

ITEM ANALYSIS AND MEASUREMENT CHARACTERISTICS OF KNOWLEDGE'S TEST IN GOLF

UDC:796.352

(Original scientific paper)

Reny Damjanova

*National Sports Academy "Vassil Levski", Sofia, Coaches Faculty,
Department – Theory of Sports Training, Sofia, Bulgaria,*

Abstract:

The purpose of this research was to establish and create standards for Selecting and Evaluating Standardized knowledge test currently used in Bulgaria for controlling and evaluating the knowledge in the field of the Etiquette and the Rules of Golf. Groups trained their knowledge were Bachelor students in the National Sports Academy, Sofia, Bulgaria who have chosen Golf as a Subject by choice in the period 2001-2006. The Golf Lesson course with those students has total 30 academic hours (2 ECTS). The first half of them are theoretical and practice sessions over the Etiquette and Rules of Golf and the second part are - physical practice and lessons with a golf-professional over the basics of the golf swing. Testing has been done by the end of the Lesson course with a test let with 3 different scales. The first scale includes 20 "truth –fouls" questions, the second scale includes 20 multiple-choice questions and the last scale includes 10 open questions. Questions in the scales are evaluated with different weight. Total time for testing - 90 min. The sample in total is n = 147 tests. Measurement characteristics of the tests used for evaluation of the knowledge of the Golf Rules have been established with methods and statistical operations for criterion-referenced knowledge tests. Those characteristics are – indexes and coefficients for reliability and validity of the tests, as well characteristics of each included type question (open, truth-fouls or multiple-choice). Item analysis and assessment of the measurement characteristics of the standardized criterion-referenced test for students showed high coefficients for validity and reliability. From this point of view the created knowledge's test could be applied for different purposes as: evaluation of knowledge's of golf referees or golf players or as a first step in the creativity process of and Computer -aided System for education (self-educational) in the field of the Etiquette and the Rules and Regulations of Golf.

Keywords: *students, reliability, validity, scales of measurements, statistical, procedures*

INTRODUCTION

After the sociopolitical changes in 1989/90 in Bulgaria there were some preconditions for new sports to start developing together with the sports that were traditional (like- gymnastics, rowing, basketball, track-and-field athletics, free-style wrestling, weight lifting etc.). One of these sports

was golf. One of the first and basic steps for the development of new sports is to represent to the audience with the world wide accepted sport rules and regulations respectively with the official accepted from USGA& R&A St.Andrews – rules and regulations and golf etiquette. The very realization of this massive educational project in

Bulgaria started right after the IV congress "Future of Golf", that was held in ASU, Tempe, AZ in 2001. This project was initiated by two recognized national institutions – The Bulgarian Golf Association and the National Sports Academy "Vasil Levski". The Bulgarian Golf Association is associate member of the R&A St.Andrews, European Golf Association (EGA) and the International Golf Federation (IGF). The BGA is the only Bulgarian institution holding an official license for administrating golf as a sport in Bulgaria and as well having the official license for translating and distributing the Rules and Regulations of Golf in Bulgaria. The National Sports Academy is educational institution with traditions and officially recognized statute of university where

Bulgarian students are trained in the field of sport in three degrees- Bachelor, Master and Doctor. In 2001 in the National Sports Academy "V. Levski" started a new program Golf which was optional subject (SIP). From 2001 to 2007, 240 students chose that golf course, 170 of them passed the practice and theory exam (Damjanova (Дамянова), 2004a, 2004b, 2004c, 2004d).

The Goal of the research that was held was to check the methodology of measurement of the characteristics of the test questions in criteria referenced tests in relation to the building of standards for selecting and evaluating of the standardized tests.

Creation of specialized test battery for checking of the theoretical knowledge of golf in compliance with the official international standards, etiquette and rules.

Setting of the measurement characteristics of the test battery and the different types of questions included in it. (Validity, standard for being successful, reliability, difficulty and etc.).

