a taxonomy is hierarchically constructed, educators benefit by becoming aware of prerequisites that are necessary for the development of various movement tasks. Teachers can also insure that they set behavioral objectives at all relevant levels of the taxonomy rather than predominantly at the lower levels. This was a common problem encountered when school curricula were examined in lieu of the cognitive taxonomy of educational objectives.

Many of the initial attempts of classification systems in the psychomotor area were concerned with categorization of behavior according to task variables (Fitts, 1962, 1964). Fleishman (1964) even went so far as to develop an extensive factor analysis to identify eleven ability and nine proficiency factors that were independent of each other but common to a variety of psychomotor skills. These experimenters were extremely concerned wih categorizing psychomotor tasks but paid little attention to the learner and the learning processes necessary to achieve the different categories of behavior. In 1970, Gagne introduced a hierarchical system of eight levels of learned behavior based on "the conditions necessary for observing and promoting each category". Two of his categories were psychomotorically oriented and he considered them to be prerequisities to the cognitive behavior levels further up the hierarchy. He called the psychomoter categories stimulus response learning, which required a specific motor response, and chaining which started with a single stimulus cue that triggered a series of motor responses. Merrill (1971 a,b) added a third category called complex skill which required the execution of a number of different chains that are each triggered by separate cues presented in varying orders. Unfortunately, neither experimenter considered environmental stimuli and the important role they play in response selection. An important discovery that the Gagne/Merrill taxonomy did assume was that the conditions and processes for learning a new single response act are very similar regardless of whether the response is the manipulation of two fingers or a gross body movement. Initially this idea was very speculative, but as more researchers became involved with developing a taxonomy for the psychomotor domain they all adopted this approach. They became increasingly interested in the similarities of the learning process across the domain of psychomotor skills rather than the diversities that exist between particular tasks or behaviors.

Simpson (1966) made one of the first attempts at devising a taxonomy specifically relating to the psychomotor domain. She hierarchically organized learning sequences according to response complexity. The initial level was perception dealing with sensory stimulation, the selection of cues and translation of this information. The second level considered the readiness of the learner according to mental, emotional, and physical set. Level three was called guided response and dealt with imitations and trial and error learning. Habituation of movement was the concern of level four which was titled mechanism. The next three levels were complex overt response, adaptation and origination of movement. Simpson's model provides a good descriptive hierarchy of the stages a learner passes through enroute to mastery of a skill. However, as Harrow (1972) explains it has limited use as a guideline for writing behavioral objectives. The first two levels are unobservable and levels three and four are inherent in skill learning but do not provide a good point at which to evaluate students because they have not yet learned the skill. The final three levels are observable but are concerned with creativity which is difficult to measure objectively. Harrow herself presented a very intricate psychomotor taxonomy that classified only observable movement behavior. The main categories were reflex movements, basic fundamental movements, perceptual abilities, physical abilities, skilled movements and non-discursive communication with a large number of sub-categories under each classification. Although these observable behaviors were sequentially ordered, mastery of one level was not necessarily a prerequisite for the evidence of behaviors at a higher level. For example, it is quite feasible that perceptual and physical abilities are developing at the same time, without one being a prerequisite to the other. Also, many of the behavior sub-categories were either innate or maturationally developed rather than learned so educational objectives would have limited use.

A more recent taxonomy for the motor domain was developed by Jewett et al. (1971). It more closely parallels the cognitive and affective taxonomies because it deals with the process of learning rather than the product which was emphasized in the preceding two models by Simpson and Harrow. In a monograph (1977) Jewett and Mullan elaborate the Purpose Process Curriculum Framework (PPCF) which was developed as a culmination of the efforts of many physical education professionals. The two major dimensions were purpose of human movement (why we move) and process of human movement (how to move). Purposes of movement in achieving the goals of man have been organized into three specific categories: individual development, environmental coping and social interaction. The second dimension of the PPCF is the process of movement. Here the concern was on understanding the learning process and differentiating between learning operations required for various types of movement.

The taxonomy began with generic movements which included perceiving and patterning. These were considered movement processes which facilitate the development of human movement patterns. The next stage was ordinative movement which includes adapting and refining motor skill according to specific task demands. The highest level of learning and performance was designated as creative movement. Here the ability to vary, improvise and compose skill became evident. These hierarchical stages of motor skill learning can be referred to when instructional objectives for a certain skill are desired. It would be difficult to outline specific objectives at the higher levels of skill, i.e., creative movement, but it should be possible to clearly define objectives for sport skills at the psychomotor levels of perceiving, patterning, adapting and refining.

The theoretical background for proficiency testing has been examined and current test construction procedures have been considered. With this information it is important to review existing volleyball tests.

Initial volleyball skill tests were developed in the 1930's and 40's and they professed to evaluate volleyball playing ability. The most commonly used skill test for volleyball playing ability has been the repeated wall volley test with a wide range of variations in specifications of test procedure (Bassett, Glassow and Locke, 1937; Crogen, 1943; Brady, 1945; West, 1957; Clifton, 1963). One of the earliest tests proposed by French and Cooper in 1937 required subjects to stand three feet away from a wall and count the number of times they could volley a ball to a target area above 7.5 feet on the wall within a 15 second time limit. Mohr and Haverstick (1955) found that the validity of the test increased as they moved the subjects away from the wall from three to seven feet. A greater degree of skill was now required to control the ball so the test was found to be more discriminating, however content validity to game playing ability is disputable. According to Jewett and Mullan's psychomotor taxonomy, repeated wall volley tests would be at the skill level of patterning while playing in a game situation would be considered a much more difficult skill, probably at the level of varying or improvising. Although highly skilled performers in a game situation have mastered prerequisite patterning movements and would score well on the wall volley, the converse is not true. Players scoring well on the wall volley would not necessarily score well in a game situation because they may be in a situation well above their skill level.

Johnson (1967) criticized the repeated wall volley tests saying that players were never required to judge a ball rebounding off a wall in a game situation. She also said that it was a difficult test to administer to an entire class because of limited wall space. She devised a high volley-to-self test. Players were required to volley to themselves for 30 seconds ensuring that the ball cleared a 10 foot rope each time. They were allowed a 10 foot by 15 foot area to move around in. A validity coefficient of .74 was calculated when test results were correlated to judges' rankings of game playing ability. The test may have been simpler to administer but content validity was still questionable. Volleying to oneself in a 10 foot by 15 foot area would again be considered a patterning movement at the lesser-skilled end of Jewett and Mullan's taxonomy.

Chun (1969) devised a reliable and valid alternative to test the overhead volley-pass with the use of a ball machine to release balls consistently and a target area on the court much like that a player would actually aim for in a game situation. Chun was careful not to generalize her results to volleyball playing ability. A validity of .81 was calculated by comparing test results to judges ratings of the subjects' ability to volley in a game situation. The Chun test score was also correlated with the Clifton wall volley test and an intertest correlation of .77 was achieved. Because Chun's test required subjects to judge an oncoming ball and move into position to volley it, the subjects were required to function at the psychomotor level of adapting or refining which is a more appropriate goal for highschool and college level players. Chun criticizes the use of testers to toss balls because the human factor negates objectivity and lowers reliability. This may be true to an extent but the ability to anticipate a toss by watching a tosser's preparation and release of the ball gives the subject a substantial amount of information in preparation for the motoric response. The release of a ball from a ball machine can be very deceptive if a subject is unfamiliar with this apparatus. There is also a problem with the lack of availability of volleyball tossing machines so the slight loss of objectivity is probably overshadowed by the practicality of using a skilled tosser.

It should also be noted that Chun (1969) analyzed her test results including all overhead passes that successfully reached the target and then repeated the analysis disallowing any passes that were not legally contacted. The reliability and validity coefficients were significantly better when only legal contacts were considered. This is an indication that technique or the process of movement should be considered along with the product of movement (objective score).

Another commonly used skill test for volleyball is serving (Lopez, 1957; Brumbach, 1969). It is not clear whether the underhand or overhand serve was evaluated. Russell and Lange (1940) combined the wall volley and serving test but both measurements only required skill at the patterning level so again volleyball playing ability was not logically evaluated.

In 1974, Sandra Fawcett reviewed the validity of existing volleyball tests. She was concerned with how well volleyball skill tests related to game performance. Fifteen female university students of varying volleyball skill background were assessed on five volleyball skill tests; the Brady Volleyball Test, the Cunningham and Garrison High Wall Volley Test, the French and Cooper Service Test, the Singer Dig Test and the Singer Spike Test. Subject numbers were low but Fawcett concluded that even the better tests (dig and high wall volley) were only moderately related to volleyball ability in a game situation. She did however postulate a subjective equation which used weighted values of the test scores according to the difficulty of the various skills and their occurrence in a game. In a subsequent study, Williams and Fawcett (1975) devised a stepwise multiple regression analysis to predict overall volleyball playing ability. They found that the high wall volley test in combination with the dig test accounted for 74% of the total variance in volleyball ability with a multiple R of .863. Caution is encouraged when interpreting these results because of the low number of subjects and the very low number of trials per test. The serve and spike tests were never considered in the multiple regression analysis because validity coefficients were too low. The concept of devising a multiple regression analysis is a good one but not at the expense of constructing valid and reliable skill tests. More research is necessary for a worthwhile equation to be constructed.

A review of previously constructed volleyball tests shows a trend from individual skill tests in an artificial environment (i.e., repeated wall volley tests) to individual skill tests in a more game-like situation requiring movement and judgement skills (Johnson, 1967; Chun, 1969) to combined skill tests (AAHPER, 1967; Williams and Fawcett, 1975). To date, very little emphasis has been given to the process of skill performance and therefore subjects are not penalized for poor technique. At the introductory level of skill acquisition proper technique is as essential as the product of movement. Both elements of performance will be evaluated in the volleyball proficiency test.

CHAPTER III

PROCEDURE

Overview

The proposed volleyball proficiency test was administered to a stratified random sample of 48 individuals who were either students or Alumni of the University of Manitoba or the University of British Columbia. Testing occurred in the spring of 1985 at both institutions. An equal number of males and females (eight) were selected for each of the three levels of skill; novice or non-instructed, instructed at the introductory level and elite or varsity players. The proficiency test evaluated four components of volleyball skill and knowledge at the introductory level:

1. cognitive aspects about skills, strategy and rules
2. performance analysis of individual and team skills
3. objective evaluation of the overhead pass, forearm pass, overhand serve and spike (product score)
4. subjective evaluation of the technique involved in the overhand pass, forearm pass, overhand serve and spike (process score).

The content validity or domain-referenced validity (Safrit, 1977) of the cognitive test was established by checking that test questions matched the table of specifications designed to describe the domain of introductory level volleyball knowledge. The video tape skill analysis was designed to assess the ability to diagnose and correct common errors in the basic skills of volleyball. Content validity was claimed due to the agreement of a panel of experts that important skill errors were represented. Logical validity was the basis for the construction of the individual skill tests. Safrit (1981) explained that a high score on a skill test should closely approximate the definition of good performance of that skill. Experts in the field were consulted to ensure the tests did parallel good performance of each of the skills.

Source of Data

The stratifications of each skill level were defined and then eight male and eight female subjects were selected in each classification. The non-instructed group consisted of university or college students who had never taken an instructional volleyball course or participated on an elite team. Participation on a highschool team was acceptable. Volunteers were solicited from a first year Physical Education course at the University of Manitoba. The remaining eight non-instructed subjects were randomly selected from intramural players either at the University of Manitoba or the University of British Columbia.

The instructed subjects were eight males and eight females who were randomly selected from a possible 47 students in the six week introductory level volleyball course at the University of Manitoba. The proficiency test was included as part of their course evaluation.

The elite players were required to have played at least one year of inter-collegiate or elite provincial team volleyball. Six of the women were first or second year varsity athletes at the University of Manitoba. The remaining two women and all eight men were Alumni of Canadian Inter-collegiate volleyball teams.

Test Construction

This section of the procedure will deal with the method of construction utilized for each test. Data analysis will follow in a later section.

1) Cognitive Test

The cognitive test was designed as a criterion-referenced test. Initially, experts in the field were consulted to help define the domain of introductory level volleyball knowledge. A pilot study was conducted by testing 59 instructed subjects on 35 multiple choice questions. An item analysis was conducted on these initial test results. Generally when multiple choice items are constucted for a norm-referenced test, a biserial correlation of .30 or above is desired as the index of discrimination (Safrit, 1981). In general, this indicates that students who score well on a question also score well on the total test. However, this stringent biserial coefficient is not realistic for a critierion-referenced test where the majority of instructed students are expected to pass the test items (Brown, 1981). Because discrimination and item difficulty were not the goals of the proficiency test, rather than using a cut-off point for the selection of items, the information from the item analysis was used to improve ambiguous questions or replace distractors that were never used or used too often.

Seven of the original questions were deleted, eight were kept and twenty were revised. The 20 revisions plus 12 new questions were tested on 73 non-instructed students from a first year Physical Education class at the University of British Columbia. Again an item analysis was conducted and the questions were scrutinized.

In order to select the most relevant questions a content balance table was constructed. (See Table I.)

Table I

Content Balance Table

	Cognitive Levels			
Content Areas	Knowledge	Comprehension	Application	Total
Skill Techniques	9, 23, 24, 26	1, 4, 6, 18, 19, 22, 25, 30	2, 13, 14, 29	16
Strategy and Tactics	5	3, 7, 16, 17	8, 15, 27, 31	9
Procedures and Conduct	37, 38	11, 12		4
Rules	10, 34	32, 35, 36, 39	40	7
Terminology		21, 28		2
History	20			1
Equipment	33			1
	11	20	9	40

The domain of introductory volleyball knowledge was divided into its content areas and 40 of the previously tested questions were selected according to Bloom's first three levels on the cognitive taxonomy; knowledge, comprehension and application. A copy of the questions can be found in Appendix A.

