

## SELF-ASSESSMENT OF LANGUAGE SKILLS AND EVALUATION OF PERFORMANCE IN ENGLISH FOR SPECIFIC PURPOSES CLASSROOMS

**Irena Darginaviciene**  
Klaipeda University, Lithuania

**Abstract.** *Although evaluation has been increasingly used in education for a long time, attention to evaluation and its uses in English language teaching is relatively recent. Essentially, evaluation may be carried out to determine if the students achieved a certain level of language proficiency. Recently self-evaluation of language skills has been used to find out how students assess their accomplishments in language learning. On the other hand, formal testing is the most common way to measure achievement and proficiency in language learning that teachers conduct during an academic year. Formal testing is considered to be an effective tool that can help teachers identify students' strengths and weaknesses and evaluate the effectiveness of teaching. This article focuses on student's perceptions of success in learning English for Specific Purposes (ESP) and formal evaluation of their performance in class activities. The methods of the research include the administration of the designed questionnaire, analysis of students' responses and their statistical treatment by a means of the Statistical Package for Social Sciences (SPSS 19). The self-evaluation data are compared with the results of formal testing. The conclusions are drawn on the merits of self-evaluation in learning ESP.*

**Keywords:** *English for Specific Purposes, formal testing, self-evaluation.*

### Introduction

Assessment and self-assessment have become an important part of learning and teaching foreign languages. Methods of assessment are based on learners' evaluation of their own learning. The importance of self-assessment is demonstrated by a number of publications in this area (Frank, 2012; Stoyhoff, 2012; Kavaliauskienė, 2005). Motivation is one of the most important factors for success in learning. Lack of interest to class activities leads to de-motivation in learning. Self-evaluation has been widely used in assessing the effectiveness of education at tertiary level (Douglas, 2000): the key features of assessment are learners' active participation in the evaluation of their own learning and the development of reflective thinking. The outcome of self-assessment shows what learners have learned in communicative classrooms. Self-evaluation is often used as the application of non-traditional ways of judging students' performance.

## **Literature background**

According to the Oxford Advanced Learner's Dictionary (2012), "assessment is a carefully considered opinion or judgment" (p. 61), and "evaluation is an assessment of the amount, quality or value of something" (p. 394). Overall, both concepts are often used interchangeably, and their application is usually contextualized. It should be mentioned that the 'assessment' is commonly used with a variety of different meanings (Bachman, 2005, p. 7). In language research it is the process of collecting information about an object of interest. Moreover, an outcome of this process is also referred to as an assessment. Another term that is often associated with assessment is 'evaluation' (Bachman, 2005, p. 9). It involves making judgments and decisions. The use of assessment for evaluation is common in educational programs, when information is used to assign grades or marks. A conventional view of language assessment considers the notion of self-assessment. Self-assessment derives its theoretical justification from a number of well-established principles of second language acquisition (Douglas Brown, 2004). The principle of autonomy stands out as a primary foundation of successful learning. It is quite common among linguists to use terms 'self-assessment' and 'self-evaluation' interchangeably. There are various types of self-evaluation. Students typically monitor either their particular competence or specific performance. Researchers view self-evaluation as **informal assessment**. Self-assessment is closely linked to instruction because it can reveal weaknesses in instructional processes. Students' reflections offer the teacher an in-depth knowledge of the student as a learner and allow the teacher to individualize instruction for the student. The most common areas of student reflections noted by Nunes (2004) are: syllabus (7 %), instruction (36 %), learning (43 %), and assessment (14 %). The students' preferences in learning were investigated by examining their written reflections (Beckett & Slater, 2005). It was found that only 20 % of the 73 participants enjoyed project work; 25 % had mixed feelings, and 57 % perceived it negatively. Moreover, the high drop-out rate from the course existed because some students found the course too difficult or believed English classes should be limited to the study of language and resented being asked to accomplish non-linguistic tasks. It has been claimed that the challenges of assessment to language learners include lower comparability and reliability and difficulty ensuring standardized testing conditions. It was pointed out by D. Nunan (1988), there is a considerable extent of the mismatch between teacher and learner perceptions of the usefulness of different activities.

