THEORETICAL AND METHODOLOGICAL ASPECTS OF THE TEST AS AN INSTRUMENT FOR MEASUREMENT AND EVALUATION OF KNOWLEDGE AND SKILLS

Katia Stoyanova[1]

[1]*Sofia University "St. Kliment Ohridski" Department for Information and In-Service Training of Teachers, 1619 Sofia, Mail: katiastoyanova@abv.bg*

ABSTRACT An important and essential question, related to the test as an instrument for measurement and evaluation of knowledge and skill of the tested persons is discussed in the paper. tThe items directly concerned with the theme of the material, namely: measurement, estimation, test and didactical test are clarified. Concrete examples are used in order to attain more clarity and profoundness with regard to their specifics and application. The paper stresses on evaluation, various types of assessment, tests, and more specifically didactic test. The following types of assessments are considered: selective, "in situ", diagnostic, the final evaluation, and their specifics are clarified. After clarifying the measurement and assessment, the attention is focused on the test. Various definitions of its meaning are presented. Short historical knowledge of the origin and development of the didactical test in our country and abroad, including Europe, is also introduced. This issue is discussed in more detail about Bulgaria.

Key words: measurement, estimation, test, didactical test

ТЕОРЕТИКО-МЕТОДИЧЕСКИ АСПЕКТИ НА ТЕСТА КАТО ИНСТРУМЕНТ ЗА ИЗМЕРВАНЕ И ОЦЕНКА НА ЗНАНИЯ И УМЕНИЯ

Катя Стоянова[1]

[1]*Софийски университет „Св. Климент Охридски" Департамент за информация и усъвършенстване на учители", 1619 София, ел.поща: katiastoyanova@abv.bg*

РЕЗЮМЕ. В настоящия материал се разглежда един значим и важен въпрос, свързан с теста като инструмент за измерване и оценка на знанията и уменията на тестираните лица. В тази връзка се изяснява същността на понятия, пряко свързани с темата на разработката, а именно: измерване, оценка, тест, дидактически тест. При изясняването им са използвани и конкретни примери с оглед по-голяма яснота и пълнота относно тяхната специфика и приложение. В материала се акцентира върху оценяването, видовете оценки, тест, в частност-дидактически тест. От различните видове оценки са разгледани: селективна, текуща диагностична, заключителна и е изяснена тяхната специфика. След изясняване на измерването и оценяването, вниманието се фокусира върху теста. Представени са различни дефиниции на това понятие. Представят се кратки исторически сведения за възникването и развитието на дидактическия тест у нас и в други страни, както в Европа, така и извън нея. По-подробно този въпрос е разгледан за България.
Ключови думи: измерване, оценяване, тест, дидактически тест

Introduction

For better clarification of the test, at first we would deal with the items that have direct connection with that measuring instrument.

Measurement

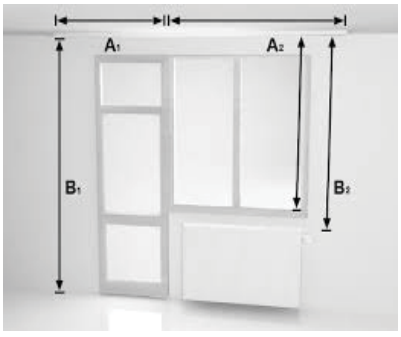

Fig.1. An example of a measuring representation

Before making an evaluation of concrete knowledge and skills, first we need to measure them. In that connection we will present a definition of measurement, which is more common and more frequently cited (S. Stevens). *The measurement is a conformation between the objects and figures, when for every object a digit is disposed by an appropriate rule.* Depending on what and how is measured, we can attach various numbers to an object. For instance, when measuring a square – m^2, dm^2 (Stoyanova, 1996). This means that in measurement the big estimation has not only the result *(digit)*, but also the measurement unit *(measure)* that is used. The various measuring units lead to various results. When we are measuring, we measure *the characteristics of the object,* not the object alone (an unique characteristics: for example, knowledge, if the object is the student).

The measurement can be direct or indirect. In direct measurement, the items that is measured, is measured directly: length, temperature, etc. For instance, direct measurement is when we measure with a tape-measuring device the length and the width of a room, which has a rectangular shape. If a

quadrature or cubature of the room must be found, then the measurement is indirect. This is because at first we are measuring directly the length, width and height of the room, then by the corresponding formula we obtain its quadrature and cubature (Stoyanova, 1996).