2) Performance Analysis

The investigator postulated that it was important for teachers and coaches to analyze skills and be able to detect relevant errors in skill execution. In order to evaluate this ability, an appropriate video tape was constructed for analysis purposes. Representative episodes were selected from two introductory volleyball classes performing the basic skills of overhead passing, forearm passing, overhand serving and spiking. These students were performing the skill tests as part of their evaluation for the course thus all errors portrayed were authentic.

The investigator, with the aid of two experts, chose 15 episodes of individual errors and five episodes of team errors to be representative of the performance analysis expected at this level. Pilot work with video analysis had revealed that open-ended questions solicited a very wide range of responses that made objective evaluation difficult. Consequently, a multiple choice format was adopted for the refined video analysis portion of the test. The testing procedure introduced each episode by first presenting the question on the video screen thus focusing the subjects on the pertinent aspects of the performance to follow. All questions were phrased as if the player on the video tape was asking the viewer for assistance, i.e., "Why can't I get the ball to go farther forward?". Three repetitions of the error were presented followed by the multiple choice responses to the previously presented question. The three repetitions of the error were repeated with a 15 second response time allotted. After 10 seconds, a tone sounded to tell the subjects to look up and prepare to attend to the next episode.

The tape consisted of three overhead pass errors, five forearm pass errors, two overhand serving errors and five spiking errors. A greater number of forearm passing and spiking errors were selected as the pilot test revealed those to be most problematic to the introductory students. A copy of the questions can be located in Appendix B.

3) Objective Performance Evaluation - Product Score

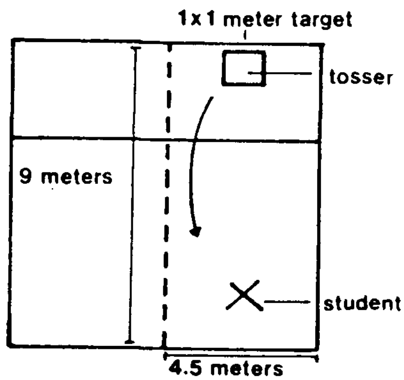
The four skill tests were constructed to require closest approximation of a good performance. After consulting the available psychomotor taxonomies it was decided that a player proficient at the introductory level of volleyball should be able to perform the four basic skills at Jewett's fourth level of motor performance (Jewett and Mullan, 1977). This level is called the refining stage and it is directed toward the acquisition of smooth and efficient control of established motor patterns, toward the acheivement of precision and toward the habituation of performance.

Although a number of different testing dates were used for different subjects, test administration was standardized. Subjects were allowed 10 to 15 minutes to properly stretch and warm up. A practice net was always available so subjects could keep warm while waiting their turn. An assistant administrator was given instruction in tossing the ball properly with a two-handed underhand motion and sufficient arc so it was possible for subjects to move into the expected position. The assistant practiced while the subjects warmed up. If at any time the assistant felt that a toss was erratic, another trial was given. All trials of all four skills were video taped so the two judges doing the ratings did not always have to be present. The video taping of all skills was done from a lateral view so forward-backward movement could be highlighted.

Before being tested on each skill, subjects received one practice trial to prepare them for what the toss would look like. Then the 10 trials were given. The skill test was always administered in the same order. Each subject performed the overhead pass followed directly by the forearm pass since the tests were so similar. Male and female subjects were separated for the overhand serving and the spiking test so the net height could be adjusted properly. International volleyball net heights were used; 2.24 metres for the women and 2.43 for the men.

a) <u>Overhead passing</u>. On one side of the net the volleyball court was divided in half from the net to the back line, i.e., nine metres long by 4.5 metres wide. A one metre by one metre box was taped to the floor as a target area. This box was 10 centimetres from the centre court line and 1.75 metres from either of the bordering sidelines. The tosser was located in this target area. The subjects were instructed to start on an X taped to the floor two metres from the back line and 2.25 metres from either sideline. Between each trial the subjects were told to return to the X. Refer to Figure 1.

Figure 1



The subjects were told that the assistant would be tossing balls for them to receive in various locations within the designated boundaries. They were responsible for moving into position and legally overhead passing the ball back to the tosser. The trial was considered successful and scored a point if the tosser could catch the ball above waist level with two hands while standing on both feet within the one by one metre target area. If the tosser had to jump to catch a ball that was going over the net, it scored zero. At some point in it's flight path the ball had to reach a height above net level. This was judged by the test administrator who was doing the video taping and had a side view of the subject, the tosser and the flight path of the ball. If the pass was not above net level, the administrator informed the tosser and the trial scored zero. At the end of the 10 trials the subject received a product score out of 10 which represented all successful trials.

b) Forearm passing. The test was carried out in the same format as the overhead passing test with all the same court markings and rules for successful passes. The only change in instructions was that the ball had to be legally forearm passed. Again the subject received a product score out of 10.

c) Overhand serving. On one side of the net the volleyball court was divided in half from the net to the back line, i.e., 4.5 metres wide by nine metres long. The subjects positioned themselves in the serving area on the opposite side of the court. Each subject was responsible for serving 10 balls in a row. The first five were to be served diagonally to land anywhere in the cross-court area. The last five were to be served straight ahead in a down-the-line position. Serves landing in the proper court or on the boundary lines were considered good and scored a point. Serves that cleared the net but landed in the wrong half of the court scored zero.

Previous serving tests allotted higher points for accuracy in different court areas. The subject had to be so concerned with accuracy that technique probably suffered. Since no subjective evaluation was undertaken, a player could use a very simple underhand serve and receive a high score. In modern volleyball the serve is used as a weapon and even more important than pinpoint accuracy is the velocity and flat trajectory of the ball. As with the other skill tests, the overhand serve test was designed to correspond to the criterion of good performance at the refining level of performance.

d) Spiking. The tosser was positioned near the net at approximately center court and used a two-handed underhand toss to simulate a set about two to three metres above the height of the net and about .5 to 1.5 metres away from the net. This tossing skill was more difficult than the one for the overhead or forearm pass so again if the tosser felt a set was erratic the subject was given another trial. Balls were tossed to the subjects on their power or on-hand side, i.e., right-handed hitters attacked the ball from the left front position and left-handed hitters attacked the ball from the right front position. This skill was not as advanced as hitting a ball that crossed the spikers body before being contacted, i.e., off-hand.

The subject was responsible for approaching and spiking 10 balls in a row over the net and into the court. No target areas were designated in the court because it was felt that the skill of properly timing a volleyball spike to accurately place it in the court 10 times was itself at the level of refining according to Jewett and Mullan's taxonomy. If the ball 'knicked' the net on it's way over but the flight path was not really altered it was considered a point. However, if the ball was hit into the tape at the top of the net and happened to roll over, it was not considered a point. If the subject committed a net fault or a center line fault upon landing, the trial also scored a zero. The subjects were given sufficient time between trials to back-off the net and prepare for the next spike so fatigue did not become a factor in their performance.

4) Subjective Performance Evaluation - Process Score

One of the goals of an introductory level volleyball course is to teach proper technique of the basic skills. Therefore, the process of performance was deemed as important as the product of performance in the proficiency test and both components were evaluated. Subjective rating scales have been devised for a number of sport skills. Suttinger (1957) presented a four point rating scale for volleyball playing ability in general (the Suttinger Volleyball Rating Scale). This was not specific enough for the present study so the investigator with the help of other volleyball experts, developed a four point rating scale for each of the four skills tested. Detailed descriptions of the overhead pass, forearm pass, serve and spike were devised at each of the four levels of performance. The judges were presented with these descriptions prior to rating the subjects and were given time to scrutinize the information and ask questions of the investigator. Both judges had five years coaching experience as university coaches or elite provincial team coaches in the province of British Columbia. A copy of the skill descriptions and the judges tally sheets can be found in Appendix C and D.

As previously stated, all trials of all subjects were video taped. The tapes were made available to the two judges so they could review them at their own speed and go back and repeat episodes if they felt it was necessary. This reduced the chance of external distractions that might have affected the accuracy of the ratings. The judges were required to give each skill trial a rating of one to four. The subject's process score on each skill took into account both judges' scores of the 10 trials, i.e., total score judge one plus total score judge two divided by 20 equaled total score out of four. Equalization of the process score with the product score (possible out of 10) was accomplished by multiplying the process score by 2.5.

Data Analysis

Construct validity of the proficiency test as a test of introductory skill and knowledge was investigated by a series of two by three factorial ANOVA's for randomized groups. Each of the four components of the test were analyzed separately and then as a total score. In addition, correlation coefficients were obtained for the relationships between the three different skill levels and each of the four test components to see if the same subjects did well in all aspects of the test. A Chi square analysis was conducted on each of the test components and on the total score to determine if the number of individuals achieving mastery differed among groups.

Since the cognitive knowledge portion of the test was constructed as a criterion-referenced test it was not appropriate to test for reliability with the Pearson Product-Moment technique which assumed a normal distribution of scores. Instead a proportion of agreement statistic was utilized to test reliability of the multiple choice questions. Problems with test-retest methods of reliability on multiple choice tests led the researcher to separate odd and even test items and analyze the scores according to the method proposed by Swaminathan, Hambleton and Algina (1974); the kappa coefficient. Reliability of the video tape analysis was determined by the Pearson Product-Moment correlation on a co-ed group of 12 instructed subjects who took the test twice with five days between testing sessions. The reliability and objectivity of the four individual skill tests was established by using the Generalizability theory and calculating G coefficients.

CHAPTER IV

RESULTS AND DISCUSSION

The analysis of data will be presented in the following manner:

1) Construct validity - ANOVA on each test component; a) cognitive test, b) performance analysis, c) product score of the overhead pass, forearm pass, overhand serve and spike, d) process score of the overhead pass, forearm pass, overhand serve and spike and e) a total test score.

2) Reliability of the cognitive test - proportion of agreement and kappa coefficient on odd/even trials.

3) Reliability of the performance analysis - Pearson Product Moment of Correlation on test-retest results of 12 instructed subjects.

4) Reliability and objectivity of the skill tests - product and process - generalizability coefficients for inter-rater reliability, inter-trial reliability and performer reliability.

5) Correlation between test components - Pearson Product Moment Correlation between the 10 test components.

6) Comparison of number of subjects achieving mastery in each skill level - Chi Square statistic.

Construct Validity

The initial concern of the volleyball proficiency test was to investigate construct validity since this had been a common problem with previously constructed volleyball tests. One method of establishing construct validity is to test for theoretical group differences. In order to accomplish this, analysis of variance was used to determine if significant differences existed between the scores of the three skill levels; novice, instructed and elite. For each component of the proficiency test the results will be presented graphically with a table of significant values and ensuing interpretations.

Cognitive Test

Figure 2

Graphic Representation of the Results of the Cognitive Test

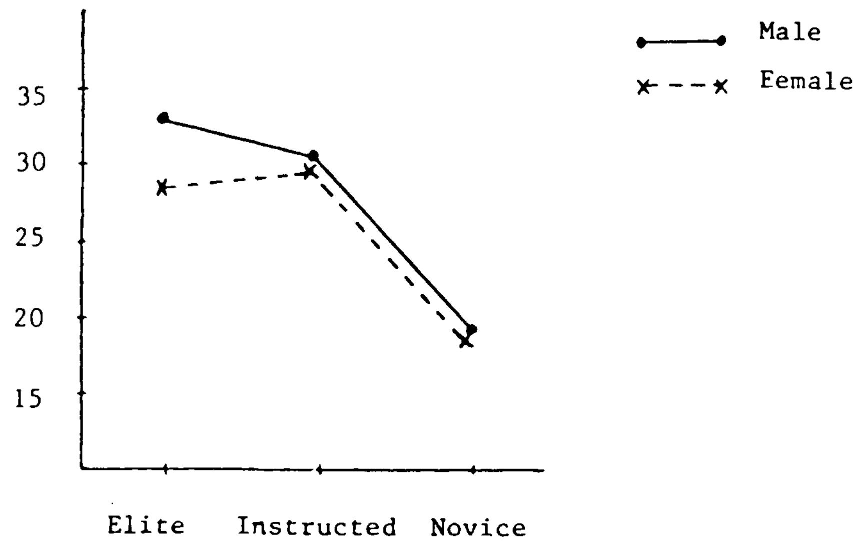


Table II

Analysis of Variance Table for the Cognitive Test

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	1414.29167	2	707.14583	41.55	0.0000
*H	50.02083	1	50.02083	2.94	0.0938
GH	49.29167	2	24.64583	1.45	0.2465
ERROR	714.87500	42	17.02083		

*G = Skill level *H = Gender

There was a highly significant skill level effect for the cognitive test with the elite subjects averaging slightly higher than the instructed (31.2 compared to 30) and both these groups much higher than the novices (19.2). A great difference was not expected between the elite and the instructed since the test was a criterion-referenced test constructed to evaluate introductory level knowledge. The significant levels effect supported construct validity of the test.