Creation of standard for a choice of a mark of the criterion tests and fare marking of the knowledge in the field of Etiquette and The Rules and Regulations of Golf through different variants of tests intended for referees, coaches, instructors or golf players(amateurs or professionals).

The research was held in two stages – preliminary research and central experiment.

METHODS

In the first model of the test, designed for students of Golf, we tried a combined variant with two

scales. In the first scale comprised of 20 questions with multiple choice questions, there were included questions of general theory and history of the play. In the second scale, comprised of 30 questions with open answers regarding terminology, etiquette and the Rules and Regulations of Golf. The test should be completed in 60 minutes. The students making the test have no right to use any textbooks or other materials while completing the two parts of the test. With the first d variant of the test were examined 47 students (n = 47).

The second variant of the test was comprised of 3 scales. The first scale included 20 "truth –false" questions, the second scale includes 20 multiple-choice questions and the last scale includes 10 open questions. Questions in the scales are evaluated with different weights. Total time for testing - 90 min. 10-15 minutes break was made between the second and the third part of the test. The tested people have no right to use any materials while filling in the first two parts of the tests. Filling in the third part they can use all possible information sources (textbooks, notes, Rules and Regulations of Golf and etc.) The sample in total is n = 100.

METHODS

In order to distinguish the measurement characteristics of each question in a test we used J. Flanagan method, 1939. This practical method gives the possibility by using 13 steps to calculate the indexes of the discrimination and the difficulty of each question also to research by the use of a table the correlation coefficient (that shows the degree of validity). At the same time with the help of the frequency and semantics analysis it was possible to decide the qualities of every destructor in every test question. The total validity of the whole test can be calculated as average to the validity of all the questions. This practical method for distinguishing the validity of the classical educational tests shortens a lot the time for administration of the tests and in this way makes it economical. What is more the using of the method gives the chance to improve the tests in compliance with the goals of the tested people- referees, coaches, instructors, or players: (Nelson, 1960); Wilcox, (1977).

The second analysis of the data of the testing was connected to the formulating of the standards for success („cut off") in the criterion referenced tests which are very suitable for urgent information regarding the degree of learning of a specific

material from the program- a chapter or unit during the course of study. In this case we were focused again on the empirical models, that does not require very complicated and time consuming calculations. For this purpose we used the method of the contrast groups suggested by Livingstoun & Zieki, (1982).

RESULTS AND DISCUSSION

The data from the results of the first test (n=47) gave us the opportunity to mark the measurement characteristics of the whole test and of each question as well as the indexes of the discrimination of the difficulty and the function of the destructors (in multiple choice questions) etc. As a whole this test has good distinguishing power (mean of the index of discrimination $D = 0.72$), very high difficulty (index of difficulty- $d = 9\%$) and good reliability ($\alpha = 0.798$). In order to improve the measurement characteristics of the test and the questions a thorough analysis was made and all the questions with poor measurement characteristics were changed before being included in the second final variant of the test.

All questions that had the following characteristics were revised:

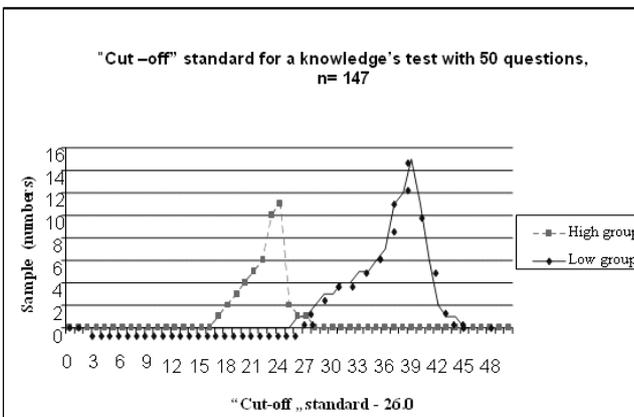


Fig 1. "Cut-off" point

Low discrimination index, because this index shows as well the low validity in relation to the program for which measurement is assigned to measure this question. In principal every wrong answer is expected to have negative correlation and the true answers –positive. In other words it is expected that within the weak group of tested students more of them will chose wrong answers;

Very high or very low index of difficulty as it is evident that if that value is high the question is

easier. The questions with index above 90 % - are very easy and those below 10 % -too difficult. The optimal difficulty of the whole test must be average between 50 - 60 %.

