COMPUTERISED SELF-EVALUATION OF PERFORMANCE IN PROFESSIONAL ENGLISH

Galina Kavaliauskienė

Mykolas Romeris University, Lithuania, gkaval@mruni.eu

Lilija Anusienė

Mykolas Romeris University, Lithuania, lilija.anusiene@gmail.com

Abstract

Purpose – *the focus of this article is to explore self-evaluation of performance in professional English that students have to do in university studies. It is clear that students can have various problems in learning English and make mistakes in listening, speaking, reading, writing, and vocabulary. Speaking and listening are the skills that are more common on an everyday basis than reading and writing. Speaking and listening skills have been known to be more difficult to master. English vocabulary presents another problem for language learners. At the university level, students study professional English, also called English for Specific Purposes (ESP). Students will need it for their future profession, so it is important for them to be able to evaluate their achievements in learning.*

Methodology/approach – *the research paper adopts the qualitative research approach. The questionnaire on learner perceptions of computerised self-evaluation of various activities was administered to students of three different specialisations. Student perceptions of success or failure in their performance were analysed.*

Findings – *the results showed that attitudes to computerised self-evaluation depended on the difficulty of tasks and chosen future specialisation. The findings illustrate that the respondents of the three investigated specialisations encounter similar*

Socialinės technologijos/Social Technologies
© Mykolo Romerio universitetas, 2014
© Mykolas Romeris University, 2014

ISSN 2029-7564 (online)
https://www.mruni.eu/lt/mokslo_darbai/st/apie_leidini/
https://www.mruni.eu/en/mokslo_darbai/st/apie_leidini/index.php

difficulties, but to a different degree. There are no specific problems in performance due to the complexity of the professional vocabulary that students must learn. The computations of Pearson's correlations coefficients demonstrate good correlational relationships within each specialisation. The ratings of Psychology (PS) students reveal higher mean values and lower scattering of Standard Deviations in Moodle tasks than the ratings of Internet Management and Communication (IMC) students.

***Research limitations** – the investigated samples contain a limited number of respondents, which might raise a question of reliability of the findings. The statistical computation of the reliability of the obtained responses will prove whether the data are reliable and may be extended beyond the studied samples.*

The hypothesis for the research: student performance might depend on their specialisation.

***Practical implications** – the respondents were asked to indicate the degree of difficulty in performing various online tasks on the Likert's scale ranging from "very difficult" to "very easy". The analysis of the responses by means of the Statistical Package for the Social Sciences (SPSS) software proves that despite the limited number of respondents, the results may be extended beyond the studied samples. It means that self-evaluation of computerised activities by students studying professional English is recommended at the university level.*

***Originality/Value** – the value of this study encompasses the statistical processing of the responses, which should prove whether the findings are reliable or not.*

***Keywords**: English for specific purposes, computerised self-evaluation, formal testing, online listening, Moodle self-checking.*

***Research type**: research paper.*

1. Introduction

It is well known that online learning of English for specific purposes presents a considerable challenge for undergraduate students worldwide. The aspects of academic communicative competence, such as reading, listening, writing and speaking skills, are crucial to students' successful progress through the syllabus. Online learning comprises all forms of electronically supported learning and teaching. It mainly refers to using electronic applications and processes, such as web-based learning, computer-based learning, virtual classrooms and digital collaboration. Lately it has become part of subject studies at the university level. Online self-testing is a novel way for students to evaluate their achievements in language learning. This article focuses on student perceptions of their computerised self-evaluation of learning English for specific purposes online. The data of self-evaluation have been compared with the results of formal testing. Moreover, the data on student performance in Moodle activities are presented and analysed. The conclusions on the usefulness of self-evaluation in preparation for formal tests have been drawn.

This paper aims at examining application of self-evaluation in learning professional English and drawing conclusions about its suitability for language learning and testing at the university level.

The aims of the research: to investigate student attitudes towards different tasks in online learning at a tertiary level for revising professional vocabulary, self-checking reading and listening comprehension of authentic materials.

Research methods used: (1) a specially designed questionnaire to examine student attitudes to self-evaluation in learning; (2) actual performance in various online assignments, including Moodle activities; (3) statistical processing of the data by the means of Software Package for Social Sciences (SPSS); (4) analysis of written student responses intended to find out their opinions on self-evaluation.

The respondents in this study are the students of three different specialisations, namely Psychology (PS), Socio-Cultural Education (SCE) and Internet Management and Communication (IMC), who studied professional English at Mykolas Romeris University, Vilnius.

