About Evaluation in Chemistry Education

Meliha Zejnilagić-Hajrić
Department of Chemistry, Faculty of Science, University of Sarajevo
Sarajevo, Bosnia and Herzegovina
mzejnilagic@pmf.unsa.ba

Abstract

Evaluation needs to be continuous process and students should perceive it as such, as a way of teaching and a way of keeping students development. Therefore, any kind of assessment should not make students afraid or anxious. This paper presents results of a research conducted among 580 primary school students (7^{th} and 8^{th} grade) and nine chemistry teachers in order to get insight in their perception of evaluation in chemistry education. We have compared students' and teachers' opinions in two cantons in Federation of Bosnia and Herzegovina using comparable questionnaires for students and teachers designed for the purpose of this study. Results show that every teacher has its own way of assessing her/his students, that performing an experiment, if performed at all, mostly does not affect students' mark. However, knowledge is not the only factor that has influence on the mark in chemistry, but the overall students attitude towards chemistry and her/his entire personality. Differences in students' and teachers' responses between cantons were not significant. Further research should be conducted in order to evaluate outcomes of teaching process and therefore to make some changes in evaluating both students and teachers. The purpose of evaluation should be a contribution to students' progress and a tool for increasing students' motivation for learning and for gaining better results.

Keywords-evaluation; primary school students; chemistry education; chemistry teachers;

1 Introduction

The task of evaluation is to ensure monitoring of the development of students to determine their weaknesses and affirmation, to highlight areas where preferred pedagogical measures are necessary, as well as to create a basis for changing the curriculum and to apply experience to meet the needs of students [1]. In addition, it enables them to identify their own goals and achievements as well as shortcomings in learning [2].

Evaluation is defined by Oxford Dictionary of Education as "the measuring of the effectiveness of a lesson, course, or programme of study, often based upon, among other source of evidence, the views and responses of the learners concerned, which constitutes qualitative evidence, as well as upon a quantitative assessment of the impact the course or lesson has had on learners' level of attainment. As well as participating in formal evaluations carried out for institutional purposes of quality assurance, teacher are encouraged to evaluate their own performance in planning, teaching, assessing, and supporting learning through a process of reflection on their professional practice. Evaluation is a process quite distinct from assessment. While student attainment may be assessed, it is the effectiveness of the processes which have contributed to their learning which are the focus of evaluation, although this may include using assessment data on attainment as one source of evidence"[3].

The evaluation has the task of determining the level of personality development, to inform students about the results they obtained, to point out its flaws and problems in the work and the causes its failure to stimulate the students' interest in the work and nurturing a culture of work, that developing talent and special skills of work etc. The evaluation should determine how achieved general, special and individual goals and tasks of education and the degree to which developed personality students [1].

Scales used in educational system shall be determined by agreement and are not equal everywhere. For example, in Bosnia and Herzegovina scale from 1 to 5 is used (1 being lowest, 5 being highest mark) in primary and secondary schools. In some countries apply the scale to 100. Sometimes in the evaluation applies only two degrees, "passed-failed" [4]. There are different marking systems in the world. For example, in the United States is 12-level scale, the lowest F (F+, F-) to the highest A (A+, A-) level. In other words, different educational systems have different ways of grading, which indirectly confirms that there is no consistent and enough good theoretical solutions for a comprehensive uniform practices [5].

In order to achieve a subjective evaluation of what we want and what it takes, in the evaluation should take into account the quantity of knowledge, the quality of the knowledge, attitudes students the work that the problems arising in the development process and that on that basis, apply appropriate pedagogical measures [1].

Aspects of teaching that are part of the final assessment include testing, weekly assignments, oral or written answers, research projects, essays, group projects and laboratory exercises. For every aspect of teaching, the important skills knowledge and achievement should be identified, and specify the ways in which the students will be evaluated [6].