Wrong answer (destructors) which are not chosen at least from 2 or 3 % of all tested people. Such answers are useless and does not help in distinguishing the successful and unsuccessful participants in the test. And it should be replaced with more attractive variant of answer.

After the main experiment was held (n = 100) then again were calculated and analyzed the measurement characteristics of the whole test and the questions included in it. As a whole they were very much improved. The reliability of the whole test was increased to $\alpha = 0.838$, the mean discrimination – to $D = 0.802$, and the difficulty of the test became $d = 59\%$.

The methodology gave is the opportunity to join the results of the two kind of tests of the examined people on golf by dividing them in two contrast groups (weak and strong) with a measure of capacity $n_1 = 47$ and $n_2 = 100$. The frequency distribution of the results of the two groups shows that in order to be successful the test there should be 26.5 solved problems (Fig. 1).

In accordance with the standard of success received in this way the people from the "success " group will be categorized as "not achieved the goal" and the people from the "Low " group that have more than 26 solved problems (total 2) will be categorized as accomplishing the goal. In other words with this standard of success 3 out of 147 tested people will be marked wrongly (this means 2.04 %). In this case the validity of the test results will be equal to 0.98, as $1 - 0.02 = 0.98$.

The analysis of the data of the two tests (the preliminary and the central) as well as the data from the enquiry made through the tested people show that the first variant of the test is too difficult for the students. The reasons for that are in the administration of the test and in the fact that golf is one of the new sports in the country and respectively its terminology, the Etiquette and Rules and Regulations should be distributed among the fast growing golf society in Bulgaria. This caused changes in the whole administration of the test (the right to use information materials, the duration of the test, etc.), as well revision of some of the questions included. The second variant of the test has as a whole very good measurement

characteristics, which shows us that the chosen methodology for researching the measurement characteristics gives for practical and easy optimization and standardization of the criteria referenced knowledge's tests.

CONCLUSIONS

The procedure used by us and the analysis of the data of the two experiments permits to develop a great number of equivalent in its measurement characteristics standardized forms of tests intend to estimate the knowledge in the domain of golf of wide spectrum of target groups – referees, trainers, instructors and players with different statute (amateurs, professionals, beginners and advanced, different age groups – Gentlemen, Ladies, Children, etc.)

As the Etiquette and Rules of Golf, together with Golf Decisions are a bank of questions themselves the methodology can be easily applied to the development of interactive systems of education and study as well with implanted on them subsystems for generating different scenarios from tests for objective and standardized rating and self marking of the knowledge in that domain.

The expected effect from the application of such system will increase not only the special knowledge of the players and their skills, but will as well make it easy for the management of the golf clubs in relation to safety of play and planning lower expenses for the golf course maintenance.

REFERENCES

Дамянова, Р. (2004а). *Ръководство за студентите от СИП голф* [Manual for students from SIP golf. In Bulgarian].

София: Национална спортна академия.

Дамянова, Р. (2004б). *Английско-български терминологичен речник по голф* [English-Bulgarian Dictionary of golf terms. In Bulgarian]. София: Национална спортна академия.

Дамянова, Р. (2004с). *Етикет и Правилник по голф (2004-2007)*. [Etiquete and Rules of golf (2004-2007). In Bulgarian]. София: Национална спортна академия.

Дамянова, Р. (2004д). *Европейска хандикап система (2004-2006)* [European Handicap system (2004-2006). In Bulgarian]. София: Национална спортна академия.