2. Literature Review

In the twenty-first century, nobody is indifferent to the technological change in teaching languages (Walker and White, 2013). Online learning is a wide and rapidly changing field in teaching English. E-learning refers to language teaching and learning using digital devices such as computers, tablets and smart phones, and it can be conducted by means of different tools, not necessarily designed for language learning. The latest are Web 2 tools that include weblogs, wikis, and some other social media. Current e-learning practices in teaching foreign languages at university level were overviewed in four different countries (Japan, New Zealand, Singapore, and the United Kingdom) by N. Cowie and K. Sakui (2013). They claim that the most common learning management systems are Blackboard, Moodle and Edmodo. The teachers use them to coordinate and monitor learners' progress as well as to post the instructions. Students use them to access materials, upload or download various tasks, communicate with peers and teachers. Moodle area adjusted to a particular course and settings is particularly popular. Moodle is a free software that by now has been adopted by universities worldwide. Moodle is an acronym for Modular Object-Oriented Dynamic Learning Environment. By January 2014, it had a user base of 66,148 registered and verified sites, serving 66,162,479 users in 233 countries (Moodle Statistics, https://moodle.org/stats/). Research shows that students want feedback (Cowie and Sakui, 2013), and Moodle provides it immediately after tasks have been accomplished. Numerous benefits of student activities in Moodle area were described in our previous research (Kavaliauskienė, 2011), and application of Moodle for language testing was recommended due to its advantages over traditional formal testing.

Traditionally, the test has been the most common way to measure achievement and proficiency in language learning (Frank, 2012). Testing is a tool that can help teachers identify student strengths and weaknesses and evaluate the effectiveness of

programmes, i.e. how well students have learned what they were supposed to have learnt. In a discussion of language assessment, Bachman (2007) reviewed testing practices of the past five decades and categorised them into seven approaches: skills and elements, direct testing and performance assessment, pragmatic language testing, communicative language testing, communicative language ability, task-based performance assessment, interactional language assessment. J. Pino-Silva (2008) examined student perceptions of computerised testing and came to the following conclusions: (1) the two most frequently reported advantages of a computerised test were instantaneous report of the grade and a fast and convenient manner in which it was taken and administered; (2) the two most frequently reported disadvantages were the need to adapt to a different way of testing and the visual fatigue that may result from reading on screen. In general, the data lead to the conclusion that student perceptions of the computerised test are very positive. A good example of progress in large-scale language self-assessments is DIALANG (Stoynoff, 2012). It is a self-directed assessment available for 14 languages, including English, that can be accessed free of charge via the internet. Examinees are able to assess their reading, writing, listening, grammar, and vocabulary abilities in a foreign language, and they receive feedback on their strengths and weaknesses as well as their language level based on the Common European Framework of Reference for Languages (CEFR).

3. Respondents

The respondents in this study are eight groups of the 1st year students (4 samples), who studied English for Psychology (PS), Socio-Cultural Education (SCE) and Internet Management and Communication (IMC) at Mykolas Romeris University in 2012 and 2013. The students are mainly native speakers of Lithuanian who have enrolled in the university after having studied general English at secondary schools. The design of the English courses reflects student needs in professional language. The courses are adjusted to the requirements for a Bachelor of Social Science degree. The levels of student proficiency are either B2 or C1 according to the Common European Framework of Reference for Languages.

4. Research Methodology

The findings have been obtained from four sources: (1) a survey completed by four samples of respondents, (2) student self-evaluation of their online performance in English classes, (3) formal testing of student knowledge, (4) comparison of grades awarded for Moodle tasks. The questionnaire was designed in accordance with the accepted standards of constructing surveys (Dornyei, 2003). It contained statements on student perceptions of computerised testing. The relevant part of the survey consists of 8 statements, to which students responded on a 5-point Likert's scale ranging from "very difficult" (1) to "very easy" (5). The responses were processed by

means of Software Package for Social Sciences (SPSS-18). Student self-assessment method was employed as a way of encouraging participants to reflect on their learning experience, their achievements or failures. Formal computerised performance testing was conducted in English classes, aimed at clarifying the point of how realistic students are of their own evaluations.

