

## Evaluation of Classroom-based Online Multimedia Language Assessment

**Ana Cetinić**

E-mail : [anacetinic@gmail.com](mailto:anacetinic@gmail.com)

**Sanja Seljan**

University of Zagreb – Faculty of Humanities and Social Sciences  
Department of Information Sciences, I. Lucica 3, 10 000 Zagreb, Croatia

E-mail : [sanja.seljan@ffzg.hr](mailto:sanja.seljan@ffzg.hr)

### Abstract

*Implementation of on-line multimedia language assessment represents a domain of high interest, although in Croatia it is still in the beginning phase. It could be used as summative testing at the end of education cycle or lesson, as formative testing during the education period, as diagnostic testing usually at the beginning of education period, i.e. as self-evaluation or peer assessment method. Bloom points out six levels of cognitive learning (including fact knowledge, understanding, implementation, analysis, synthesis and evaluation), skills and attitudes, which are to be covered by the assessment process.*

*Having certain advantages, but also disadvantages, the best results are obtained as complementary method with traditional assessment and in combination with other types of assessment as oral communication, individual research, group work, tandem learning, etc.*

*affective attitudes*

*For the purpose of evaluation of on-line multimedia language assessment, the research was conducted at the Faculty of Humanities and Social Sciences, University of Zagreb among the group of 29 Croatian students studying the Italian language. The online multimedia test was implemented using Omega (Croatian version of Moodle LMS system) containing 21 questions relating to the Italian language and culture. Questions were accompanied by various types of multimedia – audio file, video clip and a map of Italy. The test covered knowledge, as well as listening, reading and writing skills, while speaking skills were not covered. After the test, students were asked to fill out a questionnaire to express affective attitudes. The results of the test and of the have been analyzed. At the end, the guidelines for further researches are suggested.*

### 1. Introduction

In recent years there is obvious need for evaluation of on-line classroom-based assessment. Widely known language centres, such as Cambridge and Oxford claim the opportunity of computer-based testing adjusted to Common European Framework of Reference for Languages (CEFR) from A1 to C2. Oxford Associates claims to evaluate not only the language proficiency, but also communication and interpersonal skills. While some tests tend to be adapted to „real life“ communication capabilities, the others, such as the English proficiency test, tend to be business and academic-oriented testing the four language skills and grammar.

The term *evaluation* is often used in the same meaning as the term *assessment*, but according to Marsh (1994), ref. [4], there are some major differences. Evaluation includes assessment. It is often said for assessment that it is guided by the criteria or standard. Criteria-guided assessment assumes data gathering about individual in relation to a particular task or criteria. Each task has a series of criteria or abilities that need to be achieved if one wants to be successful. According to Marsh (1994), norm-guided assessment assumes comparison of success of the individual or group with others that are in the same category.

In this pilot study the term *assessment* is used in describing activities which teacher performs to get information about the knowledge, abilities and attitudes of students. Before creating tests, according to

MacGregor, ref. [3], it important to consider the type of testing, the purpose of testing, who will take the test and which types of skills and knowledge will be assessed. Also, classroom tests have less important reliability (referring to consistency with which tasks measure the same material) and validity (the extent of measuring, i.e. how much test tasks show knowledge, abilities and values which are being assessed) than large-scale standardized tests.

## 2. Types of evaluation

Online assessment is the process used to measure certain aspects of information for a set purpose where the assessment is delivered via a computer connected to a network, ref. [6]. When talking about the use of computers in assessment, the following terms are used as synonyms: Computer-Aided/Assisted Assessment, Computer-Based Assessment and Online assessment.

Some of *advantages of online assessment*, according to Rister (2006), ref. [7] are availability, privacy, feedback, repeated application, lower implementation cost, mass simultaneous testing, use of multimedia, adaptive assessment, objectivity and reliability, availability of software, digitalization, new forms of testing (e.g. portfolio), attractiveness, challenges of change, shared resources and integration into VLE (Virtual Learning Environment). The aspect of time requirement significantly changes, being considerably increased in the preparation phase, while the process of test correction is significantly reduced. Booth (2002), ref. [2] points out that online assessment is subject to several criteria that overlap with traditional assessment, such as objectivity, reliability and honesty.

*Some of the disadvantages of online assessment* are technical or conceptual, ref. [7], such as security, hardware and software problems, speed connection, cheating, authentication and plagiarism, human factor, time requirements, interoperability, dominance of technology, essay-type tasks, inflexibility and IT knowledge.

