

Docimological Principles Applied To Standardized Tests

Șișianu Ala

senior lecturer, PhD student, Technical University of Moldova

ala_at@mail.ru

Abstract - This article contains a brief introduction to the main principles which should be followed by the constructors of tests and assessments. It briefly introduces the key concepts of test validity, reliability and wash back, and provides information on docimological test history, researches done in the described field, which have shown a multitude of factors that appear purposely or unpurposefully in evaluation, generating a low objectivity in estimating the results. The work reveals the docimological principles as those needed to be used in analyzing the results and that the algorithm used while doing this does not mean subjectivity and absence of precision. There is a possibility not to sanction a very good paper for few common faults generated by the tiredness, carelessness. The methods provide student's accuracy, attention and predict that repeated wrong answer at common questions brings to the examinee a score unsatisfying his/her expectations. These factors should encourage students in using their gained learning potential and fight against cheating.

Index terms - evaluation, docimology, learning, student, results, test.

1. INTRODUCTION

One doesn't need to look very far to see how important testing and assessment have become in educational system.

Testing is more than accountability. It can be a means to improve education, itself. Standardized tests and large-scale assessments can be used, and are being used, to encourage teaching of the skills prescribed by state and local agencies. A critical component of instruction, various forms of teacher assessment permeate everyday classroom activity. Written tests provide formal feedback with regard to what has and has not been learned. The routine asking of questions and the scoring of projects and activities in the classroom are other forms of assessment that strike at the heart of instruction. Teachers' need for information is commensurate with the pace of their instructional decision making, which is probably more intense than in any other profession.

As the nation searches for ways to improve student achievement, educators and policy makers continue to evaluate and reform their education systems. Educational testing, or assessment, is a key component of all education systems. Assessments can be used in schools to monitor educational systems for public accountability; help improve curricula; evaluate the effectiveness of teaching and instructional practices; measure student achievement; and determine a student's mastery of skills. Education leaders need to think deeply about the role of standardized testing in higher schools. While many people agree that some system of responsibility is essential to maintain school quality, there is disagreement

about the role of high-stakes testing in that system. Standardized tests are designed to enable us to compare the performance of students in a relatively efficient way. But how much can tests tell us about what students actually know?

Although educational testing is a complex field, there are several basic principles that provide a foundation for further understanding.

A problematic aspect of the educational process either developed to the formal or non-formal level is represented by the evaluation of the students, a feature also highlighted in the case of e-learning. By evaluation one can obtain important information, concerning the learning results (stocked knowledge, abilities, skills, etc.). This information has a double role [7, p. 395]:

1. It confirms or infirms the expected results both by the ones who projected the learning sequence and by the pupil (child, teenager or adult);
2. It fixes the future development of the process (by an authentic feed-back effect).

Besides the evaluation, the permanent feature of the education also imposes the self-evaluation. Self-evaluation is defined by C. Stan as being "the pupil's capacity to emit and elaborate valuable appreciations concerning the own competences and performances, extended to the own person in general". To form this capacity it is necessary to elude the subjectivism in grading, a phenomenon with most negative consequences both on learning and on the involved actor's personalities.

2. ONE CENTURY OF DOCIMOLOGY

The beginning of the 20th century represent the debut of the first scientific researches in the field of

school evaluation, initiated by the H. Pierson, who names this preoccupation docimology. The term has Greek roots: "dokime" which means trial, test, and "logos" which means science, so docimology means the science of tests, exams. Exam is a form of social evaluation, by which one realizes a brief evaluation, of the end of learning period (the BAC marks, the end of the high school, the university studies end with the university degree exam, the final

exam after attending a course), and by graduating it one can obtain a diploma, which allows the possessor to occupy a "social role" [10, p. 140].

The psychologist Vasile Pavelcu considered the entire period of our existence as a succession of exams, which marks the ending of certain steps in the life of the individual [8]. Due to their social importance, these exams are criticized severely, being reproached the absence of some proper evaluation instruments and the strong subjectivism in grading.