5. Results and Discussion

This part of the article reports the findings of the survey and analyses the key points that emerged. For the sake of clarity in visual displays of the data, negative responses (very difficult and difficult) and positive responses (very easy and easy) have been added up. Naturally, neutral responses have also been accounted for in the statistical treatment. The frequencies of positive responses in percentage terms are shown in Chart 1 and those of negative responses – in Chart 2. The 1st cylinders in both Charts represent the responses by the 1st sample (PS), the 2nd cylinders – by the 2nd sample (PS). The numbers of the survey statements from 1 to 8 are displayed on the X axis in accordance with the survey descriptions, as reproduced below. Statements from 1 to 5 refer to self-testing of ESP vocabulary from online course books. Statements 6, 7 and 8 refer to self-testing of online listening from the website Breaking News English.
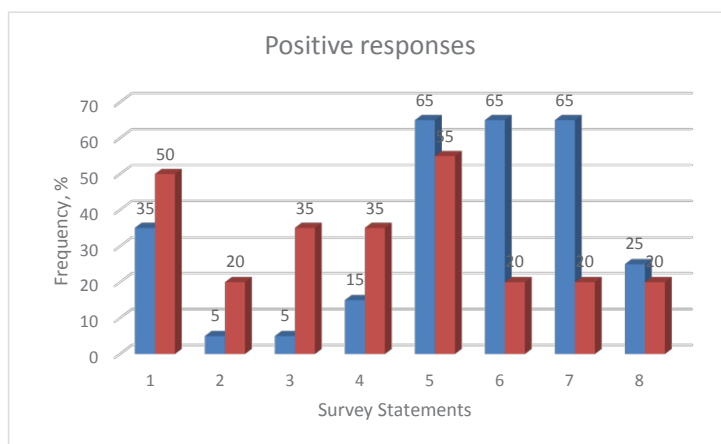

Chart 1. Self-evaluation data on computerised testing, positive responses by two samples.
The 1st cylinders show the frequency of positive responses by the 1st sample of PS students,
the 2nd cylinders show the frequency of positive responses by the 2nd sample of PS students

It is obvious that student perceptions of computerised testing differ in both samples. However, it is hard to define a certain trend in responses. Qualitatively, the most difficult areas are statements 2, 3 and 8 for both samples. The easy areas are activities described by statements 1, 5, and, to some extent, by 6 and 7. However, the existing scatter of responses is hard to interpret. It might be due to individual personalities of the respondents.
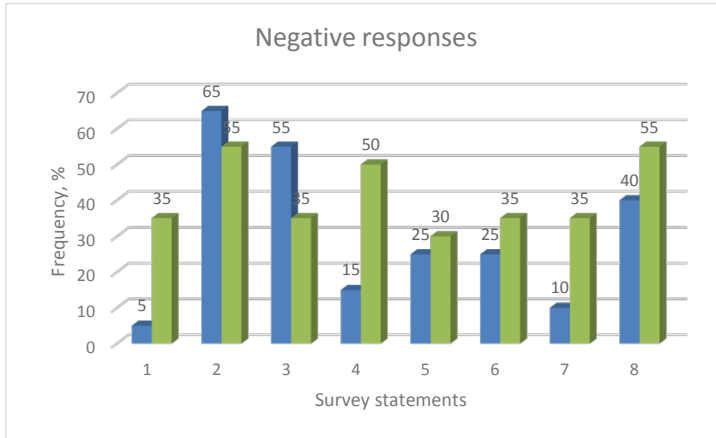
Chart 2. Self-Assessment data on computerised testing, negative responses.
The 1st cylinders show the frequency of negative responses by the 1st sample of PS students,
the 2nd cylinders show the frequency of negative responses by the 2nd sample of PS students

Survey statements were administered using the 5-point Likert's scale from (1) very difficult to (5) very easy.

1. Online course book MULTIPLE CHOICE vocabulary exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

2. Online course book GAP FILL vocabulary exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

3. Online Psychology Review WRITTEN QUESTIONS are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

4. Online Psychology Review MATCHING exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

5. Online Psychology Review TRUE / FALSE exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

6. Breaking News English Online Listening: GAP FILL exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

7. Breaking News English Online Listening TRUE/FALSE exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

8. Breaking News English Online Listening SYNONYM MATCH exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Responses of SCE and IMC students are similarly scattered and not displayed here as the data do not present any novel elements.

Traditionally, either Pearson's or Spearman's Correlation coefficients rho and the Significance levels Sig. p (2-tailed) are computed to determine any correlations between the samples. Good correlations **within** and **between** PS samples were reported earlier (Kavaliauskienė et al., 2013). Some results are reproduced in *Table 1*. As can be seen, four cases of good correlations have been detected either at the probability of 95% (Sig. p of .05) or 99% (Sig. p of .01). The values of correlation coefficients rho are either 1.000 or quite close to it, which indicates close relationships between responses.