Indirect measurement refers to the measuring of other items, which are related to the measured object. By the result of their measurement, the appropriate conclusions are made (for instance – quadrature, cubature, as above).

The higher the reliability of the instrument by which we are measuring is, the more precise and exact the measurement would be. In this moment it is necessary to point out what exactly is a reliability of the measurement. This is one of the criteria for quality of the measurement. More often the reliability represents the *exactitude, trustworthiness, correctness, precision, certainty*. Remember how important is the measurement, as we connect it with the proverb „Measure twice, cut once!".

Except reliability, another characteristic of the measurement is validity, which is a relative concept. Usually by validity we understand the degree of correspondence between the goals of the measurement and the goals for which this instrument were made (Stoyanova, 1996).

Estimation

Fig.2. The role of evaluation

Essence of evaluation

The problem with the evaluation in the school will always be an one of interest and significance. The evaluation is a necessity, that not only has a stimulating function, but provides a possibility to make important decisions for refinement of the educational process.

Another widely known definition of evaluation is related to comparing the results from the measurement with previously defined standards (criteria). At first, we need something to measure and then to evaluate. In other words, the evaluation is a process of comparing the real achievements of the students with previously defined requirements (criteria) (Visokalijska, 2015).

Assessment

The assessment in the school is expressed by a digit. In Bulgaria the six-grade system of evaluation is still operating. Actually, the evaluation of knowledge and skills is expressed in digits. This evaluation is a means for achieving a definite goal and for decision making. There is no fully objective assessment, because there are many factors which are relevant. For instance, the standards that are picked up, subjects that are making the evaluation, etc. The adequacy of the evaluation depends a lot on the standards chosen (criteria) with which we compare the results from the measurement.

It is worth to say the following: *The measurement is made with the goal to obtain the necessary information, but the evaluation of that information is needed to make adequate decisions.*

Types of assessments

In this paper we present some basic assessments, based on various indications.

Depending on the goals of decision-making and the steps of the evaluation in the education process, the assessments are:

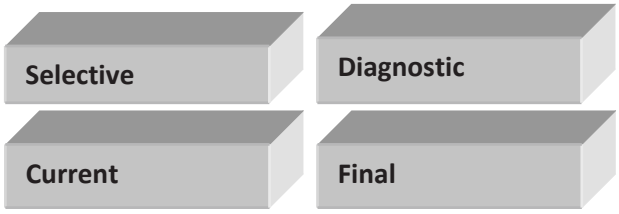

Fig.3. Types of assessments according to the goal of decision making and stage of assessments

Selective assessment

It is used when it is necessary to make decisions for the purpose of future development (training) of the pupil (child, student), i.e. estimation of his/her potential skills and preliminary training. It may be made on the basis of *criteria approach* (readiness, level of foreign language knowledge, etc.) and *normative approach* (when choosing a candidate among many others).

Current assessment

It is directed at the achievements of the students. It accompanies the process of studying and provides the feedback for its quality. It is based more often on the criteria approach, i.e. – how often the goals are achieved.

Diagnostic assessment

With this type of assessment, the goal is to establish the reasons for attaining the concrete learning goal; for instance – systematic lag of the students, etc.

Final evaluation

As with the current assessment, its goal is to establish to what extent the concrete educational goals have been achieved, but at the end of the school year, course, educational degree. It is also based on the criteria approach (Stoyanova, 1996).

Test

Fig.4. Test solution

The test is a multipurpose concept, which means a *testing, probing, testing, inquiring, method* in the pedagogic, psychology, medicine, sport, etc. According to the English dictionary, "test" means testing or probing something, in order to define its quality, value, composition, etc. In *the education, test* means measuring of the knowledge, skill, intelligence or possibilities of an individual or group of individuals (Tujarov,2009).

The term test is also used in a broad and narrow sense. The broad meaning includes probe, test, testing, as well as all other means required for this purpose.

In narrow sense, test is "…short, technical simple testing, whose solution may be attained by a quality approach and is used as an indicator for the stage in the development of a specific measured subject.

According to Ebel the test is a common term, which is used in every project or procedure for measurement of the possibilities, interests and other characteristics (Siderova, 2008)

L.R.Guy defines the test as measurement of the knowledge, skills, feelings, intelligence or skillfulness of a person or a group (Siderova, 2008). According to A. Reber – test (the more common word is procedure) is used for the measurement of a specific factor or for evaluation of a skill (Siderova, 2008).