According to primary purpose of evaluation, several basic types are to be distinguished, ref. [7]: *summative evaluation* – at the end of educational cycle, *formative evaluation* – during the educational cycle, *diagnostic testing* – commonly used before the start of the education cycle, *self-evaluation* of proper knowledge and skills and *peer assessment* between students themselves. Each one of them having some specific characteristics, should be included into education cycle, and can be used in on-line form.

Bloom, ref. [1] points out six *levels of learning*: fact knowledge, understanding, implementation, analysis, synthesis and evaluation making contained in the cognitive purpose (knowledge) of education, while another two purposes - affective (attitudes) and psychomotoric (skills) – make equally important aspects which should be included into education cycle.

## 3. Research

The research on evaluation of on-line multimedia assessment was conducted at the Faculty of Humanities and Social Sciences, University of Zagreb among Croatian students studying Italian language and literature. In the research, 29 students have completed on-line multimedia test and questionnaire, whose results have been analyzed.

### 3.1. Test

The test consisted out of 21 questions of different types: multiple choice, short answers, matching question and essays. It was intended for the intermediate level of language knowledge, covering knowledge of Italian culture (regions of Italy, market in Rome), as well as activities of listening, reading and writing, while speaking activity was not included. The general knowledge assessment included finding synonyms and Italian regions on the map. Listening assessment activities included listening with compression of the song and listening and watching of video clip with cultural elements. The reading assessment included reading of the text related to the market in Rome and answering the questions. The writing activity included writing of the short essay on the topic of representing the proper country to the foreign visitor and offering the authentic souvenir. At the end, the questionnaire on affective attitudes (6 questions) toward on-line multimedia assessment was given to the same students.

### 3.2. Evaluation

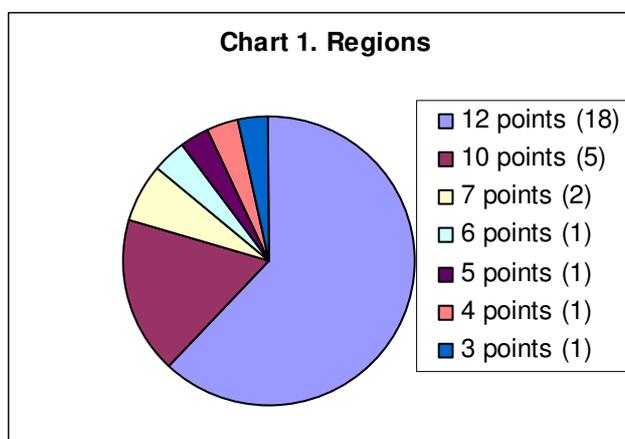
The Table 1 presents the average results of different types of activities. The average point for the first 5 questions related to the *listening activities* (typical Italian song) was 0.55 out of 1. The average result of the *listening and watching activity* (video clip – market in Rome) with included cultural elements has shown considerably better results with average score of 0.81 point (p) out of 1. In the group of

questions related to synonyms search, the students have achieved the average result of 0.86 p out of 1. In the *reading with comprehension* activity students in average achieved 7.86 p out of 8.

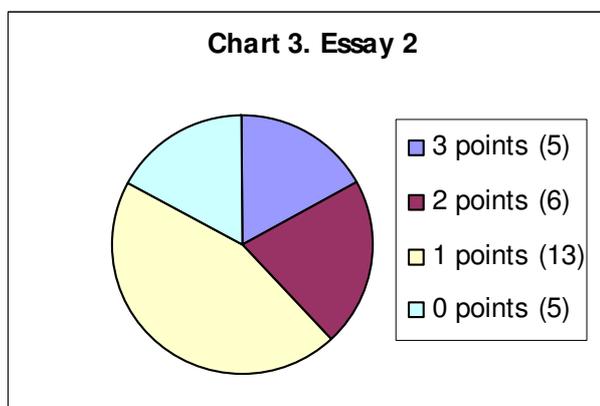
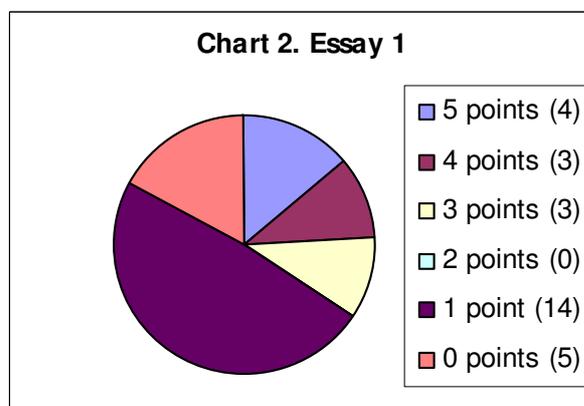
	<b>Exercises</b>	<b>Average</b>
Listening	song	0.55 / 1
Synonyms	questions	0.86 / 1
It. regions	blind map	10.27/ 12
Video	video clip	0.81/ 1
Reading	text	7.86/ 8
Essey1	presenting country	1.89 / 5
Essey2	souvenirs	1.38 / 3