Table 1. Pearson's correlation coefficients rho and Significance levels Sig. p for two samples of PS respondents

Correlations are found for the statements	Pearson's correlation coefficients *rho*	Significance levels *Sig.p* (2-tailed)
1 Online MULTIPLE CHOICE vocabulary exercises are **easy**	.853*	.031
5 Online TRUE / FALSE exercises are **easy**	1.000**	.000
2 Online GAP FILL vocabulary exercises are **difficult**	.841*	.036
8 Online Listening SYNONYM MATCH exercises are **difficult**	.908*	.012

* Correlation is significant at the .05 level
** Correlation is significant at the .01 level

Formal testing allows assessing the student learning quality. Chart 3 shows the grades that some students gained in formal testing of ESP vocabulary. It can be seen that on average grades vary within the range of 40% to 100%. It means that the majority of students have passed their formal testing with the exception of a few, who obtained 43% or 44%, which means failure.
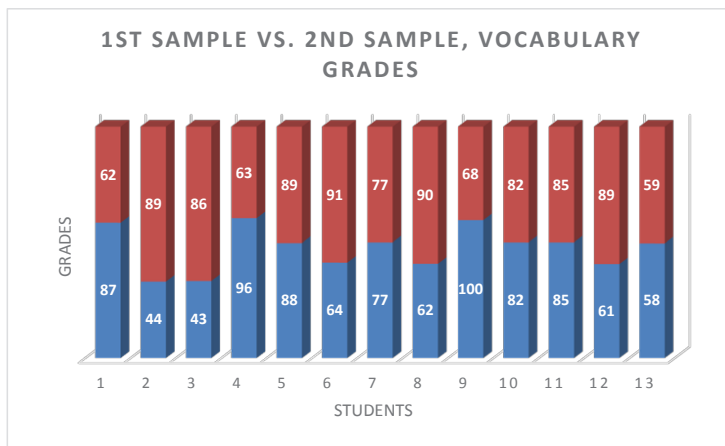

1ST SAMPLE VS. 2ND SAMPLE, VOCABULARY GRADES

Chart 3. Two samples (PS vs. IMC): Lower part – grades of some IMC students, upper part – grades of some PS students.

According to Chart 4, the grades that students obtained in formal testing of listening are within the passing limits. It is obvious that all students have passed the formal testing of listening with their grades between 50% and 100%.
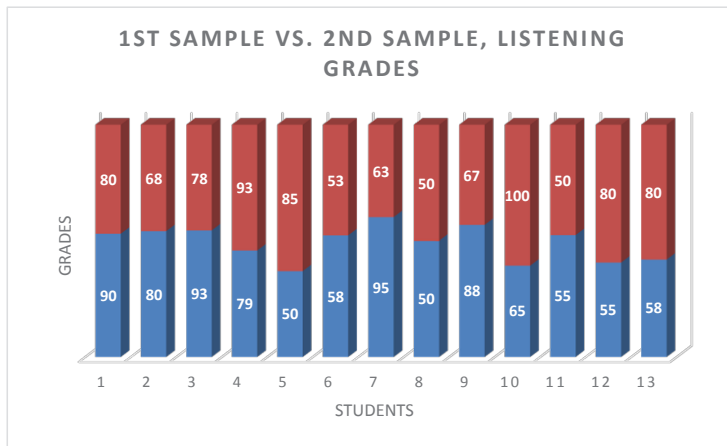
Chart 4. Two samples: (PS vs. IMC): Lower part – grades of some IMC students, upper part – grades of some PS students.

As it has already been mentioned, exercises in Moodle area give students the opportunity to revise covered materials and provide immediate feedback on their performance. Table 2 summarizes average grades of students in Moodle self-testing tasks in the academic terms of 2012 and 2013. It is obvious that the students of Psychology perform better in Moodle activities than the students of IMC or SCE: their average grade is 95% versus 85%, and the scattering of findings that is demonstrated by the value of Standard Deviation, is also smaller, 3.5 versus 10 to 13. Such discrepancy might be due to the differences in proficiency of general English, which affects the performance in professional English.