If as a measure of actual knowledge of the evaluated student is taken a joint evaluation of a large number of independent evaluators, it can be used to determine the diagnostic validity of a mark which was given by a single evaluator. Such studies have shown that the scale of

1-5, individual assessment often deviates in the range of one. So if the mark was 3 by a single evaluator, the real mark is probably somewhere between 2 and 4 (3 ± 1). The final marks 1 an 5 can deviate only on one direction, which indicates that their accuracy is probably greater than the accuracy of other marks, but the range of knowledge covered by these marks is higher than a range of other marks (further knowledge after excellent marks and further ignorance after poor marks shall not be registered). Certainly one should to keep in mind that these numbers are only replacing the names, ie. they do not indicate quantitative values. Mark 4 does not mean that whoever gets this mark knows twice as those with grade 2 [4].

Evaluation is a continuous process. The teacher monitors, evaluates and assess students' activities continuously. If the is evaluation observed as a continuous process, as a way of teaching or as a way of keeping students' development, then assessment would not enter restlessness, anxiety and fear in students [7].

Systemic factors operate constantly, and their impact can be relatively easy to define and predict. Thus, the average teacher for grades 1-5 evaluates student's entire personality, while subject teachers (grades 5-9) give advantage to individual achievements [4].

“Halo effect is the tendency of assessors to be influenced by their previous evaluation of a learner's standard of work when arriving at an assessment decision. Therefore, if a teacher has already come to the conclusion that a learner produces good (or poor) work, so the halo theory goes; the mark or grade they award, even though it might be in an unrelated subject, may be in danger of reflecting this prejudgement, rather than being based entirely on the standard of the actual work being assessed. This effect can also apply to their perception of learners' behaviour, in line with the old saying „Give a dog a bad name...“ [3].

Measuring knowledge

It is necessary to choose the appropriate instruments ie. find ways that will be able to assess how they achieved certain goals in the educational work [1]. For evaluating the achievement of educational goals, it is necessary to have accurate information on the outcome of this process.

Tests achievements can be divided into formal and informal [7]. Formal and standardized tests are more carefully constructed than informal. Their items have been revised several times by several people. In our schools, informal tests are mostly used. These are tests which psychometric properties are not known. It is therefore of great importance that teachers have sufficient knowledge of the ways of constructing tests. In use are usually two types of informal tests: a „short answer“ and „essay test“ [7]. Tests usually include questions about the matter, which is essential for the relevant discipline.

Well constructed test measures cognitive skills, not just knowledge of the facts [6]. However, it is virtually impossible to say what kind of a test has more impact on students' learning [8].

When selecting measurement forms, it is necessary to take into account several factors, such as: the available time students have for testing, the time teacher needs to evaluate tests, and the matter that is evaluated [6].

The test with short answers is an objective type test. Objectivity of the test is achieved through [7]:

- Carefully designing each item,
- Careful selection of representative items,
- Preliminary test of selected items,
- Setting norms
- Accurately determine the value of certain items.

In a test of short answers different types of items can be used. The most common are: matching and completion items, items of alternative choice, multiple choice items [7].

In one aspect, if testing is more often, the students will be more accustomed to this situation and will be less uncomfortable. Since the testing and evaluation serve as a means of increasing motivation of students, they should be administered as often as possible. In order to achieve the positive effects of the test, the feedback information should be received as soon as possible: if it is possible, immediately. It is particularly useful to explain the assessment - to say what did student know well, and where did he make mistakes. The advantage of the written test is that students usually have less fear. The lack of written tests is that it is impossible to help those who experience fear on written examination.

Practical work in primary school chemical education

Laboratory work include the aspect of reasoning, teamwork, creating experiments, data acquisition, recording, analyzing data, discussion, interpretation and reporting [6]. One way of assessing laboratory work is [9] :

- Understanding results, regardless of whether they agree or disagree with expectations,
- Application of theory when selecting the way that experiment is going to be performed,
- Method of writing, presenting and analyzing the data obtained, recording data and final reports
- The performance of the experiment.

Rondini and Feighan (1978) describe laboratory work in chemistry as laboratory assignments in which students after every lab exercise are awarded by points. These points take into account the final product, use of laboratory equipment, handling chemicals, cleanliness of products, the time required to perform laboratory exercises, techniques and taking all preventive measures. Students thus quickly find out which aspects of laboratory work should be improved.