The researches which have been done have shown a multitude of factors that appear purposely or unpurposely in evaluation, generating a low objectivity in estimating the results. These factors can be grouped in several categories, being reported to [9, pp. 23-27]:

1. Teacher: the "halo" effect, the "kind" effect, the generosity error, the Pygmalion effect, the "contamination" effect, the "contrast" effect, the examiner's personal equation, the error of central tendency, the logical error, the effect of Gauss curve, the teacher's personality factors;

2. The subject referred to: the papers in subject such as Physics, Math, Chemistry can be evaluating more objectively than the papers in subjects such as Philosophy, Literature, etc.

3. Pupil: personality particularities – temperament, aptitudes;

4. The social circumstances in which the evaluation is performed: the leaders, the mates, the parents' interventions on the examiner for a certain pupil, the tolerated deviation of cheating.

5. The authors experience allows us to stand that this negative factors are not enough known by the staff in general, but especially to ones in technical superior education. The phenomenon of the tolerate deviation of cheating, by the very mechanisms it manifests, is one of the most harmful and hard to analyze and to discourage.

6. In these conditions, the validity of the obtained information is doubtful. The quality of this information strictly depends on the objectivity of the evaluation process and on the quality of the used instruments.

The testing process will be objective if:

7. The process of applying the test will be objective - the same task given to all the students under the same condition;

8. The results will be objectively amazed by imposing a straight criterion of evaluation or a sample correct answer, subjectivity being in this way reduced to minimum;

9. The results will be objectively interpreted, meaning that the same performances are evaluated and marked in the same way by different examiners.

One of the main directions the contemporary docimology sustains concerning the objectivity of the learning evaluation is using the docimological tests, founded on the base of docimological principle. The docimological principles are "fundamental theses, general rules with descriptive and normative character, which base the evaluation project, organizing and development in order to ensure their scientific consistency and efficiency" [9, p.85].

The most relevant docimological principles, which lead evaluation activity, are:

1. The principle of evaluation objective character, refers to the structure and organization of evaluation, so that the pupils' performances to be reflected and evaluated in a real and relevant manner, limiting as much as possible the influence of external factors;

2. The principle of evaluation interactive character expresses the fact that learning evaluation are inherently connected to the and determined both by the evaluation made by the teacher and by the pupil's self-evaluation activity;

3. The principle of the pupil's performances contextualization: regards the fact that during the evaluation there have to be considered the performances and there have to be used such tasks that could reflect the reality, meaning to attach the pupil's capacity to adapt the knowledge to various situations.

3. THE DOCIMOLOGICAL TEST

As principal methods of objective evaluation, the docimological test is " a set of questions with the help of which one can check and evaluate the knowledge and the capacities acquiring to operate with them, by reporting the answers to a sample appreciation scale, previously elaborated" [7, p.401]. In specific literature we will also encountered other terms: pedagogical test, knowledge test, learning evaluation test, performance test or simple test to designate the instrument and method of evaluation which has a specific element - the item, being characterized by a greater objectivity in evaluating the results. The quality of information offered by the testing depends on two sets of attributes the test have to possess:

1. Psycho - pedagogical: the test must be appropriate to its specific purposes and comprehensive;

2. Statistical: which guarantee the perfection of the test as a measurement instrument; the most important being the accuracy and validity.

The accuracy of the test, also called constancy or exactness designates the trust we can have in the respective instrument, the degree of exactness of measurement. The principal condition a test has to have in order to possess this quality refers to the stability of the results:

1. when one pupil is being examined by different teachers, who do not know the previous results of the examinee;

2. when the conditions in which the testing is made are modified. Because of this a test applied in different objective and subjective conditions has to reach almost identical results. The most important objective conditions are: the position of the desk in the classroom and the examinee in the desk, the lighting in the classroom, the weather, the atmospheric pressure, the degree of the classroom ventilation, the moment the exams is being taken, the influences coming from the mates, and so on. I can also mention some of subjective conditions: degree of tiredness of the examinee, his/her previous experience, the social importance of the exam, the wish to pass the exam, the moral and the experience, the parents' pressure, the degree of nervousness;

3. even the appearance of the items is changed: the grammatical form, the addressing manner, the replacement of one word by its synonym, the changing in the items order;

4. in time: applied successively, the test must give the same results; if not, it is not an accurate one.