Performance Analysis

Figure 3

Graphic Representation of the Results of the Performance Analysis

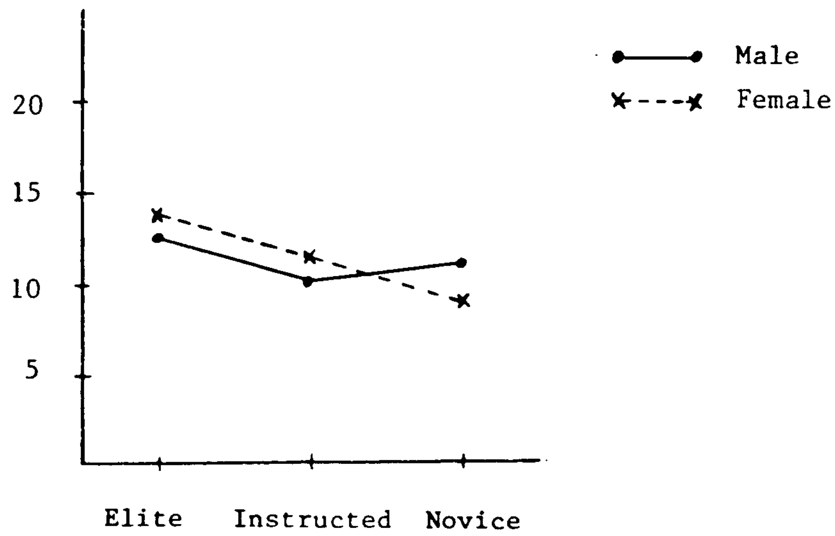


Table III

ANOVA Table for the Performance Analysis

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	63.87500	2	31.93750	6.58	0.0033
*H	0.33333	1	0.33333	0.07	0.7946
GH	23.04167	2	11.52083	2.37	0.1057
ERROR	204.00000	42	4.85714		

*G = Skill level   *H = Gender

A significant skill level effect was found for the performance analysis with elite subjects averaging 13 out of a possible 20, instructed averaging 11 and novice averaging 10.2. This significant skill level effect is consistent with the current body of knowledge concerned with observation ability (Barrett, 1979). Researchers agree that developing the ability to observe requires comprehensive knowledge of the skill being observed. Thus highly skilled and experienced players scored better than lesser skilled players.

Scores were relatively low for all subjects which may indicate that the video tape analysis was a more difficult test or that subjects were not as competent at the skill.

<u>Overhead Pass - Product Score</u>

Figure 4

Graphic Representation of the Results of the Overhead Pass - Product Score

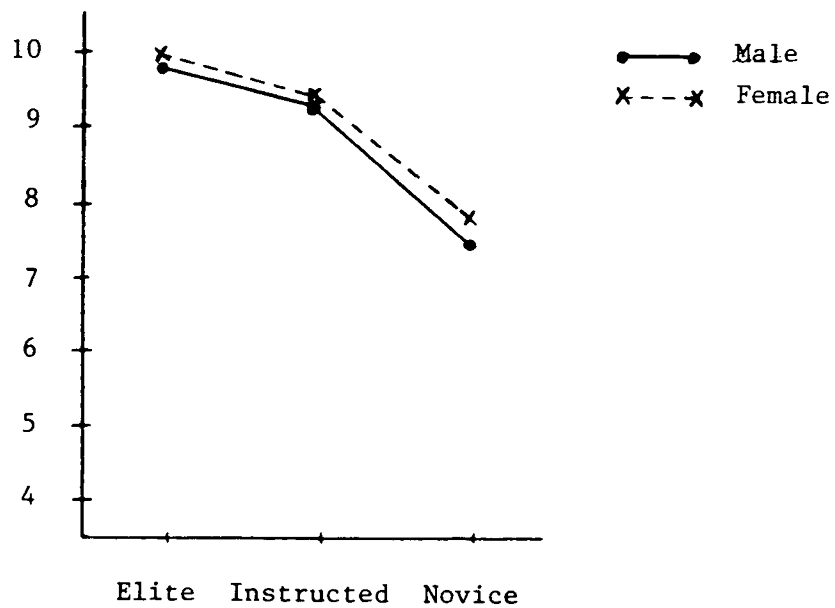


Table IV

ANOVA Table for the Overhead Pass - Product Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	44.66667	2	22.33333	18.35	0.0000
*H	0.52083	1	0.52083	0.43	0.5166
GH	0.16667	2	0.08333	0.07	0.9339
ERROR	51.12500	42	1.21726		

*G = Skill level *H = Gender

The product score refers to the ability of the subject to overhead pass the ball to a target. The trials were scored objectively; either the ball reached the target or it did not. There was a highly significant levels effect with elite scoring 9.9, instructed scoring 9.4 and novice scoring 7.6. There was no significant gender effect.

Forearm Pass - Product Score

Figure 5

Graphic Representation of the Results of the Forearm Pass - Product Score

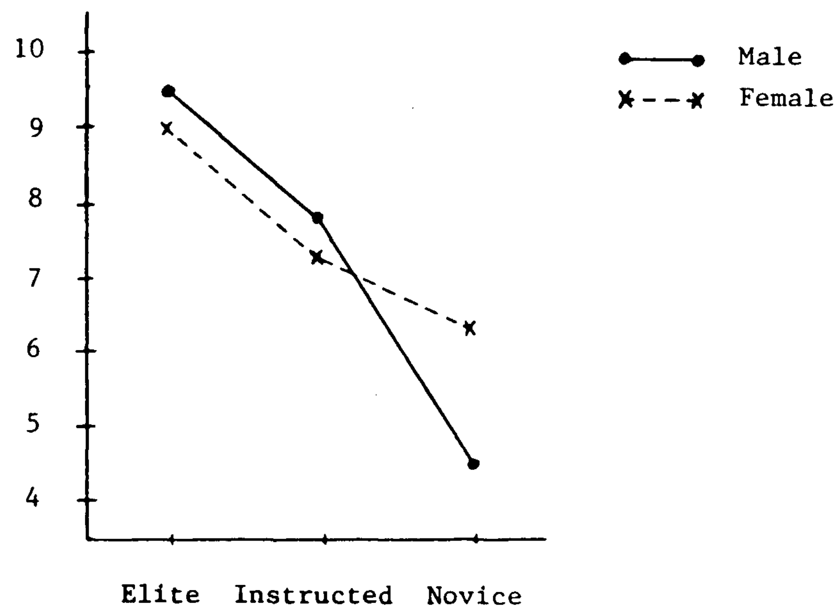


Table V

ANOVA Table for the Forearm Pass - Product Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	120.79167	2	60.39583	29.54	0.0000
*H	0.52083	1	0.52083	0.25	0.6164
GH	14.29167	2	7.14583	3.49	0.0394
ERROR	85.87500	42	2.04464		

*G = Skill level *H = Gender

The results of the Forearm Pass test showed a significant interaction between gender and skill at the .05 level. Scores for both genders decreased as skill level decreased, however females scored lower than males at the elite and instructed level but scored higher (6.3) than the males (4.5) at the novice level. A common problem with the forearm pass is to hit the ball too hard; perhaps males' greater strength was a disadvantage at the novice level.

Overhand Serve - Product Score

Figure 6

Graphic Representation of the Results of the Overhand Serve - Product Score

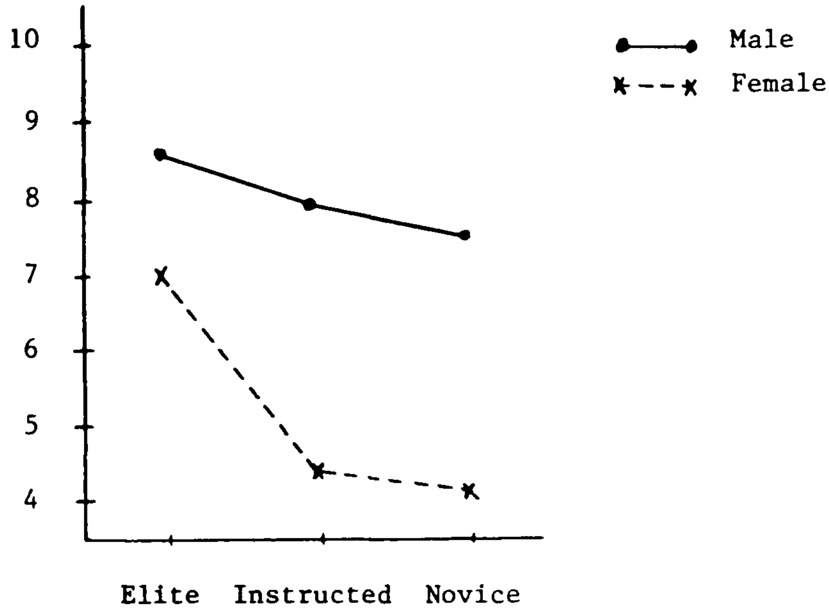


Table VI

ANOVA Table for the Overhand Serve - Product Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	34.62500	2	17.31250	4.17	0.0223
*H	102.08333	1	102.08333	24.57	0.0000
GH	10.04167	2	5.02083	1.21	0.3088
ERROR	174.50000	42	4.15476		

*G = Skill level *H = Gender

Table VI shows a highly significant gender effect and a significant skill effect at the .05 level. Figure 6 displays males performing 1.6 points better than females at the elite level, 3.6 points better at the instructed level and 3.5 points better at the novice level. Males' scores were probably much higher because of their increased upper body strength which is a definite asset in the overhand serve. As with all the tests thus far there was a significant skill effect which supports the construct validity of the proficiency test.

Spike - Product Score

Figure 7

Graphic Representation of the Results of the Spike - Product Score

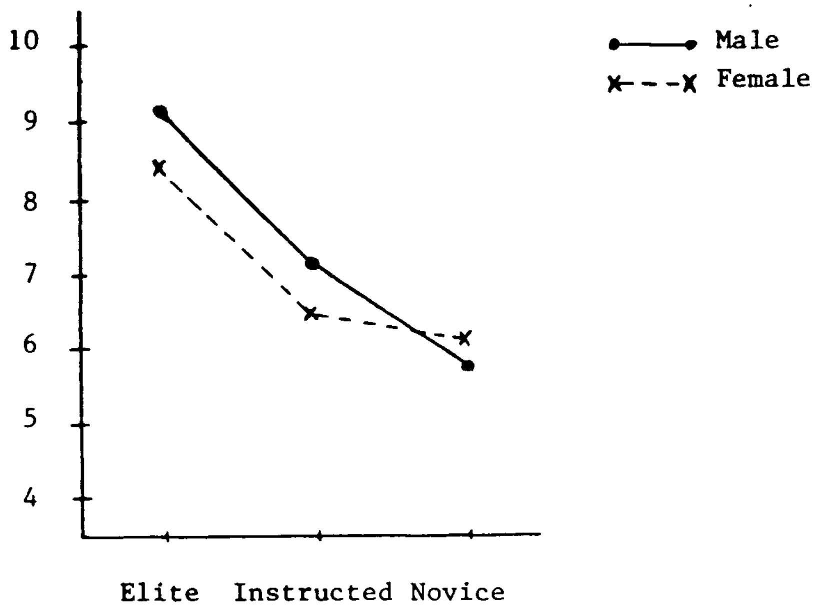


Table VII

ANOVA Table for the Spike - Product Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
* G	67.04167	2	33.52083	9.15	0.0005
* H	1.68750	1	1.68750	0.46	0.5011
GH	3.37500	2	1.68750	0.46	0.6341
ERROR	153.87500	42	3.66369		

*G = Skill level *H = Gender

The product score of the Spike test indicated a significant skill level effect with the elite scoring 8.8, the instructed at 6.8 and the novice at 6.0. Because technique was not part of this score, as long as the subjects accomplished the goal of getting the ball over the net and within the court boundaries, a point was scored. The trajectory and power of some of the spikes were questionable but because technique was evaluated in the process score, the product score was kept very objective.

Overhead Pass - Process Score

Figure 8

Graphic Representation of the Results of the Overhead Pass - Process Score

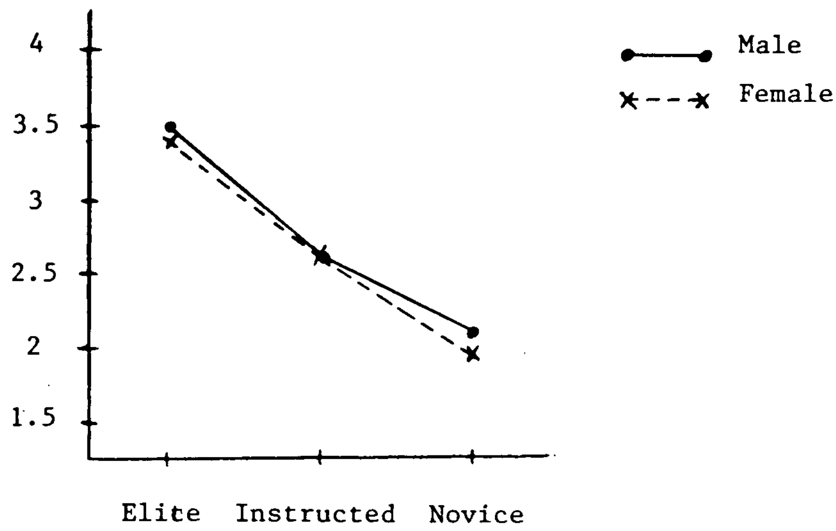


Table VIII

ANOVA Table for the Overhead Pass - Process Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	15.86531	2	7.93266	36.37	0.0000
*H	0.07521	1	0.07521	0.34	0.5602
GH	0.01260	2	0.00630	0.03	0.9715
ERROR	9.16000	42	0.21810		

*G = Skill level *H = Gender

The process score refers to the average score of both judges' ratings across all 10 trials. Table VIII shows a highly significant levels main effect. Elite subjects scored 3.5 while instructed scored 2.6 and novice scored 2.1. Differences between sexes were negligible.

Forearm Pass - Process Score

Figure 9

Graphic Representation of the Results of the Forearm Pass - Process Score

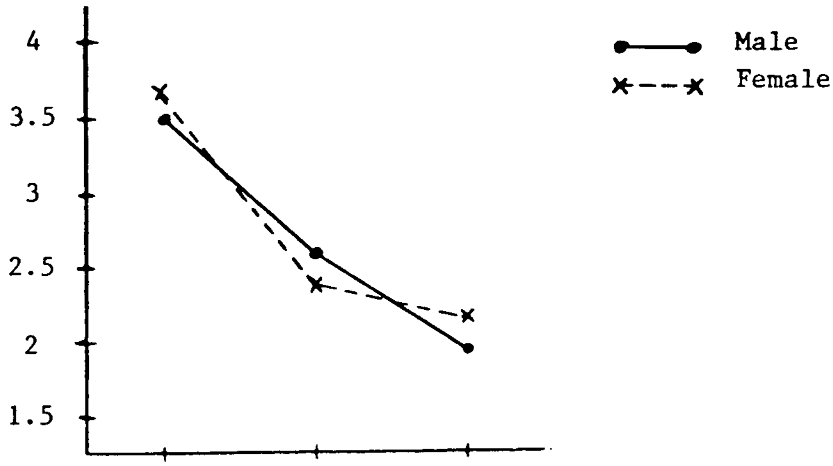


Table IX

ANOVA Table for the Results of the Forearm Pass - Process Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
* G	19.67197	2	9.83599	43.28	0.0000
* H	0.03797	1	0.03797	0.17	0.6848
GH	0.29344	2	0.14672	0.65	0.5295
ERROR	9.54531	42	0.22727		

*G = Skill level *H = Gender

Again construct validity is supported by the significant main effect of skill level for the forearm pass. No significant gender effects or interaction were identified in the study.