Group activities are difficult to assess individually. Many believe that one of the better methods of evaluation of group work is giving each member of the group the same score, then giving overall responsibility for answers, presentation and

overall group results. This encourages better students to help those who are not so good. Observing groups at work will enable getting to know each member of the group and exploring ways to cooperate with others. One approach in the evaluation of group work is that the students estimate each member's contribution, including themselves [6]. Some authors recommend that the group should work in small measures that affect the final grade [11].

2 Research Methodology

2.1 Research Instruments

Questionnaire consisted of 19 questions regarding process of grading in primary school in two Cantons to compare how teachers use criteria for evaluation knowledge of students.

Questionnaire for teachers consisted of 20 questions to see opinion about way to evaluation knowledge in chemistry by them.

2.2 Participants

Participants in this research were 7th and 8th grade primary school students, total 580 students. In two cantons we have conducted research in six primary schools, three in each Canton. There were 264 students in one canton (Una-Sana Canton, USC) and 316 students in second canton (Canton Sarajevo, CS).

3 Results and Discussion

3.1 Data obtained from Questionnaire for students

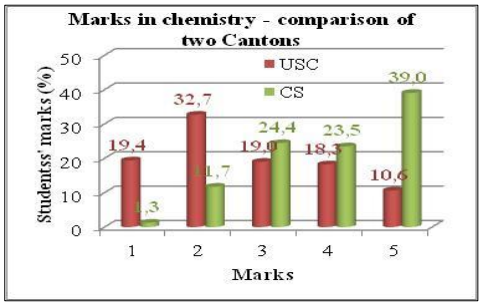

Figure 1. Marks in chemistry – comparison of two Cantons (1-the lowest mark; 5 – the highest mark)

It is obvious from the Fig. 1 that marks in CS are higher than marks in USC.

For example, there are almost four times as many highest marks (5) in CS than in USC, while there is only 1,3% of lowest marks (1) in CS and in USC almost 20%.

Table 1. Statistical parameters for markets between two cantons

Parameter	CS N=315*	USC N=263*
Mean	3,87	2,68
Standard Error	0,062	0,078
Median	4	2
Mode	5	2
Standard Deviation	1,10	1,27
Sample Variance	1,21	1,62
Kurtosis	-0,86	-0,97
Skewness	-0,51	0,36
t	12,1	

*In both cantons there was one student who did not answer this question about mark in chemistry

The mean is higher in Canton Sarajevo but mode and median show how distinct marks are in these two cantons. Statistically significant difference between two cantons is also confirmed by results of t-test (t (576)=12,1, p<0,05).

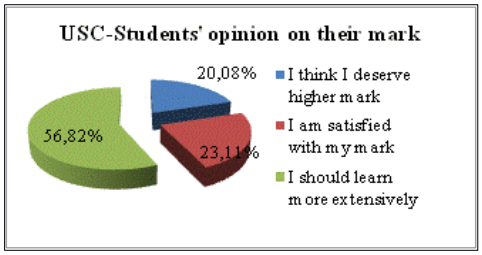

Figure 2. USC – Students' opinion on their mark in chemistry

It is obvious that only 23,11% students were satisfied with their mark in chemistry. Comparing to Fig. 1, it can be seen that there is only around 10% of students with highest mark in USC.

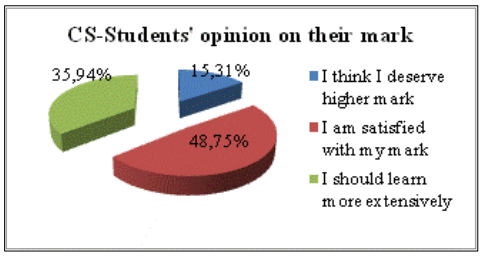

Figure 3. CS - Students' opinion on their mark in chemistry

There is a difference between USC and CS about mark from chemistry. In CS students are satisfied with marks more than students from USC.