The accurate coefficient can be calculated as follows [4, p. 78-80]:

1. repetition the applying of the test at a certain period replicating the test must not record measurement deviation, or, at least, the error must be precisely anticipated and measured; the constancy coefficient of the results will show the accurate degree of the test, by the correlation between the values obtained at two distinct and outdistanced application in time.

2. comparing the results obtained by its application with the result obtained at other equivalent tests. The qualitative and quantitative correlation shows the equivalent coefficient. So, we will establish that tests are alike or are distinguished from each other;

3. by halving, considering the even items score with odd items score, the degree of correlation obtained represents the homogeneity coefficient.

The fidelity of the test depends on the difficulty of the items. When the test contains items with high difficulty, the individual resorts to a guessing the correct answer. The greater the number of guessed answers, the more the scores distribution takes a binomial form. Therefore, we could stand that between the difficulty of the items and the quality of the test there is a reversely proportional rapport. If the test contains items with low difficulty the individuals seldom resort to guessing, and the distribution of the results is uniform. Therefore, the

test cannot be useful to evaluate all the individuals. From the theoretical point of view, the longer the test is (it contained a greater number of items), the more accuracy it has [1, p. 205].

The validity represents the most important quality of the test; show whether the instrument measures what it proposes and how well makes that.

In establishing the validity of a test there are asked two questions:

1. does the test measure what it is meant to?
2. can it be used in taken the right decisions?

Concerning the purpose, validity can be:

1. of content: the test must refer to those contents referred to during the instruction
2. of criteria: involving the rapport to an external criterion

The first operation made after applying the test is to correct and marking the answers. There are two points of view in marking the answers at the items of the brief tests [1, p. 231]:

1. The first point of view is founded on the concept of "errorless activity", used in preparing the specialists in different fields of techniques. According to this point of view, in the professional activity (programming, building plans, extraterrestrial flights) there are no options between minor and major errors. The computer program works or not. Therefore, in evaluating the specialist preparing, it is not considered the indices of difficulty of the items, but only the fact that they are solved right or wrong, omitted or partly solved is not considered.

2. The second point of view is founded on the thesis that the marking of the item must reflect its index of difficulty (the relative frequency of the individuals who have answered correctly of the all examinees who answered to that item). Thus, when marking, the teacher distributes the pupils in class of results, each of them corresponding a marking coefficient (e.g. the mark "7" will be given both to the students who have acquired 37 points and to the ones who have acquired 40). So, when giving marks, the teacher levels the scores of the pupils pertaining to a class results.

4. CASE STUDY

After several months of application of these docimological principles and tests to the students of different specialties at Technical University of Moldova, I had the opportunity to study all those elements in practice. It has been demonstrated that a correct application of docimological principles can bring innovations at the level of whole coaching and final evaluating process. Because the main pursued goal is to eliminate the cheating, for this we must try to change the mentality, both the professors and students.

The contradiction consists in fact that nobody from the professors assumes the responsibility to improve

his style of work, against the changing using the reason of previous successful generations of students. In this context, are accepted only the comparisons between the different generation of students.

Besides, it can be imposed them from the exterior what they have to do. The model of the perfect professor do not exist in the reality. The students benefit by the whole spectrum of the university values, both on the scientific plan and didactic, in the same degree like in the past. What is different and must be corrected is the inertness of the students. On the found of tolerate deviation and low socio-economic standard, the students' involvement must be forced to rise to the level imposed by his\her future profession. The instrument destined to realize this correction is just the subject of our discussion: the university degree exam.