Table 1. Average results

In the task related to find the corresponding Italian regions, the students have in average correctly marked 10.275 regions, out of 12. The distribution presented in Chart 1 marked that students had either good knowledge (62.07%), i.e. 18 of students with maximal points (12) and 5 students with 10 points, either low level (6 students between 7 and 3 p).



In the first essay question „*How would you present Croatia to a foreign visitor?*“ students wrote in average 4-5 sentences presenting the geographical position in Europe, neighbour countries, population number, capital city, some natural and national parks, culture and gastronomy. Evaluation criteria was as follows: 1 p – very short description, 2 p – a little longer description, but not enough, 3 - medium size description, 4 - subjective and objective description with some facts , 5 p – subjective and objective description with many facts. As presented in Chart 2., 17.24% students didn't at all write the answer, while almost half of students (48.28%) gained only 1 point. Nobody gained 2 points, 3 points 10.34% of students, 4 points also 10.34% of students and 5 points 13.79% of students; 24.13% of students received 5 and 4 points, 10.34% 3 points and 65.5% of students received 1 and 0 points.

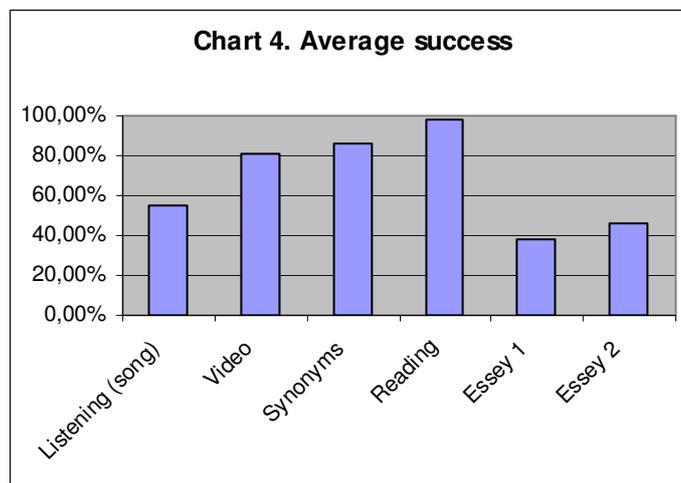


Another essay related to the task „According to your opinion, which are authentic Croatian souvenirs?“ (Chart 3) where students could gain at maximum 3 points. 1 point was given for one indicated souvenir, 2 points for 2-4 souvenirs and 3 points for 5 or more indicated items. Again, 17.24% students haven't answered this question, and almost half of students (44.83%) gained 1 point, altogether 62.07% having low results. Medium results (2 points) received 20.69% students and 3 points 17.24% of students. As presented, both writing activities indicate that students did not give enough attention to the writing task integrated into on-line test and that results would be probably much better if it was given as separate task. In both essay questions (65.52% and 62.07%) gained 0 and 1 points.

### 4.3. Results

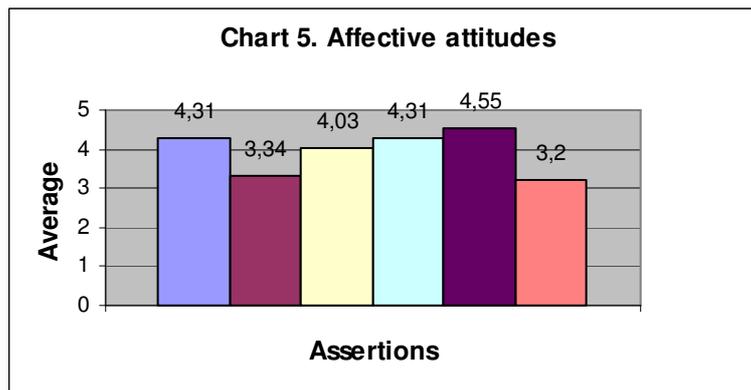
The Chart 4 presents average success in % for each type of activities. The lowest results were achieved in *writing skills* (essays). Low number of gained points, considerable number of unanswered questions and short sentences indicate that students preferred it as separate task.