Overhand Serve - Process Score

Figure 10

Graphic Representation of the Results of the Overhand Serve - Process Score

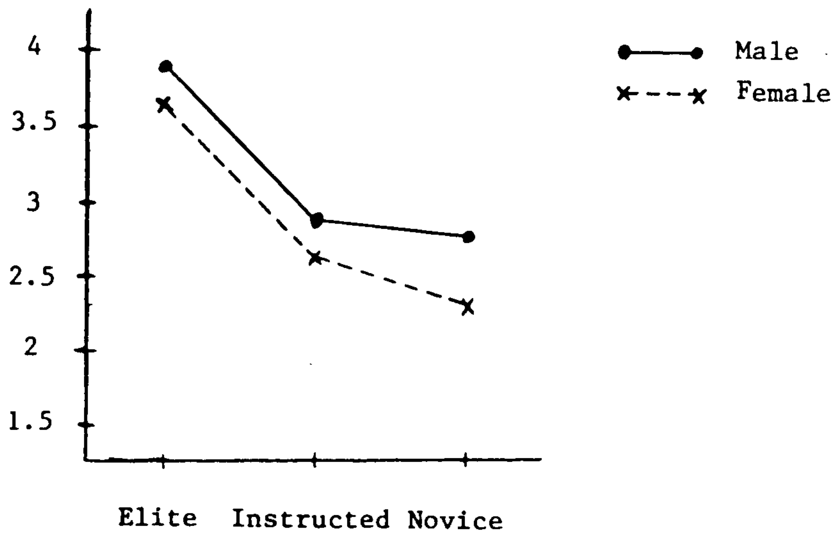


Table X

ANOVA Table for the Results of the Overhand Serve - Process Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	13.30761	2	6.65380	20.80	0.0000
*H	1.40083	1	1.40083	4.38	0.0425
GH	0.27635	2	0.13818	0.43	0.6521
ERROR	13.43500	42	0.31988		

*G = Skill level *H = Gender

Table X displays a significant skill level main effect and a significant (.04) gender main effect with males scoring higher than females. The arm motion required in the overhand serve closely resembles the overhand throw. Most males have greater experience with this action regardless of whether or not they have played volleyball. Prior experience plus upper body strength may be the explanation for novice males scoring only .1 lower than instructed males. Novice females scored .3 lower than instructed females.

<u>Spike - Proces Score</u>

Figure 11

Graphic Representation of the Results of the Spike - Process Score

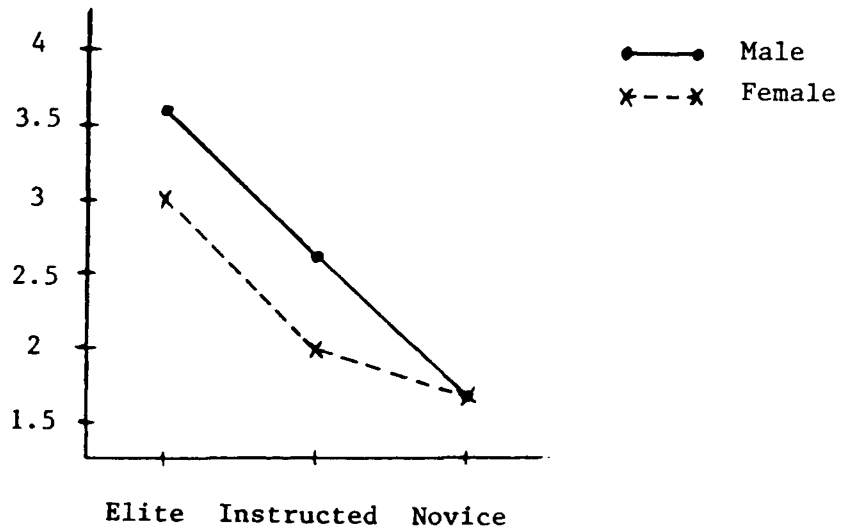


Table XI

ANOVA Table for the Results of the Spike - Process Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
* G	19.35948	2	9.67974	40.11	0.0000
* H	1.96021	1	1.96021	8.12	0.0067
GH	0.89135	2	0.44568	1.85	0.1703
ERROR	10.13562	42	0.24132		

*G = Skill level *H = Gender

Process scores for the spike indicate a highly significant skill effect with scores decreasing as skill level decreases. There is also significant gender effect with males scoring higher except at the novice level where both sexes scored 1.7. Increased strength and jumping ability probably explain the males significantly higher scores. Also the relative inexperience of the elite females compared to the elite males may surface as a factor here because spiking is such a complex skill. Elite males scored 3.6 while elite females scored 3.0.

Total Score

Figure 12

Graphic Representation of the Results of the Total Score

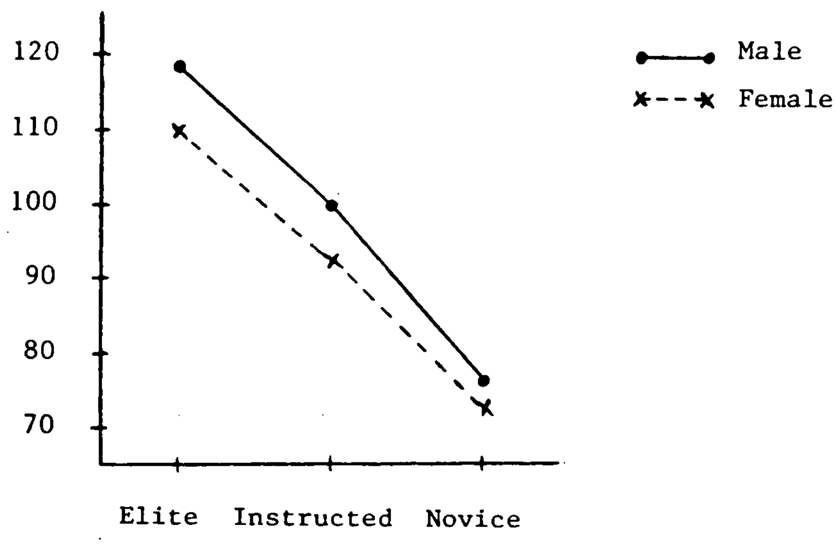


Table XII

ANOVA Table for the Total Score

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	TAIL PROB.
*G	12681.67383	2	6340.83691	53.95	0.0000
*H	588.87533	1	588.87533	5.01	0.0306
GH	40.66471	2	20.33236	0.17	0.8417
ERROR	4936.50195	42	117.53576		

*G = Skill level *H = Gender

The score of all test components was totalled to equal 140. It was initially determined that the product score and process score were weighted equally, consequently, each process score out of four was multiplied by 2.5 to give it the same value as the product score out of 10. The final equation was:

Cognitive (40) + Performance Analysis (20) + Product Scores (4x10) + Process Scores [(4x4) x 2.5)] = Total Score (140)

Table XII shows a highly significant main effect for skill level with elite subjects averaging 115, instructed subjects averaging 96 and novice averaging 75. Therefore, the proficiency test is a valid test for the construct of introductory level volleyball skills and knowledge.

The significant gender effect (.03) was not expected. In general, males scored higher than females at all three skill levels. Individual components of the test that demonstrated gender significance were the product score of the overhand serve and the technique or process score of the overhand serve and spike. As previously mentioned, the higher male scores can probably be attributed to enhanced upper body strength and jumping ability which are advantageous in the skill of serving and spiking. Another notable factor may be the relative inexperience of the elite females in comparison to the elite males. Although three of the 10 tests conducted showed males peforming significantly better than females, by looking at the graphs it can be seen that for these 48 subjects, three of the other tests (performance analysis, product score of the overhead pass and process score of the forearm pass) showed some evidence of females scoring higher than males. It is important to remember that initial development of the tests was based on the definiton of a good performance of each skill. Although males may have scored higher than females on some tests, this does not reduce the construct validity of the test. The skill level main effect was evident for both genders, therefore, no modifications in test construction are suggested.

The Reliability of the Cognitive Test

The cognitive test consisted of 40 multiple choice questions concerned with introductory level knowledge. Questions were constructed according to the content balance table that described the domain of introductory level volleyball, (refer to Table I, p. 25). The test was constructed as a criterion-referenced test, therefore, it would be inappropriate to analyze reliability using norm-referenced techniques. Reliability of a criterion-referenced test can be defined as "a measure of agreement over and above that which can be expected by chance between the decisions made about examinee mastery states" (Swaminathan, Hambleton, & Algina, 1974).

Reliability is usually analyzed from information gained in a test-retest situation. However, there are problems associated with retesting subjects on written tests because learning becomes a factor. For the present multiple choice test, results from the odd and even questions of all 48 subjects were separated and analyzed as two different tests.

Following the example of Swaminathan et al. (1974) and Safrit (1977), the proportion of agreement of mastery categorizations between the odd and even questions was determined by adding the proportions in the main diagonal. (Refer to Table XIII).

Table XIII

Proportion of Agreement Between Odd and Even Questions with an 80% Mastery Criterion

		Even					
		Mastery		Non-mastery			
		N	P	N	P	N	P
Odd	Mastery	9	.19	9	.19	18	.38
	Non-mastery	5	.10	25	.52	30	.62
		14	.29	34	.71	48	.71*

*Proportion of Agreement
N = Number of Individuals
P = Proportion of Individuals

The value of .71 was interpreted as meaning 71% of the categorizations made on the two different tests (odd and even) were in agreement. In order to account for the correct categorizations that were made purely by chance a further analysis was applied and the coefficient of agreement ($\hat{K}$) was calculated. For information on calculating the kappa ($\hat{K}$) coefficient, the reader is referred to Appendix II in Safrit (1977).

The kappa coefficient for the cognitive test with an 80% mastery criterion equalled .36. When the mastery criterion was lowered to 75% the proportion of agreement was again .71 but $\hat{K}$ increased to .44. Safrit (1977) warns that interpretation of $\hat{K}$ is not entirely clear. When marginal values are equal, $\hat{K}$ is equal to the phi coefficient and would thus be interpreted much like a correlation coefficient. The problem is that a relatively small number of misclassifications seems to yield a low kappa coefficient. At both the 75% and 80% criterion level only 14 of 48 subjects (30%) were misclassified and the resulting $\hat{K}$'s were .44 and .36 respectively. Perhaps new research will yield a better interpretation of kappa. For the present study, 71% of the subjects were categorized consistently which actually only accounts for 21% classification better than chance, so reliability of the cognitive test is questionable.

Reliability of Performance Analysis

The performance analysis was composed of 20 video-taped episodes of skill errors with appropriate multiple choice questions. Episodes were chosen to represent the four individual skills of the overhead pass, forearm pass, overhand serve and the spike. Five game situation errors were also included. It was very diffiuclt to define the domain of performance analysis for introductory level volleyball since little knowledge has been published in this area. Lacking specific performance objectives, the decsion was made to construct the performance analysis as a norm-referenced test.

The appropriate method of analysis for reliability is the Pearson Product-Moment Correlation. A co-ed group of 12 instructed subjects were tested on two dates with five days between testing sessions. The Pearson Product-Moment produced an r of .81. According to Johnson and Nelson (1979) a correlation coefficient of at least .80 is desirable for test relaiblity. Reliability of the pefromance analysis is acceptable but with the subject sample being so small and homogeneous, a difference of one mark would drastically affect the coefficient. It would be useful to collect test-retest data on a larger and more heterogeneous sample.

Reliability and Objectivity of the Skill Tests

The mean scores, averaged across subjects and observers, are presented in Table XIV for each of the skill tests.

Table XIV

Mean Values for the Overhead Pass (OP), Forearm Pass (FP), Overhand Serve (OS) and Spike (SP) - Process Scores

SKILL LEVEL		OP	FP	OS	SP
Elite	M	3.52	3.52	3.86	3.55
	F	3.39	3.66	3.63	2.96
Instructed	M	2.63	2.56	2.86	2.55
	F	2.58	2.40	2.62	1.95
Novice	M	2.09	1.98	2.83	1.73
	F	2.03	2.17	2.27	1.71

The highest possible score was four. Generally, subjects at all skill levels scored highest on the overhand serve test while the spike test produced the lowest scores for all except the elite males. Relatively speaking, the volleyball spike which includes an approach, judgement of the ball, jumping and body motion while in the air, is a much more complex motor skill than the overhand serve; thus the lower scores. In order for the volleyball skill tests to be considered useful measuring tools, reliability and objectivity had to be established.

Traditionally subjects would have to perform the tests on two occasions and the reliability would be determined by analyzing the results with a Pearson Product-Moment correlation. Objectivity or inter-observer reliability would be indicated by the percentage of inter-observer agreement. In a study such as the present one, the complex design distinguished 19 potential sources of variation. Rather than allowing for differentiation between these sources of variation, the Pearson Product-Moment correlation averages variances over all sources. It is clear that interpretation of such a correlation coefficient would be very difficult. Generalizability theory is a more sophisticated statistical procedure that was developed by Cronbach, Gleser, Nanda and Rajartnam, in 1972 to solve this problem. Basically, generalizability theory uses analysis of variance to determine the relative contribution of each source of unreliability. Variance components are estimated from the mean square values. These variance components are arranged into an equation according to facets of generalization and facets of differentiation, (Cardinet, Tourneur and Allal, 1976). A facet of generalization is a source of variation which affects the measures taken of the objects under study. A facet of differentiation is an object or characteristic which is to be compared in a study. Facets of differentiation are held constant in order to determine the amount of variation that occurs in the selected facet of generalization.