Table 2. Student performance in Moodle self-testing tasks (in 2012 and 2013). Meaning of abbreviations: PS - Psychology, IMC - Internet Management and Communication, SCE - Socio-Cultural Education

Number of respondents	Specialisation, Year	Mean value, %	Standard Deviation
34 (3 groups)	PS autumn 2012	95.41	3.48
57 (4 groups)	PS spring 2012	95.02	3.62
22(1 group)	IMC autumn 2013	84.78	13.47
12 (1 group)	SCE spring 2012	85.85	9.89

In order to assess whether the difference between the Means is due only to chance, the computation of one-way ANOVA has been employed. ANOVA allows computing the Group Means, the Overall Means, the total deviation of each score from the Group Mean known as Within Group Variation, and the deviation of each Group Mean from the Overall Mean known as Between Group Variation. Finally, ANOVA produces the F value which is the ratio Between Group Variation to the Within Group Variation. The F ratio is a measure of how different the Means are. If the difference between the Means is due only to chance, then the F ratio would be equal to 1.00. Generally, the F ratio is sometimes greater than 1 or less than 1. Therefore, it is essential to find the

critical F value. For this purpose, the Table of critical values for F distribution is available and must be used (http://homepages.wmich.edu/~hillenbr/619/AnovaTable.pdf). If the obtained value of F is equal to or larger than the critical value for experimental data, i.e. the number of samples and the number of respondents, than the result is significant at that Sig. p (or probability).

For the findings shown in Table 2, the obtained value of F is 18.936. The critical value of F is 3.44.

Thus, as the obtained value of F is larger than critical, the result is significant at the Sig. p 0.05 (probability of 95%).

6. Excerpts of student entries to portfolios

6.1. Entries by the Psychology students

http://eaidukaite.blogspot.com/ The most interesting thing about Moodle tests was that after doing one by one, you can improve your reading skills. Sure, there were a few moments when I could not do more because it was too tiring and exhausting. But I think I coped with them quite well. Online listening practice was sometimes quite difficult. Sometimes I listened to the recording and tried to hear what it was about, but it seemed impossible to follow. However, it helped me a lot. Now I can listen to the text fluently and now I make less mistakes not even in listening, but also in true-false matching, synonym matching and gap filling.

http://iadamonyte.blogspot.com/ In my point of view, Moodle tests were sometimes confusing but quite easy and my performance was really good. I think it's a good idea to solidify students' knowledge by doing these tests after each module. Online exercises from the course book were almost the same as Moodle tests. The main difference was that online exercises was a teamwork. I think my performance was quite good, but when I was working on my own on Moodle tests, my performance was better.

http://gpjanuskevicius.blogspot.com/ I have really enjoyed doing online listening. It was because the themes of recordings were really interesting, related to general Psychology. Also the fact that it was not very difficult made me like it.

http://simonakuzyk.blogspot.com/ Online listening practice including exercises (T/F, synonym matching, gap filling) - I love them mostly because it is not very hard, but also it is not very easy. I think I am quite good at it. While doing this task I extended my vocabulary.

http://vmockus.blogspot.com/ Moodle tests - this task was new for me. Despite the fact that we can do this work at home, I find this task hard to do. But the great thing about it is that you can check your knowledge and get results immediately. I did not do Moodle tests after I had read the module, so that was my mistake. Now I think that it's better to do Moodle test immediately after reading and revise the vocabulary.

6.2. Entries by the students of "Internet Management and Communication"

http://lureiciuna.blogspot.com/ Listening to online recordings: (a) Open-ended questions - It wasn't really easy I was struggling with some questions, but most of them was quite understandable. I would evaluate myself seven out of the ten in this point. (b) T/F exercises after listening - true or false exercises was one of the easiest because most of the time I manage to catch what it was all about. c) Synonym match - this exercise was most difficult for me because there were so many words that I don't know. d) Gap-filling task was easy because I remembered text from the recording. I would evaluate myself seven out of the ten.

http://mantasgrg.blogspot.com/ Listening to online recordings, answering open-ended questions and other exercises weren't very troublesome tasks for me. I've always managed to complete all these exercises with only a small amount of mistakes. In my opinion, I've improved quite well.

http://emstonkute.blogspot.com/ Answering open-ended questions was quite hard for me, because usually I have false understanding. I can not identify my development. The difficulty depends on the contents. However, T/F exercises after listening, true/false exercises are easy for me. At the beginning, I made a lot of mistakes, but now I do not make so many. Synonym match was the hardest exercise for me, because my vocabulary is limited. In this task, I do not see any progress. Gap-filling exercise is easy, sometimes I remember words or phrases, which helped me fill in the gaps correctly.

http://nvenckus.blogspot.com/ Task of listening to online recordings: (a) Open-ended questions task wasn't easy for me. I usually answer a bit more than half of them correctly, so I evaluate myself 6. (b) T/F exercises after listening. This task was one of the easiest for me, but I made some mistakes. I give 8. (c) Synonym matching task was not the hard one, I usually matched correctly 8 words out of 10, so my self-evaluation grade is 8. (d) Gap-filling task was a bit tricky. I failed to do it most of the time, so my evaluation is 6.