**Formal assessment** of students' learning is testing. It is the most common way to measure achievement and proficiency in language learning that traditionally carried out by teachers (Frank, 2012). Formal testing is a tool that

can help teachers identify students' strengths and weaknesses and evaluate the effectiveness of programs, i.e. how well students have learned what they were taught. In a discussion of language assessment, Bachman (2007) reviewed testing practices over the past five decades and categorized them into seven approaches: skills and elements, direct testing/performance assessment, pragmatic language testing, communicative language testing, communicative language ability, task-based performance assessment, interactional language assessment. A good example of progress in large-scale language self-assessments is recently reported DIALANG technique (Stoynoff, 2012). It is a self-directed assessment available for 14 languages including English that is accessed for free 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).

This research focuses on investigation of students' self-assessment of language skills, self-evaluation of their performance in class activities and its comparison with formal testing data.

### **Respondents**

The respondents in this study are the students of Klaipeda University (60 students, 2 samples), who studied English for Specific Purposes (ESP) in 2016. The students are mother-tongue speakers of Lithuanian and entered the university after having studied general English at secondary schools. The design of the ESP course reflects the students' future needs in professional language. The course is adjusted to the requirements for a Bachelor of Social Science degree. The level of students' proficiency is either B2 or C1 according to the Common European Framework of Reference for Languages.

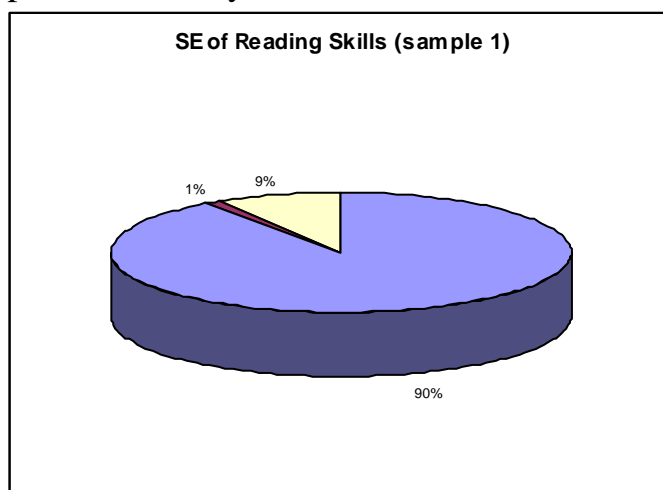
### **Research methodology**

The findings have been obtained from the survey completed by two samples of respondents. The questionnaire was designed in accordance with accepted standards of constructing surveys (Dornyei, 2003). It contained statements on students' self-assessment of their language skills, self-evaluation of their performance and evaluation of formal testing. The survey is presented in Appendix. It is comprised of the statements, to which students responded on a 5-point Likert's scale ranging from 1 "very difficult" to 5 "very easy". The responses were processed by a means of Software Package for Social Sciences (SPSS 19). Formal testing of students' performance by teachers aimed at clarifying the point of how realistic students are of their own evaluations.

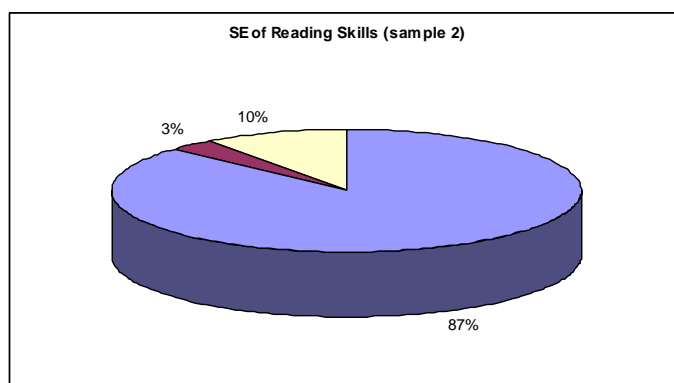
## **Results and discussion**

This part of the article reports the findings of the survey and analyzes the key points that emerged. The results of students' self-assessment of language skills are shown in pie Charts 1 to 11. It is essential to mention what is meant by positive, negative and neutral responses. The positive responses refer to "easy" and "very easy" points (i.e. numbers 4 and 5 of the questionnaire). Negative responses refer to "very difficult" and "difficult" points (i.e. numbers 1 and 2 of the questionnaire), and neutral responses refer to "unsure" (i.e. number 3 of the questionnaire).