One of the main characteristics of didactical tests, which make them suitable for pedagogical practice is the higher objectiveness as a method for measurement and evaluation of the pupil's achievements (Siderova, 2008).

On this basis it may be said, that the didactical test is a method for measurement of the pupils' achievements (the learning ones) according to the common and specific goals of a defined educational topic, or a defined methodical unit. Contrary to the traditional control and classworks, the test is made according to strictly defined methodical item.

The didactical tests measure the results related to the acquisition of certain knowledge, included in the specific scholar activity and aimed at certain goals and problems. In the specialised psihology-pedagogical literature, Gentcho Piryov defines as pedagogical texts the ones that measure the results of education by means of adopted knowledge and skill (Siderova, 2008).

One definition of a didactical text, adopted by most of the specialists in the area of testing, is that didactical text is *an instrument for measuring and evaluating the results from the learning and educational process (LEP)* according to the defined goal and objective (Bijkov, 1992).

The specifics of the didactical test will be presented in this article.

A short historical overview of didactical tests

In the specialized literature there is no uniform opinion regarding the history and development of the didactical tests. Some of the authors believe that tests could have existed four thousand years ago. The opinion of other authors is that we may speak about tests as of the beginning of XX century, more exactly from 1920s. Why are these discrepancies? They result from the different interpretations of the essence of the test. For instance, if we mean that the test is an instrument for checking and evaluation of some properties, peculiarities, or skills of the peoples, the origin of the tests may be traced even to the antiquity.

There is information in the specialized literature that the didactical test originated in the second half of XIX century aiming at overcoming some of the weaknesses of the traditional system for checking and assessing the knowledge and skills of the students. After the Second World War, the application of tests increased in the Western countries, , not only for various research in a number of areas, but also in the educational practice itself.

The creator of the first standard didactical test is the English teacher Fisher in 1864, but no attention was paid to his product because of the slowdown of scientific thought at that time. The beginning of the movement for objective didactical tests was made by the American Rike in the same year.

Nowadays, didactical tests, which are refined more and more, are widely used in education.

Short information about the creation and implementation of tests in Europe (Byjkov, 1982) is presented hereinafter.

In Germany – the creation and application of tests began in XX century and was connected with the experimental pedagogy of W. Lay & E. Moiman.

In Sweden the implementation of tests is regulated in the educational documentation, self-made and standard tests are applied, mainly in the area of diagnostic (for scholar maturity, on native and foreign language, when choosing the school, etc.).

In Belgium there are strong traditions connected with the creation and use of tests, mainly with the pedagogic consulting and directing of tests, related mainly with the pedagogic consulting and directing of the students.

In Switzerland, tests are applied in school activities (psychological and didactic) and in the professional orientation of students as well when choosing profession.

In Japan from the 1920s till the beginning of the WW2, various tests were developed, which were standardized. There is broad application of tests in the professional appointment of

the students, entry exams in the higher degree and at national level for evaluation of the knowledge and skills of students.

In Poland, the studies in the area of tests are provided mainly by the Research Institute for pedagogic studies at the Ministry of Education.

In Bulgaria the creation and implementation of didactic tests does not have great traditions. More merits in this area belong to prof. D. Katzarov and his assistant and later – to the professor and corresponding member of the Bulgarian Academy of Science – Gentcho Piryov.

Papers and developments of didactical tests started to be published in the 1970s. The so called "first white swallow" was developed by prof. Encho Gerganov, who elaborated the first Bulgarian didactical test in Bulgarian language for the IVth classes of the UMPS (unified middle polytechnical school).
In the 1980s some researchers from the Central Institute for In-Service Training of Teachers and Leaders "Vela Blagoeva", now DIUU - Department for Information and In-Service Training of Teachers, started the development and application of didactic texts aiming at supporting the qualification of Bulgarian teachers.

Later, in 1989 and 1990, various tests were developed by the group of Scientific and Research Laboratory (NIL) on pedagogic diagnostic. The evaluation and diagnostic activities were provided on the results of apprehension of new school content in the I-VIII classes, which are not standardized and because of that are not widely applied by the teachers in the of education process.

Today tests are an intrinsic part in the process of education. They are used in various phases of this process and most of them are automatized.