The best results were achieved in *reading skills* (98.25%) and finding *synonyms* (86%) indicating that on-line assessment is preferred for the tasks of reading with comprehension and short answers. High success was also achieved in *video* use (81%), connecting listening and watching with cultural elements, contrary to the pure listening actives (55%). This indicates that well-chosen video clips and pictures are motivating and have considerable impact on the results of language learners. In both writing activities (2 essays) students achieved in average 41.9% which were the lowest results.



At the end, the students were given a questionnaire on affective attitudes using the scale from 1 (I do not agree with assertion) to 5 (I completely agree). The assertions were given as follows:

- 1 – I like this test.
- 2 – This type of test is enough to assess the knowledge.
- 3 – For complete assessment, it is necessary to use oral evaluation.
- 4 – This type of test was more interesting than classic assessment methods.
- 5 - It was easy to work with multimedia on-line test.
- 6 – This assessment can present the real knowledge.



The Chart 5 presents the average scores of the affective attitudes. The best scores were given for the assertion 5 (4.55), assertions 1 and 4 (4.31), indicating that students enjoyed working in multimedia environment and found it more interesting than classic assessment methods. The score 4.03 is given to the assertion 3, suggesting that students mostly consider necessary to use oral evaluation for speaking skills. The lowest results are given for the assertion 2 (3.34) and the assertion 6 (3.2) suggesting that this type of assessment can not present the real knowledge and that this type of assessment requires complementary methods.

## 5. Conclusion

The conducted pilot research has shown that on-line testing could be used as complementary method to the traditional examination methods. The best scoring is achieved in *reading with comprehension* activities and finding synonyms as short answers.

The *writing skill* (longer essay) should be taken as separate task, and probably implemented in traditional method, where students could be more concentrated on written expression, follow the logic order of thoughts and would not use proofing tools and checkers. *Speaking skill* was not included in the research, since it requires additional equipment (headphones and microphone). *Listening* activity is better to be taken as video clip, including *cultural elements* in order to motivate students on better results and on noticing of some specific cultural characteristics, contrary to the pure listening activity.

The questionnaire on affective attitudes toward on-line multimedia assessment indicates the positive attitude, easy manipulation and motivation. It is perceived through realistic opinion pointing out that this type of test is not enough to assess the student's knowledge and that traditional oral evaluation is necessary.

The conducted pilot research brought insight into evaluation of on-line multimedia assessment of language knowledge and skills, having positive effects on reading skills with comprehension, individual tasks asking for shorter answers (synonyms) and use of video clips with cultural elements. On the other side, essay writing is considered to be better if it was given as separate task. While speaking skill was not included, the pure listening skill gained low scores, suggesting that oral presentation and communication skills are to be assessed by traditional methods. The overall affective attitude towards this type of assessment is positive, although it is considered as complementary tool for the overall knowledge and skill assessment.

## References

- [1] Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co Inc.
- [2] Booth R., Clayton B., Hyde P., Hartcher R. & Hungar S.; Creative, Quality Online Assessment. Progress report to FLAG, July 2002.
- [3] MacGregor, D. *Second Language Proficiency Assessment*. CAL Resource Guide Online. Center for applied Linguistics.
- [4] Marsh, Colin J. (1994). *Kurikulum*. Zagreb:Educa
- [5] McLoughlin C. & Luca J.; *Assessing Students' Self-Regulatory Skills* in „Meeting at the Crossroads“. Proceedings of Australasian Society for Computers in Learning in Tertiary Education (ASCILITE). Melbourne 2001.
- [6] Online assessment (2010). Wikipedia, The Free Encyclopedia. [http://en.wikipedia.org/wiki/Online\\_assessment](http://en.wikipedia.org/wiki/Online_assessment) (15.1.2010.)
- [7] Rister Damir (2006). *Online ispitivanje*, VI. <http://edupoint.carnet.hr/casopis/41/clanci/1?CARNetweb=...>