Self-assessments of reading skills are displayed in two charts: sample 1 in the pie Chart 1, and sample 2 in the pie Chart 2. In spite of differences in percentages of responses, visually charts seem similar.



**Chart 1 Self-assessment of reading skills (Sample 1). Positive 90 %, negative 1 %, neutral 9 %**



**Chart 2 Self-assessment of reading skills (Sample 2). Positive 87 %, negative 3 %, neutral 10 %**

The pie Chart 3 and the pie Chart 4 show the students' self-assessment of listening skills. There is a noticeable difference between these charts: there are no neutral responses by the students of the 1<sup>st</sup> sample.

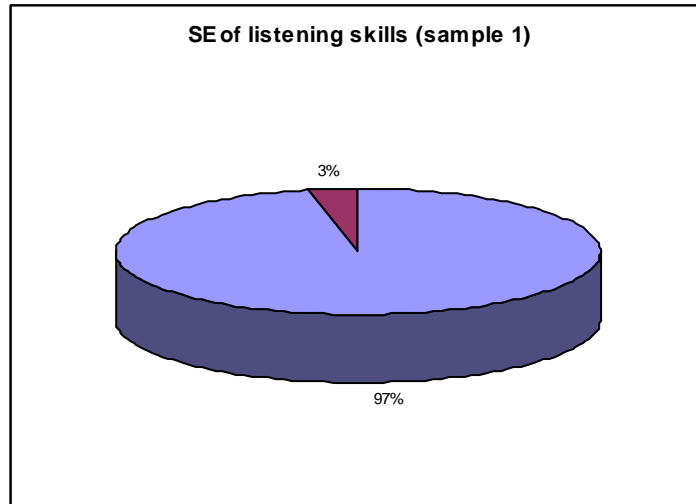


Chart 3 **Self-assessment of listening skills (Sample 1). Positive 97 %, negative 3 %**

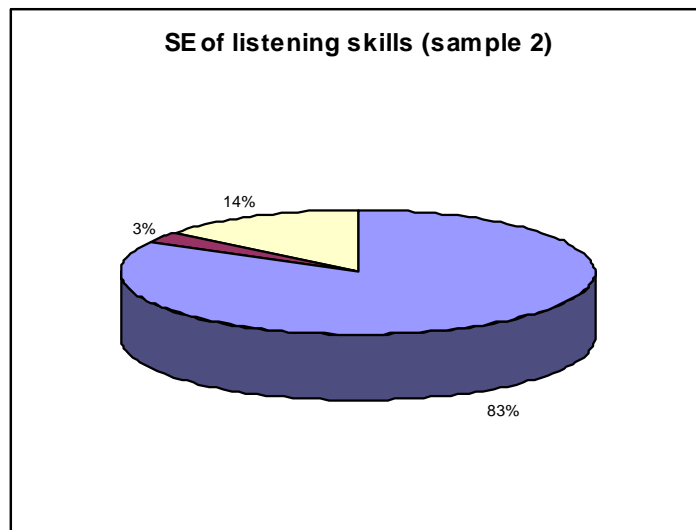
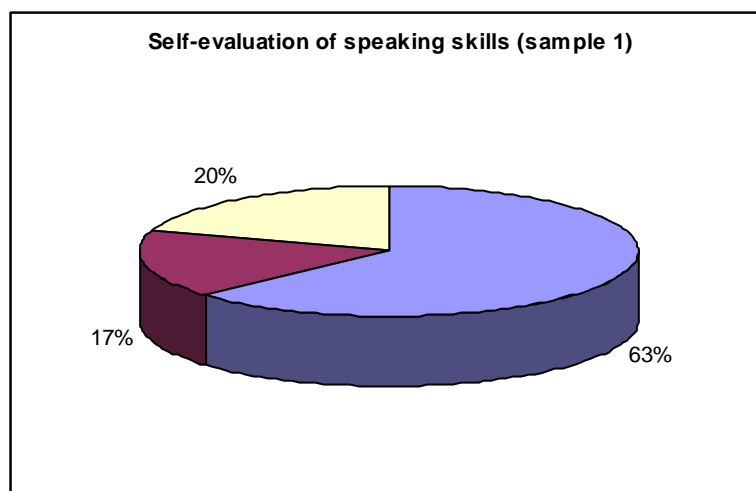
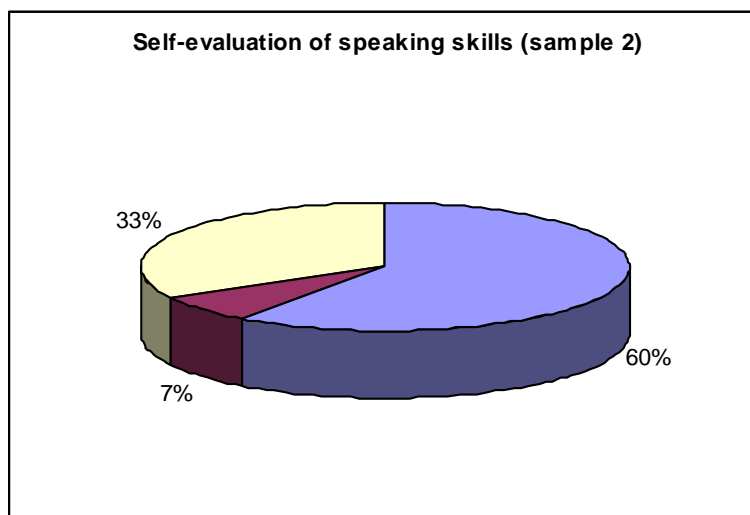