Klauer, K. J. (1987) *Kriteriumsorientierte Tests*.

Verl. Psychologie. Toronto, Zuerich: Dr. J. Hogrefe, Goettingen, Toronto, Zuerich: Livingstone, S. A. & Zieky, M. J. (1982).

Passing score. A manual for setting standards of performance on educational and occupational tests. New Jersey: Educational Test Service

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3 – 19.

Novick, M. R. & Lindley, D.V..(1978) The use of more realistic utility functions in educational applications. *Journal of Educational Measurement*, 15,181 – 191.

Poggio, J. P., Glassnap, D.R. & Erros, D.S. (1981) An empirical investigation of the Angoff, Ebel and Nedelsky standard setting methods. *Paper presented at the Annual meeting of the American Educational Research Association, Los Angeles*.

Popham, W. J.(1978). *Settings Performance Standards*. Los Angeles: Instructional Objective Exchange (b).

Popham, W. J.(1978). *Criterion-referenced Measurement*. Englewood Cliffs. New Jersey: Prentice - Hall (a).

Popham, W. J. & Husek, T.R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, (1), 1- 9.

Stoimenova, E. (1995). On the LFC natural classification rule under a new loss function. *Statistics & Decisions*, 13(1), 39-51,

Vijin, P. & Molenaar, I. W. (1981). Robustness region for dichotomous decisions. *Journal of Educational Measurement*, (6), 205 - 235.

Vygotsky, L. S. (1978). *Mind in society; the development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press,

Waglow, I. F. & Rehling, C. H. (1953). A Golf Knowledge Test. *Research Quarterly for Exercise and Sport*, 24, 463 – 470.

Wilcox, R. R. (1977) Estimating the likelihood of false-positive and false –negative decisions in mastery testing: An empirical Bayes approach. *Journal of Educational Statistics*, (2), 289 - 307(b).

АНАЛИЗА НА ПРЕДМЕТОТ И МЕРНИТЕ КАРАКТЕРИСТИКИ НА ТЕСТ ЗА ПРОЦЕНУВАЊЕ НА ЗНАЕЊАТА ВО ГОЛФОТ

УДК:796.352

(Оригинален научен труд)

Рени Дамјанова

*Национална спортистка академија “Васил Левски”, Софија,
Тренерски факултет, Катедра Теорија на спортиста,
Софија, Бугарија*

Апстракт:

Целта на истражувањето беше да се утврди и да се придонесе за создавање на стандарди за избор и проценка на тестовите кои моментно се користаат во Бугарија за контрола и оценување на знаењата во областа на Етикетата и Правилникот за голф. Беше истражувана групата на студенти од Националната спортистка академија кои како изборен предмет го слушале во периодот 2001-2006 година. Курсот по голф за тие студенти содржи 30 часови (2 ЕКТС). Првата половина од тие часови претставуваат теоретски и практични вежби за етикетата и правилникот за голф, а втората, практични часови кои ги одржува професионалец за голф. Истражувањето е реализирано на крајот од курсот, при што е користен конструиран тест, составен од 3 различни скали. Првата скала имаше 20 прашања од типот “точно-неточно”, втората имаше 20 прашања со алтернативни одговори, а третата беше скала-вена од 10 прашања. Прашањата во секоја скала беа оценувани со различна тежинаска вредност. Примерокот вкупно изнесуваше N=147. Мерните карактеристики на тестовите, се искористени за оценување на знаењата на Правилникот за голф. Тие беа утврдени со методи и статистички постапки, насочени според критериуми за проценување на знаења. При тоа, пресметани се индекси и коефициенти на релјабилност и валидност. Од таа гледна точка, конструираните тестови може да бидат применливи за различни цели (оценување на знаењата за судиите и играчите во голфот, појдовна основа за создавање на компјутерски систем за обучување и самообучување во областа на етикетата и правилникот за голф слично.

Клучни зборови: студенти, релјабилност, валидност, скали на мерња, статистички постапки