The degree to which a set of scores can be generalized across the facets of generalization results in a generalizability coefficient which can be interpreted much like a reliability coefficient (Mosher and Schutz, 1983). The researcher must decide which sources of variation it is important to generalize over.

The design for the study was a two by three by two by two factorial (Gender by Levels by Observers by Trials) with repeated measures on trials and observers. In order to determine reliability, trials were divided into one to five and six to ten for the analysis. The random effects were Observers, Trials and Subjects while Gender and Levels were fixed. Mosher and Schutz (1983) conducted a similarly designed study for the Overarm Throw and indicated that because both random and fixed effects were present, the design was considered to be a mixed model. As such, F ratios were not calculated for all sources of variance because appropriate error terms were not available from the BMD P8:V program. The same is true in the present study thus Quasi F ratios were constructed as required for the sources of variation (Kirk, 1968). Each of the four volleyball skills (overhead pass, forearm pass, overhand serve and spike) were analyzed separately. An analysis of variance table is presented for each skill. It includes significant effects, variance estimates and the percent of total variance accounted for by each source of variation. G coefficients for all skill tests are discussed in a later section.

Overhead Pass

Table XV

Analysis of Variance and Variance Estimates: Overhead Pass

Source	DF	MS	F	P	Variance Estimate	Percent of Total Variance
Level (L)	2	31.73100			.47510	57.00
Gender (G)	1	0.30083			.00000	0.00
Observers	1	1.02080			.00839	1.00
Trials (T)	1	1.203300			.01104	1.50
LG	2	0.025208			.00000	0.00
LO	2	0.587710			.01526	1.90
GO	1	0.520830			.00859	1.00
LT	2	0.033958			.00020	0.02
GT	1	0.053333			.00035	0.04
OT	1	0.120000	2.550	.1181	.00152	0.20
S(LG)	42	0.872380			.17679	21.60
LGO	2	0.406460	*3.908	<.0500a	.01891	2.42
LGT	2	0.012708	*0.397	>.0500	.00000	0.00
LOT	2	0.004375	0.090	.9116	.00000	0.00
GOT	1	0.013330	0.280	.5977	.00000	0.00
OS(LG)	42	0.142140	3.020	.0003a	.04750	5.80
TS	42	0.070238	1.490	.1003	.01155	1.54
LGOT	2	0.008958	0.190	.8276	.00000	0.00
OTS(LG)	42	0.047143			.04714	5.80
						100%

* = quasi F test a = significant effects

Table XV presents the ANOVA information for the overhead pass. Quasi F ratios were calculated for the third level interactions of Level by Gender by Observers and Level by Gender by Trials. A significant interaction was found for the LGO term which means that observers were not consistent in scoring genders over skill levels. Although the interaction was significiant it is only responsible for 2.42% of the total variance. The significant second level interaction for observers by subjects cannot really be interpreted because of the previous higher order interaction. For the same reason F's were not calculated for the main effects of Levels, Gender, Observers and Trials. However, important information about the reliability and objectivity of the overhead pass test can be gained from the percentages of total variance for each effect. The variability in skill level accounts for 57% of the total variance of the overhead pass. This is positive support for the construct validity of the test. Subjects within Levels and Genders contributed 21.6% of the variance which simply demonstrates interindividual variablity. Observers by subjects describes 5.8% of the total variance while all other sources of variance are negligible.

Forearm Pass

Table XVI

Analysis of Variance and Variance Estimates: Forearm Pass

Source	DF	MS	F	P	Variance Estimate	Percent of Total Variance
Level (1)	2	39.344000			.60010	65.00
Gender (G)	1	0.151880			.00000	0.00
Observers	1	1.960200			.01822	2.20
Trials (T)	1	0.226880			.00098	0.10
LG	2	0.586880			.00000	0.00
LO	2	0.101460			.00000	0.00
GO	1	0.046875			.00000	0.00
LT	2	0.114380			.00224	0.24
GT	1	0.001875			.00000	0.00
OT	1	0.091875	3.690	.0616	.00140	0.15
S(LG)	42	0.909080			.18095	19.80
LGO	2	0.056870	*4.260	<.0500a	.00000	0.00
LGT	2	0.105620	*1.928	>.0500	.00318	0.30
LOT	2	0.001875	0.080	.9276	.00000	0.00
GOT	1	0.060208	2.420	.1275	.00147	0.18
OS(LG)	42	0.144480	5.800	.0000a	.05976	6.50
TS(LG)	42	0.065740	2.640	.0011a	.02042	2.40
LGOT	2	0.013958	0.560	.5752	.00000	0.00
OTS	42	0.024910			.02491	2.80
						100%

* = quasi F test   a = significant effects

Table XVI also shows a significant interaction for Levels by Genders by Observers but this source of variability did not account for any of the percentage of total variance in the forearm pass. The second level interactions of Observers by Subjects and Trials by subjects are significant and contribute 6.5% and 2.4% of the total variance. This means observers did not score subjects consistently and subjects did not perform consistently over trials. Variability in skill level again contributes the highest percentage of variance at 65%; Subjects within Levels and Genders are responsible for 19.8% of the variance while variability of the observers accounts for 2.2% of the total variance.

Overhand Serve

Table XVII

Analysis of Variance and Variance Estimates: Overhand Serve

Source	DF	MS	F	P	Variance Estimate	Percent of Total Variance
Level (L)	2	26.615000			.39424	48.0
Gender (G)	1	5.603300			.04539	5.5
Observers	1	0.187500			.00105	0.1
Trials	1	1.080000			.01036	1.3
LG	2	0.552710			.00000	0.0
LO	2	0.208130			.00466	0.6
GO	1	0.083330			.00053	0.0
LT	2	0.035625			.00000	0.0
GT	1	0.020833			.00000	0.0
OT	1	0.030000	1.2100	.2773	.00011	0.0
S(LG)	42	1.279500			.28554	34.7
LGO	2	0.046458	*0.6651	>.0500	.00000	0.0
LGT	2	0.018958	*0.2780	>.0500	.00000	0.0
LOT	2	0.001875	0.0800	.9272	.00000	0.0
GOT	1	0.000830	0.0300	.8553	.00000	0.0
OS(LG)	42	0.081905	3.3100	.0001a	.02857	3.5
TS(LG)	42	0.080238	3.2400	.0001a	.02774	3.4
LGOT	2	0.012708	0.5100	.6023	.00000	0.0
OTS(LG)	42	0.024762			.02476	3.0
						100%

* = quasi F test   a = significant effect

Results in Table XVII show that significant interaction was found for Observers by Subjects and Trials by Subjects with the former contributing 3.5% of the variance and the latter responsible for 3.4%. It is interesting to note that variability due to gender contributed 5.5% of the total where previously in the overhead and forearm pass gender differences were too small to be considered (0%). Variability in skill level accounted for only 48% of the total variance in the serve compared to 57% for the overhead pass and 65% for the forearm pass. Examination of the cell means shows that males scored consistently higher than females in all three skill levels and also that there was little difference between the score of novice (2.83) and instructed (2.86) males. This would explain the somewhat lower percent of variance due to skill level. These findings correspond to the general statements made earlier that all subjects scored higher on the overhand serve test. Subjects within Levels and Genders contributed a relatively high 34.7% of the total variance which means interindividual variability was high. These findings are consistent with the results of the analysis of variance conducted for construct validity.

Spike

Table XVIII

Analysis of Variance and Variance Estimates: Spike

Source	DF	MS	F	P	Variance Estimate	Percent of Total Variance
Levels (L)	2	38.719000			.59176	60.3
Gender (G)	1	7.840800			.07293	7.4
Observers	1	0.030000			.00000	0.0
Trials (T)	1	0.067500			.00000	0.0
LG	2	1.782700			.02472	2.5
LO	2	0.015625			.00000	0.0
GO	1	0.040833			.00000	0.0
LT	2	0.016875			.00000	0.0
GT	1	0.000000			.00000	0.0
OT	1	0.000830	0.0400	.8435	.00000	0.0
S(LG)	42	0.965300			.20720	21.1
LGO	2	0.057709	*0.6712	>.0500	.00000	0.0
LGT	2	0.150620	*1.2490	>.0500	.00187	0.2
LOT	2	0.015208	0.7200	.4928	.00000	0.0
GOT	1	0.030000	1.4200	.2401	.00037	0.0
OS(LG)	42	0.061480	2.9100	.0004a	.02018	2.0
TS(LG)	42	0.096131	4.5500	.0000a	.03750	3.8
LGOT	2	0.045625	2.1600	.1281	.00306	0.3
OTS(LG)	42	0.021131			.02113	2.2
						100%

* = quasi F test a = significant effect

Table XVIII again indicates significant interactions for Observers by Subjects and Trials by Subjects. Contribution to total variance is 2% for OS(LG) and 3.8% for TS(LG). Variability due to skill level accounts for 60% of the total variance of the scores on the spiking test. Gender differences contribute 7.4% of the total variance and cell means show that males score consistently higher than females at all skill levels although differences at the novice level are negligible; males - 1.73, females - 1.71. It seems that superior strength and jumping ability afford male subjects an advantage over females at the same skill level. However a lack of skill at the novice level cannot be compensated for by strength as seemed to be the case for novice males in the overhand serve. Variability of Subjects is responsible for 21.1% of the total variance in comparison to accounting for 34.7% of the variance for the serve. Note that variability due to Observers or Trials is too small to be considered in the total variance indicating the test for spiking has high reliability and objectivity.

Generalizability of Results

The 48 generalizability coefficients in Table XIX were tabulated by following the "Rules of thumb for estimating reliability coefficients using generalizability theory", (Rentz, 1980). Equations were developed for the three types of generalization that were considered to be important in this study. Inter-rater reliability (objectivity) is the degree to which any other set of observers would obtain the same results if they saw the same subjects performing the exact same trials. Inter-trial reliability refers to the degree to which the same subjects, observed by the same judges, would receive the same score on a different set of trials. Performer reliability considers the degree to which the same subjects would receive the same score if they performed another set of trials for different judges.

The variance due to Levels was not included in any of the G coefficient equations. Like Mosher and Schutz's Overarm Throwing Test (1983), future use of the Volleyball Proficiency Test will be for a fairly homogeneous group so inclusion of the Levels effect will unrealistically inflate G coefficients.

Table XIX

Generalizability Coefficients for Each Skill of the Volleyball Proficiency Test

Type of Reliability	Overhead Pass	Forearm Pass	Overhand Serve	Spike
1) Inter-rater Reliability				
G_1: 2 observers, 10 trials	.86	.85	.96	.96
G_2: 2 observers, 5 trials	.85	.85	.95	.96
G_3: 1 observer, 10 trials	.75	.75	.91	.93
G_4: 1 observer, 5 trials	.74	.74	.91	.92
2) Inter-trial Reliability				
G_1	.98	.98	.99	.99
G_2	.97	.97	.98	.99
G_3	.98	.98	.99	.99
G_4	.95	.96	.98	.99
3) Performer Reliability				
G_1	.85	.85	.95	.95
G_2	.83	.83	.94	.93
G_3	.74	.74	.91	.92
G_4	.72	.72	.90	.90

In Table XIX the generalizability coefficients for inter-rater reliability demonstrate that the objectivity of the skill tests is good. The highest coefficients are seen for the overhand serve and the spike. In fact, reducing test protocol from two observers, 10 trials to one observer, five trials only reduces the G coefficient for the serve from .96 to .91 and the G coefficient for the spike from .96 to .92. Judges' agreement was not quite as high on the overhead pass and the forearm pass where G coefficients for two observers and 10 trials were .86 and .85, respectively. When only one observer was used, the coefficient dropped to .75 for both skills. For all four skills the difference in using 10 trials versus five trials was very negligible. The greatest reduction in reliability occurred when the number of observers was reduced from two to one. However, even the lowest G coeffecent of .74 is still acceptable in terms of observer reliability. These results indicate that one trained observer could reliably evaluate a class of volleyball students using the four skill tests and the corresponding rating scale.

As witnessed in the second set of G coefficients, inter-trial reliability for all skills was so high that very little if any extra information was gained by increasing the number of trials from five to 10. From a practical point of view this is very positive for the university instructor who must test 20-30 students in a one or two hour testing session. With G coefficients ranging form .95 to .99 it seems to make little difference whether one or two observers are being used. Again the overhand serve and spike skill tests show the highest inter-trial reliability (.99) indicating that the same subjects performing another set of trials with the same observers would score the same.

The performer reliabilities are also very high. The G coefficients refer to the degree to which performers would achieve the same scores if they participated in another set of trials with different observers. Again the overhand serve and spike seem to be the most reliable with values ranging from .95 for two observers and 10 trials to .90 for one observer and five trials. Performer reliabilities for the overhead pass and forearm pass are not quite as high. Both coefficients are .85 when two observers and 10 trials are used and .72 when only one observer and five trials are used. This may be because errors in these two skills are not as easy to differentiate as they are in the spike and overhand serve.

In general, it can be concluded that the four volleyball skill tests are reliable and objective instruments. Generalizability coefficients for the four skill tests conducted under the protocol of two observers and 10 trials were all .85 and higher. The reduction of the number of trials from 10 to five only slightly reduced the G coefficient (.02 or less). When only one observer is used the coefficients show a greater decrease with values for five trials ranging from .72 to .99.

Correlation Between Test Components

A correlation matrix was computed by using the Pearson Product-Moment on the scores of all 48 subjects on the 10 test components.