The university degree exam is used both in the various goals and levels of engineering preparation. The justification of this kind of exams is important, but the fact that such a brief exam can be useful for the final correction in university preparation. The final correction means:

1. Covering the minimum knowledge required by the practicing of that profession, in the instruction process;
2. Assignment of a relevancy for this kind of exams by praising the deserving students and clear establishing the score of the promoting.
3. These two principles assure the individual's motivation. It is well known that motivation (positive or negative) is one of the aspects of the success. The docimological criteria of the university degree exam have the role of stimulating the motivation mechanisms.

Way of work:

- Translation of the individual responsibilities towards the collective system of the university degree exam
- Principle for scaling results of the evaluation : Gauss distribution situated between n% repel and m% maximum rating
- Method of work: questionnaire focus on the engineering specialties, on extreme complexity :
- INFERIOR Extreme
- SUPERIOR Extreme

The indicated way of work means the fixing of the precise level in the area of the obligatory minimum engineering knowledge (physics phenomena, measures, calculus relations, etc.).

The accomplishment of the students' evaluating can become efficiently if the following aspects characteristic to the docimological tests are charged:

Specific aspects:

- % Variable weight for each answer

- % Vague algorithms for the results interpretation
- Answers with positive bonus
- Answers with penalty bonus, for the fault case
- Variable and particularized tolerance
 - % Clear definition for the admissibility threshold
 - % Questionnaire with a supplementary number of item (300 – 350)
 - % Informatics system for generating the questionnaires
 - % Eluding the psychological obstacles:
- free access for the self-evaluation

The variable weight of the answers is necessary because it is working both with very simple and very complicate knowledge.

The vague algorithm used in the analyzing of the results does not mean subjectivity and absence of precision. There is a possibility to not sanction a very good paper for few common faults generated by the tiredness, carelessness; but repeated wrong answer at common questions brings to the examinee a score unsatisfying.

To fight against cheating, the number of the items must be in excess. The number of the good answers being given, will be also a relative criterion. In order to avoid the inherent suspicious about the secret of elaborating the questionnaires, these are realized in the morning of the day exam, by random generating. The eliminating of psychological obstacle can be realized by students' very well, knowing of the process of evaluation. In this way can be realized, also a self-instruction, this being the final goal of the whole learning process.

5. CONCLUSIONS:

The using of the performing system of final evaluation can contribute to the improvement of the whole educational process.

Strategic results:

- The graduation mark comes into prominence
- The graduation exam becomes an incentive event

The signify of the exams degree mark means not only a supplementary outdistanced between the examinee, but also the possibility to eliminate the candidate which risk to be in the situation of professional imposition, whether they will receive a diploma which they don't deserve.

Even the opinions about cheating are disputed, using of this testing system, which realized a real competitor frame, is the one that can change the mentalities.

6. REFERENCES:

- [1] Cabac, V., 1999, Evaluarea prin teste în învățământ, Ed. Universității "Al. Russo" Bălți
- [2] Holban, I., 1995, Testele de cunoștințe, EDP, București
- [3] Ionescu, M., 2001, Didactica modernă, Ed. Dacia, Cluj Napoca
- [4] Jinga, I. (coord.), 1999, Evaluarea performanțelor școlare, Ed. Aldin, București
- [5] Moise, C., Evaluarea școlară, note de curs
- [6] Muster, D., 1970, Verificarea progresului școlar prin teste docimologice, EDP, București
- [7] Nicola, I., 2000, Tratat de pedagogie școlară, Ed. Aramis, București
- [8] Pavelcu, V., 1968, Principii de docimologie, EDP, București
- [9] Stan, C., 2001, Autoevaluarea și evaluarea didactică, Ed. Presa Universitară Clujeană, Cluj Napoca
- [10] Strugă, C., 1999, Evaluarea școlară, Ed. de Vest, Timișoara
- [11] Bonciu, Stefan , 1997, Copiatul la examene ca devianță tolerată, în vol. Câmpul universitar și actorii săi, sub coord. Adrian Neculau, Editura POLIROM , Iași
- [12] Niculescu, Rodica Mariana, 2000, Formarea formatorilor, Editura ALL EDUCATIONAL, București