Chart 4 **Self-assessment of listening skills (Sample 2). Positive 83 %, negative 3 %, neutral 14 %**

The pie Chart 5 and the pie Chart 6 display the self-assessment of speaking skills. It can be seen that the percentage of positive responses is similar, but negative and neutral responses amount to quite a significant discrepancy.



**Chart 5 Self-assessment of speaking skills (Sample 1). Positive 63 %, negative 17 %, neutral 20 %**



**Chart 6 Self-assessment of speaking skills (Sample 2). Positive 60 %, negative 7 %, neutral 33 %**

The pie Chart 7 and the pie Chart 8 show self-assessment of writing skills. There does not seem to be essential differences between two samples. It is well known that to a considerable degree writing presents a number of difficulties to language learners, mainly due to the inability to get to the point and use authentic English.

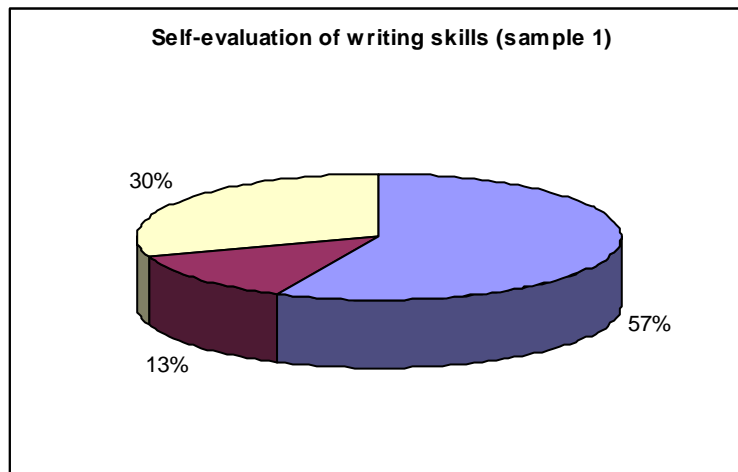


Chart 7 **Self-assessment of writing skills (Sample 1). Positive 57 %, negative 13 %, neutral 30 %**