Table XX

Correlation Between Process Scores

	Overhead Pass	Forearm Pass	Overhand Serve	Spike
OP	1.00			
FP	.85	1.00		
OS	.74	.78	1.00	
SP	.78	.80	.75	1.00

Table XX displays the finding that subjects scoring well on technique for one skill scored relatively well on technique for all skills with the relationship between the overhead pass and the forearm pass being highest at .85. This finding supports the previously observed significant skill level effect and thus the construct validity of the test. The relationships between process scores represent the highest correlations between any test components, but they are not high enough to indicate that testing only one skill would provide adequate information to generalize to all skills.

Another interesting relationship occurred between the product and process scores of each skill. (Refer to Table XXI).

Table XXI

Correlation Between Product and Process Scores for All Skills

Process Scores	Product Scores			
	Overhead Pass	Forearm Pass	Overhand Serve	Spike
OP	.60a	.68	.48	.61
FP	.52	.75a	.38	.68
OS	.41	.56	.62a	.56
SP	.58	.67	.50	.74a

a = highest correlation for each skill

The highest relationship in each case was found between the product and process score of the same skill, i.e., the OP product score correlated higher with the OP process score than with any other process score. This was an encouraging finding because it meant that subjects with the best technique (process) were also getting the best accuracy score (product). If these correlations were too high then both tests would be evaluating the same characteristic and one of them would therefore be redundant. This, however, was not the case as the correlations fell between .60 and .75. Therefore, only 35% to 45% of the variance in one score was accounted for by the variance in the other score.

Results for the performance analysis test revealed relatively low correlations from .15 with the FP product score to .46 with the OS process score. It seems that the performance analysis evaluated a skill ability quite different from the other components in the proficiency test. This finding in combination with the fact that a significant skill level effect was discovered in the analysis of variance indicates that performance analysis is a distinct component of introductory level volleyball. Barrett (1979) reviewed current literature related to performance analysis and stressed that both the ability to observe and the ability to analyze movement are important for teachers and coaches engaged in performance analysis. She concludes that "the need for teaching observation as a specific skill for effective teaching was considered essential.", (Barrett, 1979, p. 67). The video-tape developed for the present study should be an invaluable tool for this purpose.

<u>Comparison of Number of Subjects Achieving Mastery in Each Skill Level</u>

One of the purposes of the proficiency test was to exempt students from an introductory level volleyball course if the material was already mastered. To investigate the association between skill level and achievement of mastery, a Chi Square analysis was conducted. The mastery criterion was set at 80 %. Genders were collapsed because non-significant gender effects were found for most of the test components. Table XXII presents the Chi Square scores and levels of significance for each test component.

Table XXII

Chi Square Values and Levels of Significance of Test Components

Test Component	Chi Square	Degrees of Freedom	Significance
Cognitive	7.312	2	.03a
Performance Analysis	2.043	2	.36
Product Score - OP	10.666	2	<.01a
FP	25.210	2	<.01a
OS	2.032	2	.36
SP	13.500	2	<.01a
Process Score - OP	26.063	2	<.01a
FP	39.999	2	<.01a
OS	20.202	2	<.01a
SP	13.137	2	<.01a
Total Score	28.541	2	<.01a

a = significant effect

A significant Chi Square provides fairly conclusive evidence that achievement of mastery differentiates between individuals on the basis of their skill level, (Ferguson, 1976). Readers are cautioned that some expected cell frequencies were less than five. When this occurs significant differences become inflated as may be the case with some of the significant effects.

Results of the performance analysis test did not show an association between skill level and mastery. Examination of the Chi Square table shows that only one elite subject achieved mastery while all other subjects were classified as non-masters. This corresponds to the initial results of the analysis of variance where scores were low for all subjects.

The product score of the overhand serve also demonstrated no association between mastery and skill level. Results showed nine elite, seven instructed and five novice subjects achieving mastery. These differences were obviously not great enough to be considered significant. This result was probably due to the very large gender difference found in the analysis of variance. Males scored well at all levels while female scored poorly at all levels. The sexes were combined in the Chi Square analysis and thus no significant difference was evident. All Chi Square tables can be located in Appendix E.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to construct a reliable and valid assessment tool to determine the cognitive and psychomotor level of proficiency possessed by an individual.

The test evaluated four components of introductory level volleyball skill: 1) knowledge of skills, strategies and rules, 2) performance analysis ability, 3) objective skill ability (product score) and 4) subjective skill ability (process score). The four volleyball skills utilized to determine product and process scores were the overhead pass, the forearm pass, the overhand serve and the spike. Subjects performed 10 trials of each skill and were subjectively rated by two judges on technique demonstrated during the trials.

The subject pool consisted of 24 females and 24 males divided evenly into three levels of skill ability: elite, instructed, novice or non-instructed.

Construct validity for the tests was established by a series of two by three analysis of variance computations for each test individually and then as a total score.

Reliability of the cognitive criterion-referenced test was computed by a proportion of agreement test and the kappa coefficient. Reliability of the norm-referenced performance analysis was computed by the Pearson Product-Moment Correlation. Reliability and objectivity of the skill tests were determined by generalizability coefficients.

- The correlation between test components was investigated by using the Pearson Product-Moment Correlation and the Chi Square statistic was employed to determine if there was a relationship between skill level and a student's ability to achieve mastery.

Major Findings

The following were major findings of the study:

1) Analysis of variance revealed a significant skill level effect for each test component: cognitive, performance analysis, product score for overhead pass, forearm pass, overhand serve and spike, process score for overhead pass, forearm pass, overhand serve and spike. The overhand serve product score was significant at the .02 level of probability and all others were significant at .01.

2) Analysis of variance showed a significant gender effect for overhand serve product scores (<.01), overhand serve process scores (.04), spike process scores (<.01) and total test scores (.03).

3) Analysis of variance displayed one positive skill and gender interaction for the forearm pass product score (.03).

4) Reliability of the cognitive test showed 71% of the mastery categorizations made between odd and even questions were in agreement when either a 75% or 80% criterion was used. To account for categorizations made purely by chance the kappa ($\hat{K}$) coefficient was determined. With the criterion set at 80%, $\hat{K} = .36$; with a criterion of 75%, $\hat{K} = .44$.

5) Correlation between a test-retest of the performance analysis resulted in a reliability of .81.

6) The generalizability coefficients for inter-rater reliabilities of two observers and 10 trials were: .86 for overhead pass, .85 for forearm pass, .96 for overhand serve and .96 for the spike.

7) The generalizability coefficients for inter-trial

reliabilities of two observers and 10 trials were .98 for overhead pass, .98 for forearm pass, .99 for overhand serve and .99 for the spike.

8) The generalizability coefficients for performer reliabilities of two observers and 10 trials were .85 for overhead pass, .85 for forearm pass, .95 for overhand serve and .95 for the spike.

9) G coefficients were also determined for two observers - five trials, one observer - 10 trials and one observer - five trials. Generally, the coefficients showed a decrease when one observer was dropped from the data but reducing the number of trials from 10 to five had very little effect on the G coefficients. (Refer to Table XIX, p. 66).

10) Correlation between test components showed subjects scoring well on technique for one skill (process score) scored well on technique for all skills. Correlations ranged between .74 and .85.

11) Correlations between the product and process score of each skill were higher than any correlations between skills, i.e., .60 to .75.

12) The highest correlation between the performance analysis and any other test component was .46 with the overhand serve process score.

13) Chi Square values were significant for nine of the 11 test components: cognitive, product score of the overhead pass, forearm pass and spike, process score of the overhead pass, forearm pass, overhand serve and spike and total score. These significant effects show that achievement of mastery differentiates between individuals on the basis of skill level.

14) Chi Square results for the performance analysis and overhand serve product score were not significant therefore providing no evidence of a relationship between mastery and skill level.

Conclusions

From the results attained in this study the following conclusions appear warranted:

1) All components of the Volleyball Proficiency Test are valid measures of introductory level volleyball skill.

2) Reliability of the cognitive test under the present method of analysis is questionable.

3) The performance analysis is a reliable measure.

4) The psychomotor skill tests are reliable and objective measures of introductory level volleyball performance.

5) Test components are related but not redundant.

6) Nine of the 11 test components indicated a significant relationship between achievement of mastery and skill level.

Recommendations

1) It is recommended that further reliability studies of the cognitive test and the performance analysis be conducted on a larger and more heterogeneous sample population.

2) It may be necessary to modify the performance analysis to make it less difficult. Although construct validity was evident, only one subject was able to achieve a mastery score when the criterion was set at 80%.

3) It is suggested that the proposed evaluation tool be used as a practical measure of proficiency for introductory volleyball courses at the college and university level.

4) It is hoped that the theoretically based volleyball proficiency test will serve as an example and a stimulus for experts in other sporting areas to construct suitable tests for their activities.

Bibliography

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement. Washington, DC: American Council on Education.

Barrett, Kate R. (1979). Observation of movement for teachers - a synthesis and implications. Motor Skills: Theory into Practice, 3(2), 67-76.

Block, J.H. (Ed.). (1974). Schools, society and mastery learning. New York: Holt, Rinehart and Winston.

Block, J.H. (1979). Mastery learning; the current state of the craft. Educational Leadership, 114-117.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: David McKay Company.

Bloom, B. S. (1968). Learning for mastery. Evaluation Comment, 1(2), 1-8.

Brown, F. G. (1981). Measuring classroom achievement. New York: Holt, Rinehart and Winston.

Cardinet, J., Tourneur, Y., & Allal, L. (1976). The symmetry of generalizability theory: application to educational measurement. Journal of Educational Measurement, 13, 119-135.

Carroll, J. B. (1971). Problems of measurement related to the concept of learning for mastery. Chapter 3 in J. H. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart and Winston.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajartnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Fawcett, Sandra A. (1974). Validity of volleyball tests. New Zealand Journal of Health, Physical Education and Recreation, 7(3).

Ferguson, George A. (1976). Statistical analysis in psychology and education (4th ed.). New York: McGraw Hill Book Company.

Fitts, P. M. (1962). Factors in complex skill training. In R. Glasser (Ed.), Training research and education. Pittsburgh: University of Pittsburgh Press.

Fitts, P. M. (1964). Perceptual motor skill learning. In A. W. Melton (Ed.), Categories of Human Learning. New York: Academic Press.

Fitts, P. M. (1967). Human performance. Belmont, California: Brooks/Cole Publishing Company.

Fleishman, E. A. (1964). The structure and measurement of physical fitness. Englewood Cliffs, New Jersey: Prentice Hall.

Gagne, R. M. (1965). The conditions of learning. New York: Holt, Rinehart and Winston.

Gagne, R. M. (1970). The conditions of learning (2nd ed.). New York: Holt, Rinehart, and Winston.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. American Psychologist, 18, 510-522.

Godbout, P. & Schutz, R. W. (1983). Generalizability of ratings of motor performances with reference to various observational designs. Research Quarterly for Exercise and Sport, 54(1), 20-27.

Gronlund, N. E. (1973). Preparing criterion - Referenced tests for classroom instruction. New York: Macmillan.

Harrow, A. J. (1972). A taxonomy of the psychomotor domain. New York: David McKay Company.

Jewett, Ann E. (1973). Physical education objectives out of curricular chaos. Proceedings of the AAHPER Regional Conference. Pennsylvania.

Jewett, A. E., Jones, L. S., Luneke, S.M., & Robinson, S. M. (1971). Educational change through a taxonomy for writing physical education objectives. Quest, 16, 32-38.

Jewett, A. E., & Mullan, M. R. (1977). Curriculum design: Purposes and processes in physical education teaching - learning. Washington, DC: AAHPER publications.

Johnson, B.L., & Nelson, Jack K. (1979). Practical measurements for evaluation in physical education (3rd ed.). Minneapolis: Burgess Publishing Company.

Johnson, Judith A. (1967). The development of a volleyball skill test for highschool girls. Unpublished master's thesis, Illinois State University.

Kautz, Elizabeth Marie. (1976). The construction of a volleyball skill test for the forearm pass. Unpublished master's thesis, University of North Carolina, Greensboro.

Kim, Hogwon. (1971). Mastery learning in the Korean middle schools. UNESCO Regional Office for Education in Asia, 6(1), 55-66.

Kirk, Roger, E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole Publishing Company.

Krathwohl, D. R. et. al. (1964). Taxonomy of educational objectives. Handbook II: The affective domain. New York: David McKay and Company.

McGee, R., & Drews, F. (1974). Proficiency testing for physical education. Washington: AAHPER publications.

Merrill, M. D. (Ed.). (1971). Instructional design: Readings. Englewood Cliffs, New Jersey: Prentice Hall.

Merrill, D. M. (1972) Psychomotor taxonomies, classifications, and instructional theory. Chapter 14 in R. Singer (Ed.), The psychomotor domain: Movement behvior. Philadelphia: Lea and Febiger.

Mosher, R. E. & Schutz, R. W. (1983). The development of a test of overarm throwing: An application of generalizability theory. Canadian Journal of Applied Sport Sciences, 8(1), 1-8.

Popham, W. J. (1975). Educational evaluation. Englewood Cliff, New Jersey: Prentice Hall.

Popham, W. J. (1973). Evaluating Instruction. Englewood Cliffs, New Jersey: Prentice Hall.

Rentz, R. R. (1980). Rules of thumb for estimating reliability coefficients using generalizability theory. Educational and Psychological Measurement. 40, 575-592.

Safrit, M. J. (1977). Criterion-referenced measurement: Applications in physical education. Motor Skills: Theory into Practice, 2(1), 21-35.

Safrit, M. J. (1981). Evaluation in physical education (2nd ed.). New Jersey: Prentice Hall.

Shepard, Lorrie. (1980). Standard setting issues and methods. Applied Psychological Measurement, 4(4), 447-467.

Simpson, E. J. (1966). The classification of educational objectives, psychomotor domain. (Vocational and Technical Education Grant, Contract No. OE 5-85-104). Office of Education, United States Department of Health, Education and Welfare.

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a critierion-referenced test. Journal of Educational Measurement, 13, 265-276.

