FUNDAMENTAL ASPECTS OF EVALUATION IN THE CONTEXT OF ENHANCING THE QUALITY OF THE EDUCATIONAL ACT

Lucia-Larissa PALEA
"Nicolae Bălcescu" Land Forces Academy, Sibiu, Romania
luciapalea@yahoo.com

Simona BOŞTINĂ-BRATU
"Nicolae Bălcescu" Land Forces Academy, Sibiu, Romania
mbostinabratu@yahoo.fr

Alina NEGOESCU
"Nicolae Bălcescu" Land Forces Academy, Sibiu, Romania
alina.negoescu@yahoo.com

ABSTRACT
Evaluation is a very complex process, which involves a continuous effort to reduce subjectivity from the evaluation procedure in giving grades, scores, marks, percentages etc. The assessment of the instructional process is a controversial topic due to the human factor involved in this process of valorization of learning outcomes. The issue in question breeds discussions on morality, since the future of the testees depends on the facts of examination and of the factually proved objectivity. Evaluation is therefore a pedagogical concept that transcends the definition of measure and the appreciation of the students' performance. It refers to the process of determining and assessing learning outcomes by revealing the contribution of each resource in the teaching-learning process. The purpose of this paper is to point out some of the major elements of the evaluation process in the context of enhancing the quality of the educational act.

KEYWORDS: evaluation, objectivity, evaluation errors, grid tests

1. Introduction

The assessment of the instructional process is a controversial topic due to the human factor involved in this process of valorization of learning outcomes. The issue in question breeds discussions on morality, since the future of the testees depends on the facts of examination and of the factually proved objectivity. Evaluation is therefore a pedagogical concept that transcends the definition of measure and the appreciation of the students' performance. It refers to the process of determining and assessing learning outcomes by revealing the contribution of each resource in the training process, of the modalities of improvement, of the outcome-objective relationship etc.

Although a single scoring criterion is used in evaluating the individual performance of the students by the evaluator, we can observe a variability in the measurement of their results, which seems to be involuntarily generated by the degree of subjectivity involved in the assessment, materialized in errors of correction and scoring.

2. Some Fundamental Aspects of Testing

There are four fundamental aspects of testing: evaluative, theoretical, practical and instructional. These aspects can be thought of as either determinants of testing, prescribing the form a test should take in a given situation and thus limiting what it can do, or functions of testing, defining what the purpose of testing is.

2.1. The evaluative aspect can be devided into three components. First of all, we can speak of measurement. Quantification of learning is one of the classic functions of testing. A good test is supposed to measure consistently and accurately whatever has or has not been learned. This is what is meant by the term *test reliability*. A good test is by definition reliable. An objective test is likely to be more reliable than its subjective or communicative counterpart. There are techniques, like multiple-scoring, by means of which even subjective tests may be made to acquire a sufficient degree of reliability.

Secondly, there is comparability. Tests results obtained as a consequence of administering a presumably reliable testing instrument should be comparable with test results obtained from any other similar test, even though the tests in question were taken by different groups at different times. Thus, measurement is to be consistent across different tests and testee groups, in addition to being so on a single test with a single group.

Thirdly, discrimination can be mentioned. Closely related to reliability and comparability are discrimination and grading. A test should provide scores on the basis of which we can discriminate between well prepared and less prepared students fairly objectively and place them in a reliable rank order. In this way, the teacher should come to know, as a result of evaluation, the achievement level of his or her students.

2.2. Regarding the theoretical aspect, the form a test takes is also significantly controlled by the theory or theories espoused by the tester or reflected in the textbook or the syllabus in use. The test-theory compatibility is named by testing experts *construct validity*. What a test aims at assessing is relatively easy to determine in the case of a content subject. For example, it would not be difficult to judge whether what a mathematics test is trying to assess is or is not mathematics. That is not the case of language learning. A structural-behavioral view of langusge lays greater emphasis on linguistic form, an aspect of language that lends itself more readily to *discrete-point testing*. On the other hand, a communicative view is heavily biased towards the testing of authentic language behaviour, which is difficult to translate into effective assessment procedures.

2.3. Related to the practical aspect are administrability and the testing environment. On one hand, a test that is both valid and reliable is no good if it is problematic with respect to ease of administration and practicability of performance of the learner. On the other hand, the testing environment should be congenial. A student taking an examination in a room where the temperature is very high or low is most probably not destined to do well on that exam.

2.4. Finally, one of the most important aspects of the testing process in the instructional one. This is basically concerned with the relationship of the test to the course. An important feature is how testing influences the way of teaching by providing insights into the learning process. Feedback is the term used for this kind of test-to-course flow of information.