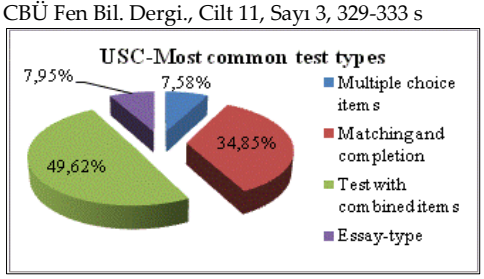

Figure 4. USC – Most common test types in chemistry

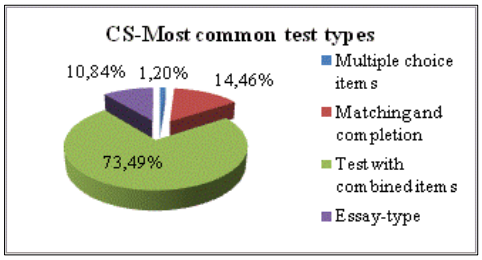

Figure 5. CS – Most common test types in chemistry

Results show that most common test types in both cantons are tests with combined items (USC: 49,62%, CS: 73,49%).

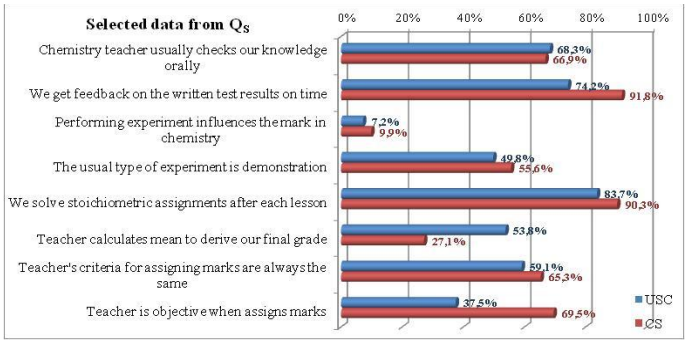

Figure 6. Selected data from Questionnaire for students – comparison between two Cantons

It is interesting that most students said that chemistry teacher usually checks their knowledge orally. Teachers from CS are somewhat more efficient in giving feedback from written test of knowledge. For some reasons teachers do not evaluate students skills in performing experiments since they do not affect the mark in chemistry. It is significant that nearly 50% of students have opinion that teachers usually made demonstration. After each lesson students solve stoichiometric assignments. Calculating mean for deriving final grade is more common in USC than in CS. More than half students believe that their teachers' criteria are always the same when he/she assigns marks. Students think that teachers from CS are more objective when they assign marks.

3.2 Data obtained from Questionnaire for teachers

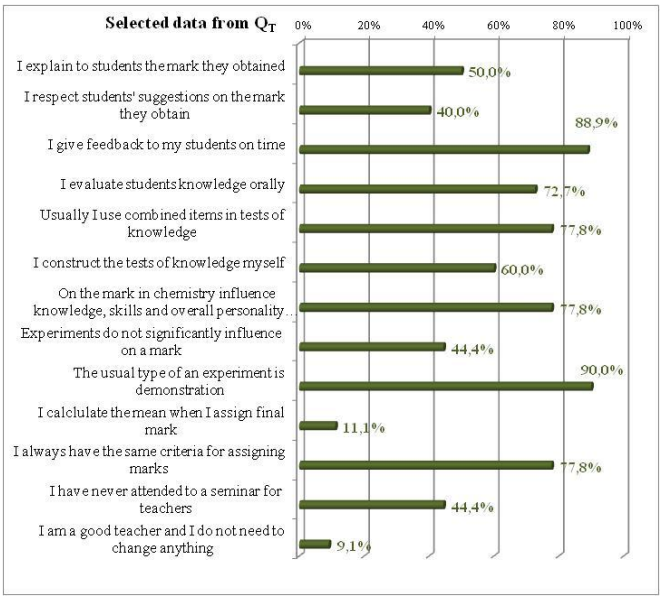

Figure 7. Selected data from Questionnaire for teachers

Teachers usually give their students feedback on time on their test results, which is also confirmed by results of QS. The same situation is with oral evaluation of students knowledge – results are similar for teachers and students. However, performing experiments do influence mark in chemistry (66,6% of teachers), but students do not agree with this statement (USC: 7,2%; CS: 9,9%). Also, there is difference between students and teachers opinion about deriving final mark: only 11% of teachers do that by calculating mean, but students believe (USC: 53,8%; CS: 27,1%) that their teachers do that more often. It is also significant that 44,4% of teachers have never attended seminar for teachers.