Suttinger, J. (1957). A proposed predictive index of volleyball playing ability for college women. Unpublished study, University of California.

Swaminathan, H., Hambleton, R. K., & Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 11, 263-268.

Thorndike, R. L. (Ed.). (1971). Educational measurement (2nd ed.). Washington, DC: American Council on Education.

Williams, L. R. T., & Fawcett, S. A. (1975). A multiple regression analysis of volleyball skill tests. The Australian Journal for Health, Physical Education and Recreation.

Appendix A

VOLLEYBALL PROFICIENCY TEST

PART I - Cognitive Knowledge

Multiple Choice - Pick the best answer for each question and fill it in the appropriate space on the answer sheet. Do not write on the question sheets.

1. Which serve is used most frequently by highly skilled players?

   a) spike serve
   b) overhand spin serve
   c) underhand float
   d) overhand float serve

2. From the following group of errors, which will prevent your serve from floating?

   a) wrist too stiff on contact
   b) wrist too loose on contact
   c) contact is off-centre
   d) b and c
   e) a and c

3. Where is the major weakness of the W serve-receive formation?

   a) the centre front area of the court
   b) the sideline
   c) the centre back area of the court
   d) the corners of the court

4. Which technique is usually the most effective when passing the ball to a spiker?

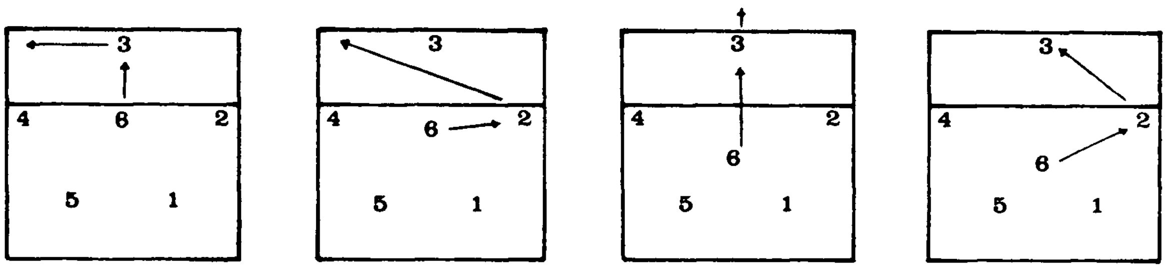
   a) a jump set
   b) an overhead set
   c) a bump
   d) a back bump

5. Which pattern represents the most basic offense in volleyball?

   a) set-spike-block
   b) pass-set-attack
   c) set-spike-cover
   d) pass-set-tip

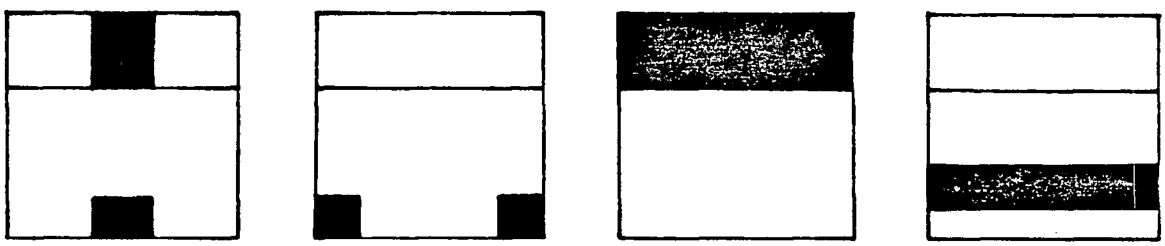
6. Which skill does a player have the most control over?
   a) setting
   b) serving
   c) bumping
   d) blocking

7. Which diagram indicates the best strategy when the centre back makes the first contact in a 4-2 centre specialized system?



a) b) c) d)

8. Which darkened area shows the <u>best</u> region for serve placement?



a) b) c) d)

9. Which part of the players body gives the best surface for contact and control of the bump pass?

   a) the fleshy part of the inner arms
   b) the wrists
   c) the forearms
   d) b and c

10. Which situation illustrates an illegal play?

   a) a player reaches under the net to play a ball falling from the net on her side
   b) a player reaches over the net to block a ball that has been attacked by the opponents
   c) a player steps over the centre line during play but does not interfere with the opponent's play
   d) a player grabs the shirt of a teammate and pulls her back to prevent her from falling into the net

11. Which score indicates a completed game?

    a) 15-14
    b) 22-20
    c) 11-10
    d) 11-9

12. How many points has Team A scored if they have rotated through the following serving order with Jackie being the first server of the game: Jackie served 3 times, Blaine served 4 times, Mike served 1 time, Roy served 2 times, Gerry is ready to serve for the fourth time?

    a) 9
    b) 8
    c) 13
    d) 14

13. If a blocking player can only reach so that half of her hands extend above the height of the net, she should:

    a) take a one-step approach and reach with one hand
    b) keep trying until her jump improves
    c) not block; stay at the net and turn to face her teammates to be ready to make the second contact
    d) soft block so that the ball will deflect up into her back court

14. What is the <u>key</u> to landing safely after making an emergency dig?

    a) using your knee pads
    b) rolling
    c) relaxing when you contact the floor
    d) diving

15. A 6-up defense works <u>best</u>:

    a) against a team that tips or hits half-speed shots
    b) against a team that hits deep over the block
    c) for a team with an inconsistent 2 man block
    d) for a team with poor tip diggers

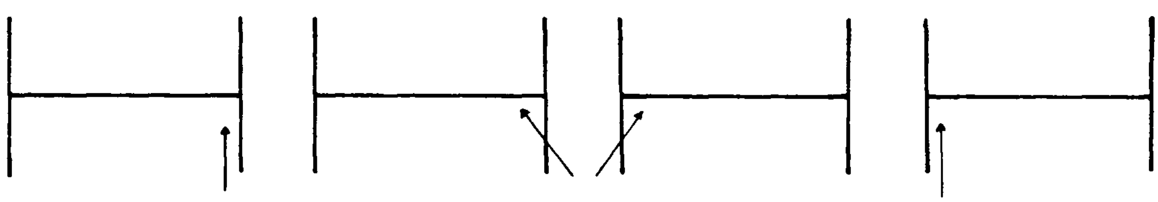
16. In a 6-up defense, whose responsibility is it to tip dig in centre front when the other team is attacking from their position #4.

    a) #3
    b) #4
    c) #6
    d) #4 and #6

17. What is #5's responsibility when her teammate #4 is attacking the ball?

    a) covering at mid-court in case the ball is blocked
    b) switching to her defensive position
    c) covering just inside the 3 meter line in case the ball is blocked
    d) watching to see where the holes in the opponents' defenses are

18. What is the best angle of approach for a right handed spiker from their power side?



a) b) c) d)

19. What is the fastest way to move across the width of the court?

    a) sidestep
    b) forward sprint
    c) stutter step
    d) cross-over step

20. The sport of volleyball was initiated in:

    a) Japan
    b) Czechoslovakia
    c) Cuba
    d) U.S.A.

21. Which hitter is executing an off-hand spike?

a) a right-handed hitter spiking from LF
b) a left-handed hitter spiking from RF
c) a right-handed hitter spiking from CF
d) a left-handed hitter spiking from LF

22. Which serve has no spin and moves in an erratic path as it approaches the receiver?

a) overhand-hit with heel of hand - no follow through
b) sidearm-hit with open hand
c) spike serve
d) overhand-hit with heel of hand - follow through

23. Where is the contact point for the overhand pass?

a) at chin level
b) directly overhead
c) right off the nose
d) near the forehead

24. How are the legs positioned when executing a forearm pass?

a) front-back stride, knees bent
b) side stride, knees bent
c) front-back stride, legs fairly straight
d) side stride, legs fairly straight

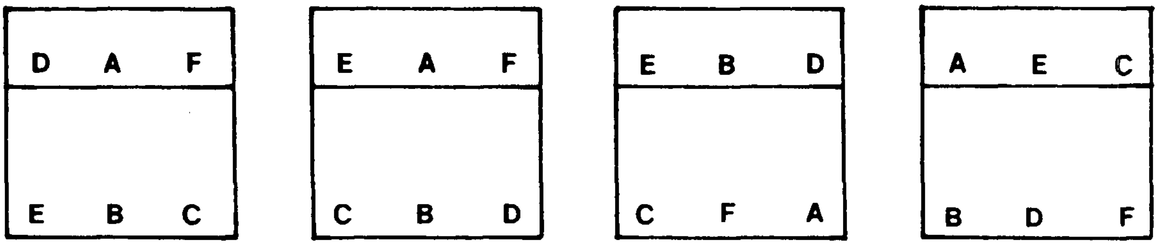
25. Which of the following increases the power of a spike?

a) contacting the ball in front of the body
b) follow through with the hand
c) speeding up the arm action
d) rotating the hips laterally after take-off

26. Which technique is recommended for successful spiking?

a) one foot take-off, cupped hand
b) two foot take-off, cupped hand
c) two foot hop, open hand
d) two foot take-off, open hand

27. Players A and B are setters. Players C and D are the best hitters. Which line-up is most advantageous?



a) b) c) d)

28. What are the responsibilities of the '2' in a 6-2 offensive system?

a) hitting and defense
b) digging and setting
c) hitting and blocking
d) setting and hitting

29. A player consistently spikes the ball into the net. Taking for granted that the sets are adequate, which correction should be offered?

a) hit the ball a little later
b) take a longer approach
c) hit the ball sooner
d) decrease the follow through of the arm

30. Which of the skills listed below utilizes primarily the legs to increase the distance that the ball travels?

a) serve
b) dive
c) set
d) dig

31. Who controls the offense?

a) captain
b) setter
c) hitters
d) coach

32. The attack area shall

a) be 3 metres from and parallel to the centre line
b) end at the sidelines of the court
c) extend indefinitely parallel to the center line
d) a and b
e) a and c

33. The heights of the nets for men and women respectively at the centre of the net shall be

a) 2.24 m; 2.49 m
b) 2.43 m; 2.00 m
c) 2.43 m; 2.24 m
d) 2.49 m; 2.24 m

34. A team is permitted

a) four substitutions per game
b) six substitutions per game
c) six substitutions per match
d) 12 substitutions per match

35. Any player beginning a game in a match may be replaced

a) once by any substitute and may not re-enter the same game
b) once and may re-enter the same game once
c) twice during the game provided the same player exchanges with him
d) at the beginning of the next game but not before

36. After a ball is served

a) each player may move to any section of his team's court
b) the backline players only may switch positions in the backline
c) the front line players only may switch positions in the front row
d) both b and c

37. The linesmen are responsible for

a) signalling balls 'in or out' of court
b) checking the height of the net before the match begins
c) indicating if a ball has been contacted by a player before landing outside the court
d) both a and b
e) both a and c

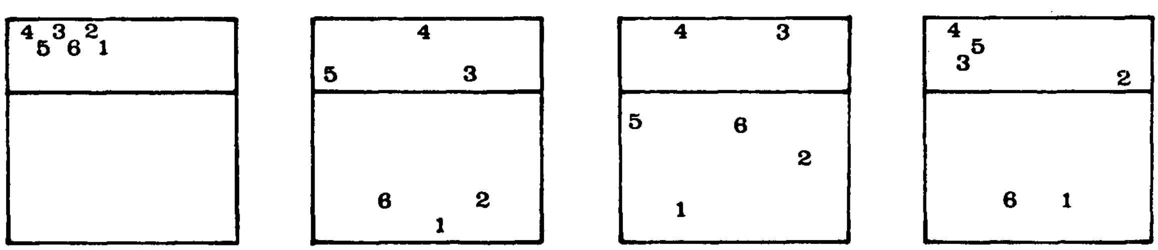
38. A third time out for rest is requested; what happens?

a) time out is granted but the captain or coach making the request shall be warned
b) time out is granted but the opponents receive a point
c) it shall be refused and the opponents receive a point
d) it shall be refused, and the captain or coach making the request shall be warned

39. A simultaneous hit by opponents allows the team on whose side the ball enters the court

a) three more hits
b) two more hits
c) one more hit
d) a replay

40. Pick out the serve receive pattern that constiutes an overlap.



a) b) c) d)

e) none of the above

Appendix B

VOLLEYBALL PROFICIENCY TEST

PART II - Performance Analysis

Pick the best answer for each question and fill it in the appropriate space on the answer sheet.