To sum up, the four fundamental aspects of testing presented above are essential within the evaluation process.

3. The Most Common Evaluation Errors

The main evaluation error is the *"halo" effect (of irradiation)* (Stangl, 2010, p. 50) and the related stereotypes, which

consists in the over- or under-appreciation of the results of some students under the influence of a good overall impression on their performance. Positive school reputation leads to the emergence of the so-called *"halo"* which generates the overlook of some mistakes or of instances of poor performance. The dimensions of the *"halo" effect* are materialized in the *"gentle effect"*, translated as the tendency of applying an indulgent appreciation of the known people, as compared to the other students; *the generosity error*, justified by the need to enhance the results of the students in order to obtain a positive evaluation of the activity; the *"contamination effect"*, when the assessment is influenced by the grades obtained in other subjects; the *"Pygmalion effect"*, also known as *the anticipation or oedipal effect*, which is a subjective notation, determined by the installation of a fixation of the opinion formed about a certain student (this didactic behavior induced by the educator can ultimately lead to school failure) (Kaufmann, Zehnder, Vanderheiden & Winfried, 2008).

Another evaluation error is the *"contrast effect"* or *"order effect"* which consists in increasing or decreasing the grade or mark awarded to a student (or his/her work), depending on the better or worse quality of the performance (work) of the other students, due to the succession of different evaluations. *The central trend error* results in the rejection of very high or very low grades, from fear of not making a mistaking or out of desire to please everyone.

The logical error involves replacing some assessment criteria which correlate the students; performance with the the teaching objectives, with their adjacent / subjective variables: the graphic form, the legibility of the work, the originality in presenting it etc.

The effect of Gauss's curve results from the teacher's desire that the distribution of their students' results reflect the percentages established by Gauss's curve. This requires different scoring requirements, due to the level of performance achieved by each student. Thus, in a very good group, the demand increases, while in a group with generally weaker results, the demand decreases. Because of this evaluation error, the two groups with a qualitatively different global performance, have only two or three very good students and two or three very weak students (Artelt & Moschner, 2005, p. 45).

We can also mention the *constant individual error* (differentiated assessment of each teacher according to the level of personal exigency and his/her conception regarding the role and functions of the learning outcomes) or the examiner's *personal equation* (similarity error which consists in the tendency of the evaluator to appreciate by reference – in contrast to similarity – to his/her own person), due to which the students practice a mimetic behavior, similar to the evaluator's, in order to determine their generosity in scoring.

4. Ensuring Objectivity of Evaluation by Applying Grid Tests

Complete elimination of subjectivity in evaluation is not possible in the case of the human evaluator, and consequently, for the objectivity of the evaluation, it is necessary to apply grid tests. For these tests, the evaluator only determines the respondent's response option in relation to the model answers provided by the correction and scoring scales and gives him the appropriate score.

Evaluation is therefore a much more complex process, which involves a continuous effort to reduce subjectivity from the evaluation procedure in giving "grades" (scores, marks, percentages etc.). The grid tests seem to be the method of evaluation by which subjectivism is sensibly annihilated, eliminated from the context of valorization and notation itself.

The difficulty of using and generalizing this type of test is that it takes a long time to design the tests; that their application is also a chronophagous process; and that errors can be made when determining the results, due to the human factor involved in applying the evaluation tests. However, the automation of the results-based process appears as a natural objective of modernizing the instructional process, which is increasingly dependent on the computer on daily basis, but this is conditioned by the immediate improvement of the reliability of the results obtained and the minimization of the time consumed.

The errors that may occur in the process of applying the grid test may be related to the human factor, but are different in nature. Depending on the actual working mode, errors may occur when transcribing the results, matching the score options, or calculating the total score of each test. A solution used in the past consisted in introducing redundancy: all operations were carried out twice by different people. As a result, the reliability of the whole process of assessment through grid tests has increased, but a disadvantage has also appeared, as such tests have proven to be time-consuming and stress-inducing for the people involved in the process. However, since this type of assessment includes routine operations, these can be automated by using computing.