Internal structure of a didactical test

Every didactic test has three main structural elements:

1) Information about the person tested (exit data, instructions, direction for a mode of operation, etc.). This information aims to relax the person tested and direct him to the proper and unique solution of the defined task.
2) Question or/and task (formulated as a question and/or task, related to the information which is checked; it is possible to make an instruction what exactly is asked from the person tested, i.e. –prescriptions are made what exactly to be done).
3) Response (depending on the kind of a question and/or the task, the tested person is required to formulate his/her own answer (to reach the goal of the task, the result that is searched/ or to pick up the correct answer from the given options) (Bijkov, 1992).

Theories of tests

First, let's focus on the meaning of test theories. This is a defined system of judgments for the relations and dependencies between the phenomena and processes measured by tests .
In the specialized literature there are two main theories for tests (psychological and didactical ones), which are based upon:

- The dependence of theoretic (ideal) value of the measurement by testing and the error of that measurement;
- The dependence of the approaches by which this relation and dependence are defined (Bijkov, 1992).

The two theories are:

- Classic (normative);
- Probability (stochastic).

It may be said, that the two theories are interconnected and are not opposed, irrespectively of their difference (Siderova, 2008).

Main characteristics of didactical tests

In the measurement we should mention reliability and validity. In this context every didactic test, which is composed according to certain requirements, must possess the following base characteristics: objectivity, reliability, validity, comparability, cost-efficiency, relevance, etc.

A) Objectivity (*value of independence from the measurement*)
It is not correct the person creating the test, to evaluate himself. It should be clarified that so far the teachers have been developing tests in the subject they teach. The same tests are applied to confirm the level of knowledge and skills acquired by the students at a certain level of education (usually for a given time). Outer assessment has been introduced, but it is in the "output" of the whole process of education of student (for instance, in the state exam on theory and practice in the profession and the specialty). A test may be wholly objective if various researchers reach the same results in relation to the person given (Bijkov, 1992). We may conclude that the objectivity of a test has a great importance for its quality.

B) Reliability – this is the main criteria for quality; this means exactness, precision of the measurement. This is one of the most important characteristics of the test, because it is directly related to the error from the measurement. The smaller the error, the more reliable the test is (Stoyanova, 1996).

C) Validity (correspondence with the goals of measurement and goals for which the instrument is created) - the validity of a test shows the degree of exactness with which it is measuring the object in question. For that purpose the results obtained are compared with the previously defined external criteria, which may be a normative requirement, scholar program, results of an investigation, etc.

In the specialized literature on testology various types of validity are known, such as content, criteria and construct.

- The content validity shows the degree to which a test is checking the school content.
- The criteria validity shows the extent to which the test corresponds to the concrete goal (predictive validity). This correspondence is defined through the method of expert evaluation, or comparing the pupils' achievements with their previous ones.

- The construct validity permits to show which are the defining factors on the basis of the test results obtained (Tujarov, 2009).

D) Comparability (comparing the results); shows that the results of a given test, obtained in various places, may be compared and on that basis valuable conclusions with regard to content to be made. Thus, adequate control decisions can be made (Tujarov, 2009).

E) Cost-efficiency (it is related mainly to the possibility to make a test for testing many students for a short period of time). It permits mass examinations with the participation of many people. On basis of the tests' results it is possible to take important decisions on various levels (school, regional, national). Recently, the newest information and communication technologies have made possible the en masse testing of knowledge and possibilities of scholars and students, as well as other persons. This may be achieved any time, when such test is necessary. In testing on line, there are no expenditures on multiplication and copying of tests and that leads to economy of this resource (Tujarov, 2009).

F) Relevancy of the test.

This characteristic of the test is directly related with the definition of goals. The more strictly and properly defined the tests are, the higher the relevance of the test is. The test must be constructed in such a way, so as to measure the necessary school results, and that will be happen when the goals are clearly and exactly defined (Tujarov, 2009).

Although this list is not exhaustive of the test characteristics, these are the main ones which must be taken into account when constructing tests for knowledge and skills.