6.3. Entries by students of "Socio-Cultural Education"

http://surbonaite.blogspot.com/ Moodle tests were quite easy. My favorite task in Moodle tests is filling the gaps. Multiple choice is more difficult for me. Self-evaluation grades are between 7 and 8.

http://vmozeikaite.blogspot.com/ I think my performance in all tasks has been either very good or excellent, I have done my best to study hard. English is very important to me, because I often visit seminars, camps, and at school I participated in "Youth in Action" project to Romania. I am leaving to Serbia the day after tomorrow to take part in another project. English is an effective tool to communicate and create connections, understand other cultures, explore the world.

http://gmazilyte.blogspot.com/ Moodle tests demanded a lot of my time, but I think they were worth it. I had a great chance to revise everything that I've learnt. My self-evaluation grade in various tasks is about 7.

7. Conclusions

The hypothesis for the present research was the notion that student performance might depend on the chosen specialisation. The findings show that the undergraduates of three specialisations and four investigated samples encounter similar problems, albeit to a different degree. In other words, there are no specific difficulties due to the diversity and complexity of professional vocabulary. The ratings of Psychology students reveal higher grade mean values and lower range of Standard Deviations than the ratings of Internet Management and Communication students.

Computations of Pearson's correlation coefficients demonstrate some good correlational relationships within each specialisation. Statistical processing implies that the obtained results are not due to chance and may be extended beyond these four samples despite the limited number of respondents. Application of one way ANOVA to the interpretation of student self-evaluated Moodle grades confirms that there is a statistically significant difference between the Means. Student perceptions of self-evaluation have demonstrated that online activities present certain difficulties depending on the type of the task. Formal testing of professional vocabulary and listening has been successful for the majority of students, which suggests that self-testing is beneficial in learning professional English.

References

Books and articles:

Bachman, L. 2007. "What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment" in Fox J., Wesche M., Bayliss D., Cheng L., Turner C., Doe C. (eds.). Language Testing Reconsidered. Ottawa, Canada University of Ottawa Press.

Cowie, N.; Sakui, K. 2013. It's never too late: an overview of e-learning. *ELT Journal*, 67/4: 457–467.

Dornyei, Z. 2003. Questionnaires in Second Language Research. Lawrence Erlbaum Associates, Inc., Publisher.

Frank, J. 2012. The Roles of Assessment in Language Teaching. English Teaching Forum, 3: 32.

Kavaliauskienė, G. 2011. Moodle in English for Specific Purposes at Mykolas Romeris University. *Socialinis darbas*, 10(1).

Kavaliauskienė, G.; Anusienė, L.; Puodžiūkaitienė, Ž. 2013. Perceived Difficulties in E-Learning During the First Term at University. *Social Technologies*, 3(1): 25–38.

Pino-Silva, J. 2008. Student perceptions of computerized tests. *ELT Journal*, 62(2): 148–156.

Stoynoff, S. 2012. Looking Backward and Forward at Classroom-based Language Assessment. *ELT Journal*, 66(4): 523–532.

Walker, A.; White, G. 2013. Technology Enhanced Language Learning: Connecting Theory and Practice. Oxford University Press.

Internet sites:

Banville, S. Breaking News English website for listening practice [accessed on 2014-01-10]. <http://www. breakingnewsenglish.com/>.

Moodle Statistics [accessed on 2014-01-20]. <https://moodle.org/stats>.

Table of critical values for F distribution [accessed on 2014-01-14]. <http://homepages.wmich.edu/~hillenbr/619/AnovaTable.pdf>.

Questionnaire:

Breaking News English Online Listening: GAP FILL exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Breaking News English Online Listening TRUE/FALSE exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Breaking News English Online Listening SYNONYM MATCH exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Online course book MULTIPLE CHOICE vocabulary exercises are: (1) very difficult, (2) difficult, (3) not sure, 4) easy, 5) very easy.

Online course book GAP FILL vocabulary exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, 5) very easy.

Online Psychology Review WRITTEN QUESTIONS are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Online Psychology Review MATCHING exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.

Online Psychology Review TRUE / FALSE exercises are: (1) very difficult, (2) difficult, (3) not sure, (4) easy, (5) very easy.