1. Why can't I get the ball to go farther forward?

   a) you are not using your legs
   b) you are not extending your arms on contact
   c) you are not contacting the ball above your forehead
   d) a and b
   e) b and c

2. I can't seem to control how far forward I bump the ball - why?

   a) arms bent on contact
   b) transferring your weight backwards on contact
   c) contacting the ball with your fists
   d) a and b
   e) b and c

3. Why do I always have to jump when I forearm pass the ball?

   a) you are too close to the ball
   b) one leg is too far in front of the other
   c) you are not using enough arm swing on contact
   d) a and b
   e) a and c

4. Why does the ball fall short of my target?

   a) because you are off balance prior to contact
   b) because you have no forward lean in your trunk
   c) because your arms are parallel to the floor
   d) b and c
   e) a and b

5. How can I spike the ball cross-court?

   a) jump sooner and reach for the ball
   b) approach the ball at a 45 degree angle to the net
   c) point your left arm and shoulder to the ball on takeoff
   d) a and b
   e) b and c

6. I'm having trouble with my timing - what's wrong?

   a) you are approaching from too far away
   b) you are not using enough arm swing
   c) you are taking too many steps
   d) a and b
   e) b and c

7. Why do I always seem to hit the ball behind my head?

   a) you are drifting under the ball after your two-foot takeoff
   b) your arm swing is too late
   c) you are taking off too close to the net
   d) a and c
   e) a and b

8. Why does my serve go so high over the net?

   a) toss is too low
   b) toss is too close to your body
   c) hitting the bottom of the ball
   d) a and b
   e) b and c

9. My serve seems to keep hitting the top of the net, why?

   a) toss is too low
   b) elbow is bent on contact
   c) toss is too far in front of you
   d) b and c
   e) all of the above

10. How can I jump higher on my spike jump?

   a) use a forceful upward armswing
   b) use a heel-toe rocking action to plant on takeoff
   c) feet should be perpendicular to the net on takeoff
   d) a and b
   e) all of the above

11. Why are all of my spikes going out of the court?

    a) because you are jumping too late
    b) because your elbow is bent on contact
    c) because you are taking off too close to the net
    d) a and b
    e) a and c

12. Why can't I get the ball to go farther?

    a) not using your legs
    b) not contacting ball above forehead
    c) not following through in the direction you want the ball to go
    d) a and b
    e) all of the above

13. Why can't I control where the ball goes?

    a) contacting it with your fists
    b) one leg is too far in front of the other
    c) arms are too parallel to the floor
    d) a and b
    e) a and c

14. Why do I have a hard time controlling where the ball will go?

    a) your steps to the ball are too long
    b) you are contacting the ball at chin level
    c) your fingers and hands are too relaxed on contact
    d) b and c
    e) all of the above

15. Why does the ball go straight up instead of forward?

    a) you are backing away from the ball on contact
    b) the ball is hitting your fists
    c) your follow through is up and over your head
    d) b and c
    e) all of the above

16. In a 6-up defensive system such as the players are using who has the responsibility to dig this ball?

    a) #2 after landing from block
    b) #6
    c) #1
    d) #1 or #6

17. Why was this player unable to spike the ball?

    a) set was too high
    b) set was too far outside the antenna
    c) attacker did not back out of the court in preparation for the set
    d) b and c

18. The player in position #1 has the second contact on the ball - who would be the best player for him to set?

    a) #4
    b) #2
    c) #3
    d) #2 or #4

19. How could the setter in center front have made a better set that was closer to the net?

    a) by using a jump set
    b) by using more leg extension
    c) by turning her body parallel to the net while setting the ball
    d) by turning her body parallel to the net before setting the ball

20. In a W serve receive pattern who has the responsibility to receive this ball in deep center back?

    a) either the left or right back depending on who can get there faster
    b) the player in the center position - she should back up if the ball is going deep
    c) the player in left back because both the left front and center player turned to show him it was his ball
    d) any of the 3 backrow players depending on who called the ball first

Appendix C

SUBJECTIVE RATING SCALE

A) OVERHEAD PASSING

4 - Excellent - demonstrates ease of movement with control and accuracy.

- the player positions his or herself properly in relation to the oncoming ball; feet, hips, and shoulders face the target and the player neither has to reach nor feel constricted as they play the ball.
- body should be balanced on ball contact with one foot slightly in front of the other.
- there is a smooth transfer of weight and momentum from the legs to the arms and forward into the ball.
- ball is contacted above forehead and arms follow through upward in direction of pass.
- fingers are firm and contact is legal.

3 - Average to Good - generally has control over the ball but one component of the pass is performed incorrectly so the fluidity of movement is missing.

- player may have judged incorrectly and finds his or herself too close or too far away from the ball but is still able to adapt and perform an overhead pass successfully.

- body may not be square to the target but otherwise the overhead pass is performed smoothly.
- feet may be parallel rather than one in front of the other therefore providing a very small base of support and loss of balance while performing the pass.
- transfer of momentum from legs to arms may not be sequential.
- body movement prior to contact is correct but on contact body weight is transferred backwards.
- body position is correct but hands are dropped below forehead level or are too far back above head.
- footwork and body movement are smooth but hands are not kept firm enough for legal contact

2 - Poor to Average - performance is inconsistent due to a combination of two errors.

- movement to the ball is inadequate and therefore the player contacts the ball in an unbalanced position - forward follow through is still evident.
- player is unbalanced prior to contact and follows through in a backward motion.
- player contacts ball at chin level and follows through backwards with body.
- transfer of momentum from legs to arms is not sequential and follow through of arms is directly upward instead of forward.

1 - Poor - demonstrates erratic body control and thus consistency and accuracy are not evident.

- a combination of three or more of the previous errors mentioned would result in poor performance.

B) FOREARM PASSING

4 - Excellent - demonstrates ease of movement with control and accuracy.

- the player positions his or herself properly in relation to the oncoming ball; feet, hips, and shoulders face the target and the player neither has to reach nor feel constricted when they play the ball.
- body should be balanced on ball contact with one foot slightly in front of the other.
- there is a smooth transfer of weight and momentum from the legs to the arms and forward into the ball.
- arms are straight on contact and almost parallel to the floor.
- very little upward follow through occurs after contact.
- the ball should be contacted on the forearm area two to four inches above the wrist.
- the ball is contacted simultaneously with both arms.

3 - Average to Good - generally has control over the ball but one component of the pass is performed

incorrectly so the fluidity of movement is missing.

- player may have judged incorrectly and finds his or herself too close or too far away from the ball but is still able to adapt and perform a forearm pass successfully.
- body may not be square to the target but otherwise the forearm pass is performed smoothly.
- feet may be parallel rather than one in front of the other therefore providing a very small base of support and loss of balance while performing the pass.
- transfer of momentum from legs to arms may not be sequential.
- body movement prior to contact is correct but on contact body weight is transferred backwards.
- body position is correct but ball contacts arms closer to elbows than to wrists.
- footwork and body movement are smooth but there is an exaggerated upward armswing on follow through.

2 - <u>Poor to Average</u> - performance is inconsistent due to a combination of two errors.

- judgement and movement to the ball are inadequate and therefore the player contacts the ball too high on the forearm - forward follow through is still evident.
- player is unbalanced prior to contact and follows through in a backward motion.

- player contacts the ball too high on the forearm and follows through backwards with the body.
- transfer of momentum from legs to arms is not sequential and player uses exaggerated armswing to get the power to lift the ball.

1 - Poor - demonstrates erratic body control and thus consistency and accuracy are not evident.

- a combination of three or more of the previous errors mentioned would result in poor performance.

C) OVERHAND SERVING

4 - Excellent - demonstrates ease of movement with control and accuracy.

- ball is tossed with a controlled lifting action of the arm.
- the height of the toss is just slightly higher than the extended hitting arm.
- a smooth forward transfer of weight occurs just prior to ball contact - this can be from back foot to front foot or from heels to toes.
- ball is contacted in front of or directly above hitting shoulder.
- arm is extended and wrist is stiff on contact for an overhand floater serve. If a topspin serve is attempted the wrist is snapped over the ball on contact.

- there is very little extraneous movment in either backswing or forward swing of arm action.
- the arm comes forward quickly but follow through is limited.

3 - <u>Average to Good</u> - generally has control over the ball but one component of the serve is performed incorrectly so the fluidity of movement is missing.

- toss may be too low or high or too close to body but server adapts and performs a successful serve.
- action of the tossing arm may lack control but the remainder of the serving action is smooth.
- arm action may be very smooth but not accompanied by any forward weight transfer.
- although toss is accurate, arm action may be slow causing contact with a bent elbow.
- body control may be smooth but wrist is loose on contact.

2 - <u>Poor to Average</u> - performance is inconsistent due to a combination of two errors.

- toss is too close to body so server adapts by bending elbow and wrist to contact ball.
- tossing arm action lacks control so forward transfer of weight is not evident.
- tossing arm action lacks control so ball is not contacted

directly in front of hitting shoulder.

1 - Poor - demonstrates erratic body control and thus consistency and accuracy are not evident.

- a combination of three or more of the previous errors mentioned would result in poor performance.

D) SPIKING

4 - Excellent - demonstrates ease of movement with control and accuracy.

- player positions his or herself outside the court at a 45 degree angle to the net in preparation for the toss.
- player takes a short step to help in timing the approach.
- from this initial step a long, low, forceful step is taken landing in a two-foot takeoff position with toes facing 45 degrees to the net.
- the arms are brought back behind the attacker as the step is taken.
- as the heel-toe rocking action of the takeoff occurs, the arms are forcefully swung forward and upward to aid in vertical lift.
- the hitting elbow is pulled back high in preparation for the attack.
- strong trunk rotation and flexion precede the forward arm action of the hitting arm.

- the ball should be contacted in front of the hitting shoulder with the arm extended.
- judgement of the set is critical so that ball is attacked at highest point of the jump.
- arm should follow through across the body.
- balance should be regained on landing with knees flexed and feet shoulder width apart.

3 - <u>Average to Good</u> - generally demonstrates good body control and ball contact but an error in one component of the spike makes the attack less effective than it could be.

- proper technique is demonstrated but the timing of the jump is too early or too late.
- proper form is demonstrated after takeoff but he angle of approach to the net is incorrect <u>or</u> there is not evidence of a long, low step prior to takeoff.
- the approach is performed correctly but elbow is bent on contact <u>or</u> ball is contacted behind the hitting shoulder.
- footwork is correct but no forceful upward armswing is evident on takeoff.
- footwork and contact point are correct but there is little or no trunk rotation or flexion preceding ball contact.

2 - <u>Poor to Average</u> - performance is inconsistent due to a

combination of two errors.

- no long, low step is evidenced prior to takeoff and no upward armswing is used.
- angle of approach is incorrect and ball is not contacted directly in front of hitting shoulder.
- feet do not takeoff simultaneously and jump is either too early or too late.

1 - Poor - demonstrates erratic body control and thus consistency and accuracy are not evident.

- a combination of three or more of the previous errors mentioned would result in poor performance.

Appendix D

JUDGE - ____________________

SUBJECT NAME - ________________
VOLLEYBALL LEVEL - ____________

VOLLEYBALL RATING SCALE

TRIALS

	1	2	3	4	5	6	7	8	9	10	TOTAL
A) OVERHEAD PASSING											
4 - Excellent											
3 - Average to Good											
2 - Poor to Average											
1 - Poor											
B) FOREARM PASSING											
4 - Excellent											
3 - Average to Good											
2 - Poor to Average											
1 - Poor											
C) OVERHAND SERVING											
4 - Excellent											
3 - Average to Good											
2 - Poor to Average											
1 - Poor											
D) SPIKING											
4 - Excellent											
3 - Average to Good											
2 - Poor to Average											
1 - Poor											

Appendix E

CHI SQUARE TABLES

Cognitive Scores

SKILL	M 1	NM 2	TOTAL
E 1	5	11	16 33.3
I 2	6	10	16 33.3
N 3		16	16 33.3
COLUMN TOTAL	11 22.9	37 77.1	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
7.31204	2	0.0258	3.667	3 OF 6 (50.0%)

NUMBER OF MISSING OBSERVATIONS = 0

Performance Analysis Scores

SKILL	M 1	NM 2	ROW TOTAL
E 1	1	15	16 33.3
I 2		16	16 33.3
N 3		16	16 33.3
COLUMN TOTAL	1 2.1	47 97.9	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
2.04255	2	0.3601	0.333	3 OF 6 (50.0%)

NUMBER OF MISSING OBSERVATIONS = 0

Overhead Pass Product Score

SKILL	M 1	NM 2	ROW TOTAL
E 1	16		16 33.3
I 2	14	2	16 33.3
N 3	9	7	16 33.3
COLUMN TOTAL	39 81.3	9 18.8	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
10.66667	2	0.0048	3.000	3 OF 6 (50.0%)

NUMBER OF MISSING OBSERVATIONS = 0

Forearm Pass Product Score

SKILL	M 1	NM 2	ROW TOTAL
E 1	16		16 33.3
I 2	7	9	16 33.3
N 3	2	14	16 33.3
COLUMN TOTAL	25 52.1	23 47.9	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
25.21043	2	0.0000	7.667	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Overhand Serve Product Score

SKILL		M 1	NM 2	ROW TOTAL
E	1	9	7	16 33.3
I	2	7	9	16 33.3
N	3	5	11	16 33.3
COLUMN TOTAL		21 43.8	27 56.3	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
2.03175	2	0.3621	7.000	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Spike Product Score

SKILL		M 1	NM 2	ROW TOTAL
E	1	14	2	16 33.3
I	2	5	11	16 33.3
N	3	5	11	16 33.3
COLUMN TOTAL		24 50.0	24 50.0	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
13.50000	2	0.0012	8.000	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Overhead Pass Process Score

SKILL	M 1	NM 2	ROW TOTAL
E 1	13	3	16 33.3
I 2	3	13	16 33.3
N 3		16	16 33.3
COLUMN TOTAL	16 33.3	32 66.7	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
26.06249	2	0.0000	5.333	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Forearm Pass Process Score

SKILL	M 1	NM 2	ROW TOTAL
E 1	16		16 33.3
I 2	1	15	16 33.3
N 3	1	15	16 33.3
COLUMN TOTAL	18 37.5	30 62.5	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
39.99999	2	0.0000	6.000	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Overhand Serve Process Score

SKILL		M 1	NM 2	ROW TOTAL
E	1	15	1	16 33.3
I	2	4	12	16 33.3
N	3	4	12	16 33.3
COLUMN TOTAL		23 47.9	25 52.1	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
20.20174	2	0.0000	7.667	NONE

NUMBER OF MISSING OBSERVATIONS = 0

Spike Process Score

SKILL		M 1	NM 2	ROW TOTAL
E	1	8	8	16 33.3
I	2	2	14	16 33.3
N	3		16	16 33.3
COLUMN TOTAL		10 20.8	38 79.2	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
13.13684	2	0.0014	3.333	3 OF 6 (50.0%)

NUMBER OF MISSING OBSERVATIONS = 0

Total Score

SKILL		M 1	NM 2	ROW TOTAL
E	1	11	5	16 33.3
I	2		16	16 33.3
N	3		16	16 33.3
COLUMN TOTAL		11 22.9	37 77.1	48 100.0

CHI-SQUARE	D.F.	SIGNIFICANCE	MIN E.F.	CELLS WITH E.F.< 5
28.54054	2	0.0000	3.667	3 OF 6 (50.0%)

NUMBER OF MISSING OBSERVATIONS = 0