A variant of the automatic evaluation system is on-line testing, such systems being already implemented around the world, with assessment being conducted through the Internet. In Australia, for example, the on-line evaluation system is viable because of the great distances that students should go to get to school, so the solution is to use automated learning and testing. However, this option requires an adequate technical infrastructure needed to allow a large number of candidates to access this evaluation system simultaneously. Under the current conditions in Romania, the

implementation of such a system involves quite large expenditures. What could be done during an exam or competition in the first stage and then as a general practice, is the automation of the evaluation stage of the written grids. For this purpose, we need a scanner to electronically translate the candidates' grids, a software component to "identify" their answers (including instances where no response or multiple selections were selected), a component that compares the answers to the evaluation scale and counts the score for each test, and a component to ensure the printout of the evaluation results. Due to the obligation to classify the grids in an exam or competition (sticking a corner of the answer containing the applicant's identification data), this remains a step that needs to be done by the human resource involved in the organization of the test: opening the papers, correlating them with the scoring lists and displaying of results. The advantages of this system are: limiting the time required for the entire process (since the end of the exam until collecting the grids and then displaying the results), the accuracy of the results (reflecting the candidates' competencies), stress reduction for the candidates and evaluators alike (Friedrich & Mandl, 1992, p. 61).

The attempt to automatize the evaluation tools began with the use of software applications on the computer. Some of these programs are designed to make the instructional process in language learning more interactive, but not only (Wild, 2000, p. 25).

Their usefulness is particularly noticeable in the disciplines for which there are model lessons in the virtual library of the software. For humanities, the offer of model sessions is limited, so the teacher has the opportunity to manifest his/her creativity in designing computer lessons. In the tree structure of the program, the material required for the design of the teaching session is stored on two levels:

resources and session frames (task-based components of the lesson).

The computer-based lesson is organized according to the instructional events of any didactic scenario, namely: *organizational moment* – connecting to the program, accessing the virtual classroom of the proposed teaching session, connecting students/learners to the program. Once students/learners enter the virtual class, the secretariat automatically registers absences; *the anchor moment*: reporting the topic (*Deutsche Sitten und Bräuche*) to other previously discussed topics (*Die deutschsprachigen Länder, Deutsche Kultur und Zivilisation*). *The learning direction* comes next – scrolling down through the lesson frames. The teacher uses the front exposition method to present the information that students/learners follow on the computer monitor. *The feed-back moment* is provided by the application framework on *Deutsche Sitten und Bräuche* support text, where students/learners can view German customs and traditions.

The instructional sequence of *retention and transfer* is materialized in the ten- task structured test, which consists of questions with only one correct answer. At the end of the test, the computer generates individual slides for each student/learner that contain the correct number of responses and displays the test items with valid responses. The teacher receives the overall score of the class and thus writes down the points/grades in the class/group catalog.

The conclusions drawn after the lesson are translated into "strong points" and "weak points". The application "weaknesses" would be the monotony of the teaching-learning process, due to the use of the expository strategy, since the program does not allow the student/learner to modify the documents. Thus, the act of communication becomes unilateral: source – computer, receiver – the student/learner. Teacher intervention is limited to directing (introduces the lesson frames to fit in the time resource). Also, a minus of the program is that it eliminates the creativity factor in receiving the educational offer, and the student/learner cannot access the frames of the session individually, cannot return to them, the resources of the lesson can be printed only before their inclusion in the program, so there are no conditions for a personalized feed-back.

5. Conclusion

Positive aspects of the evaluation with grid test with automatic result processing are as follows: the program allows the condensation of strictly informative lessons in favor of the allocation of a wider time resource for application lessons; the program obliges to update the literary and critical information for the realization of didactic offers, customized alternatives to the course support or seminar. Also, the online resource diversifies the didactic strategies in the classroom, and the computerized instructional approach forces the student/learner to concentrate constantly during the session in order to be able to respond efficiently at the end of the lesson – the grid test evaluation. The program also has another advantage, namely creating the premises for learning the lesson in the classroom or in the laboratory.

In conclusion, no matter what forms of discrimination the teacher chooses to employ in the evaluation process, the values that animate the process of student/learner valorization are the most important, by relating them to the process of evaluation itself: the objectivity of the evaluation, the ability to discriminate between different types of results, the accuracy of performance measurement, the emphasis laid on added value.

REFERENCES

Artelt, C. & Moschner, B. (2005). *Lernstrategien und Metakognition. Implikationen für Forschung und Praxis*, Münster: Waxmann, 45.

Friedrich, H. F., & Mandl, H. (1992). *Lern– und Denkstrategien. Analyse und Intervention*, Göttingen: Hogrefe, 61.

Kaufmann, S., Zehnder, E., Vanderheiden, E., & Winfried, F. (2008). *Qualifiziert unterrichten. Fortbildung für Kursleitende Deutsch als Zweitsprache,* Band 2, Didaktik Methodik, Hueber Verlag, Ismaning, 50.

Stangl, W. (2010). *Halo-Effect. Lexikon für Psychologie und Pädagogik*, available at: http://lexikon.stangl.eu/1655/halo-effekt.

Wild, K. P. (2000). *Lernstrategien im Studium*, Münster: Waxmann, 25.