3 Conclusion

The comparison between cantons within this research was not made to say if one canton is better than the other, but to see if there are differences in common chemistry teachers' evaluation practices. Due to small number of teachers we have not discriminated them by schools or cantons they teach in but there was satisfying number of students to get some overall insight.

Some earlier research we conducted indicated insufficient number of performed experiments in chemistry classes in our country, both demonstration and students' hands-on activities, at least comparing to the number of prescribed experiments in chemistry curriculum. Therefore it was expected for results to show that performing an experiment, if performed at all, mostly does not affect students' mark. There was a discrepancy between students' and teachers' answers for this particular question, but results of Q_S were very similar in both cantons.

Teachers also say that skills and overall personality of a student affect the mark. The truth is that a student cannot know exactly what affects the mark, but approx. 40% of students believe that their teacher calculates the mean from their marks when deriving a final grade, while only one teacher confirmed this statement.

However, knowledge is not the only factor that influences on the mark in chemistry, but the overall students' attitude towards chemistry and her/his entire personality.

In order to make changes in their evaluation practices, it is necessary to include hands-on activities into chemistry classes, but this does not depend on teachers only. For that purpose, we need to make sure to continuously educate our teachers since 44,4% of them have never attended seminar for teachers, but only one of them does not believe that they need them (this teachers believes that he is a good teacher and there is no need for changing anything). Moreover, the school management needs to recognize the importance of practical work in chemistry classes and to make sure that teachers and student have the necessary resources for it.

Further research should be conducted in order to evaluate outcomes of teaching process and therefore to make some changes in evaluating both students and teachers. The purpose of evaluation should be a contribution to students' progress and a tool for increasing students' motivation for learning and for gaining better results.

Finally, it can be concluded that every teacher has its own way of assessing students.

3 References

[1] Mandić, P.; Vilotijević, M. Vrednovanje rada u školi, 1980, Sarajevo: IGKRO Svjetlost.

[2] Grgin, T. Evaluacija edukacijskih dostignuća in Edukacijska psihologija, 1997, Jastrebarsko: Naklada Slap

[3] Wallace, S. A Dictionary of Education, 2008, Oxford: Oxford University Press

[4] Andrilović, V.; Čudina-Obradović, M. Procjenjivanje, ocjenjivanje, vrednovanje in Psihologija učenja i nastave, 1996, Zagreb: Školska knjiga.

[5] Simić, K. Bitne komponente pri izvođenju zaključnih ocjena kao dio rješenja problematike tradicionalnog i modernog ocjenjivanja. Didaktički putokazi. 2007; 13 (45), pp. 61-67.

[6] National Research Council, "Science Teaching Reconsidered: A Handbook", 1997, Washington, DC: The National Academies Press.

[7] Slatina, M. Nastavni metod, 1998, Sarajevo: Filozofski fakultet Univerziteta u Sarajevu.

[8] Ebel, R.L.; Frizbie, D.A. Essentials of Educational Measurement, 4th ed.,1986, Englewood Cliffs, N.J.: Prentice-Hall.

[9] Kandel, M. Grading to motivate desired student performance in a descriptive laboratory course. J. Col. Sci. Teaching. 1989; 18(4), pp. 249-251.

[10] Rondini, J.A.; and Feighan, J.A.An ongoing grading technique for laboratory courses. J. Chem. Educ. 1978; 55(3), pp. 182-183.

[11] Johnson, D.W.; Johnson, R.T.; Smith, K.A. Active learning: Cooperation in the College Chemistry Classroom, 1991, Edina, Minn.: Interaction Book Co.