Classification of didactic tests

In the specialized literature, directly connected with didactic tests, there isn't unique classification because of various points of view and various criteria for differentiation. The more common types can be represented as follows:

- Standardized (made by professionals on the basis of existing theory and fixed procedures;);
- Non-standardized (made by teachers for direct application in the area of one school; lower objectivity – the person who created the tests, will evaluate them);
- Normative (they establish the individual status of the person tested compared with the achievements of others, which are used as etalon, norm for comparison);
- Criteria (they measure and evaluate the achievement of students according to the goals and tasks of UVP, the criteria defined by the state educational documents (state educational requirements (standards) (SED), scholar programs, etc.);
- Overall tests for achievements (measure achievement of a student for a certain time);
- Diagnostic tests (establish the errors and cause of their origin);
- Preliminary /sorting/ (they are introduced in the beginning of education);
- Processing /formatting/ (they are applied after apprehension of part of the material – section, theme, methodic unit; no evaluation is made, but the errors in the knowledge and skill are checked);
- Final (for "final" at the end of the term, end of the school year, at the end of the educational degree; to check the achievement of the goals and what the efficiency of the scholar-pedagogic process is (SPP);
- Tests for diagnostic of practice skills (check a degree of apprehension of basic skills: details, models, drafts, etc. are manufactured).

In the educational process diagnostic tests are often applied. The teacher uses them to test the level of adoption of basic knowledge and skills in the definite stage of education in the object(s) given. Thus, the necessary adequate measures can be taken for overcoming errors and ignorance of students, as well to establish this ignorance at the moment of testing of students.

Advantages of didactic tests

Didactic tests have various advantages and in the following text we will present some of them, which in our opinion are more essential, namely:

- Cost-effiency (for short time are checked many people); objectivity, reliability, validity, permits individual and group checking; a degree of achievement may be obtained; norm of achievement may be exposed; computer-aided check of the results may be performed;
- There are strictly defined requirements for objectivity, reliability and validity;
- They may be used in a more prolonged time;
- When standardized, *they give more objective assessment and clear presentation of the achievement of students,* compared with the traditional methods of checking the educational process (Iliev).

Shortcoming of didactical tests

- They do not present the causes for one or another result; they do not enlist every personal characteristic of students, but only part of it; there is an element of randomly checking the answers; they have short prognostic life;
- The tests, which are not standardized and are used often in the practice, are not objective and reliable and in such a way, they cannot measure with accuracy the achievement of the students (tested persons).

Conclusion

In the material presented, the inherence of measurement, evaluation and on that basis – the didactical test are explained. Basic questions concerning the specifics of tests and more precisely, the specifics of didactical tests, are clarified.

References

Бижков, Г. Теория и методика на дидактическите тестове. С., Просвета, 1992. -20-46 с. (Bijkov, G. Teoriya i metodika na didakticheskite testove. Sofia, Prosveta, 1992. -20-46 p.)

Бижков, Г., Ф. Стоянова. Теория и методика на

дидактическите тестове. С., Просвета, 1996. -36 с.. (Bijkov, G., F. Stoyanova. Teoriya i metodika na didakticheskite testove. Sofia, Prosveta, 1996. -36 p.)

Илиев, А., ПГ „Проф. д-р Асен Златаров", Свищов. Дидактическите тестове - съвременен метод за измерване и оценяване на постиженията на учениците поматематика. (Iliev, A., PG "Prof. d-r Asen Zlatarov", Svishtov. Didakticheskite testove – savremenen metod za izmervane i otsenjavane na postizheniata na uchenitsite po matematika)
http://www.diuu.bg/ispisanie/broi16/16kt/16kt6.htm

Сидерова, И. Дидактически тестове – В: Педагогически факултет, ВТУ, „Св. Св. Кирил и Методий", 2008. (Siderova, I., Didakticheski testove – B: Pedagogicheski fakultet, VTU „ Sv. Sv. Kiril i Metodii", 2008.)

Стоянова, Ф. Тестология за учители. С., Атика, 1996,

Тужаров, Х. Тестология, *2009.* (Stoyanova, F. Testologiya za uchiteli, S., Atika, 1996,

Tujarov, H. Testology,
2009.).http://tuj.asenevtsi.com/EL09/EL49.htm

Високалийска, М. Проект „Обучение на педагогически специалисти за формиране на знания, умения и компетентности за оценяване на учениците",2015. (Visokaliiska, M. Proekt Öbuchenie na pedagogicheski spetsialisti za formirane na znaniya, umeniya i kompetentnosti za otsenyavane na uchenitsite", 2015)

http://prezi.com/ze_do5cnd0k4/presentation/

The article is reviewed by Assoc. Prof. Dr. Dz. Dzelev and Assist. Prof. M. Legurska.