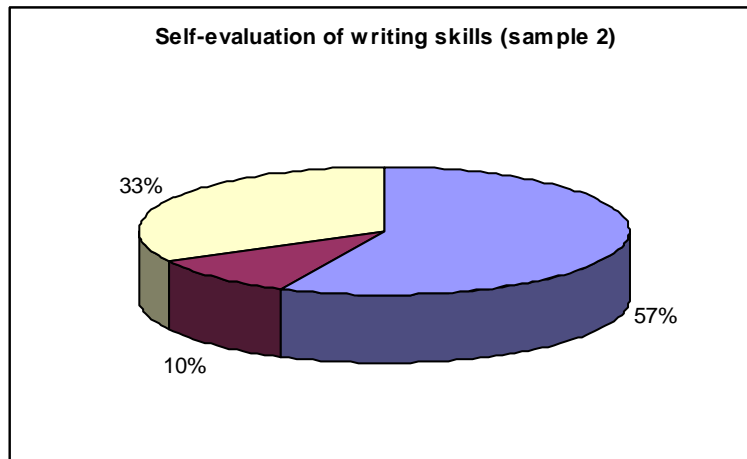
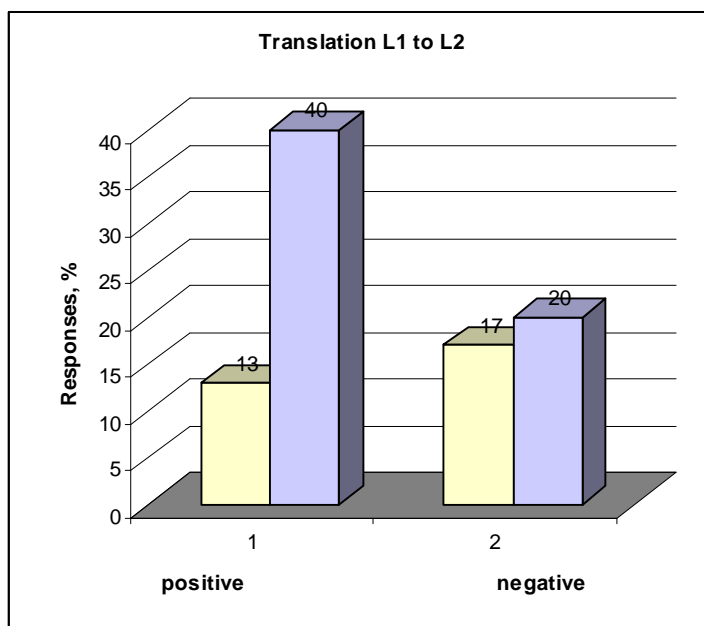
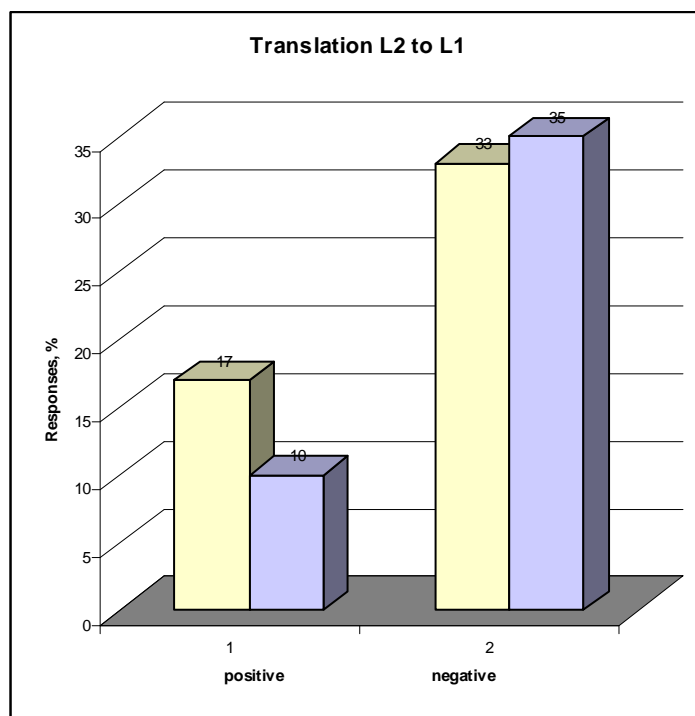


Chart 8 **Self-assessment of writing skills (Sample 2). Positive 57 %, negative 10 %, neutral 33 %**

Column Chart 9 and Chart 10 display students' self-assessment of translation skills. Chart 9 refers to translation from the mother tongue L1 to English L2, and Chart 10 refers to translation from L2 to L1. 1<sup>st</sup> columns in both charts show responses by the students from sample 1, and 2<sup>nd</sup> columns – from sample 2. It is quite obvious that translation from L1 into L2 is problem-oriented, and students are apparently aware of it.



**Chart 9 Self-assessment of translation from L1 into L2. 1<sup>st</sup> bars (13 % and 17 %, respectively) show responses by the respondents of sample 1, 2<sup>nd</sup> bars (40 % and 20 %, respectively) show the responses by the respondents of sample 2**



**Chart 10 Self-assessment of translation from L2 into L1. 1<sup>st</sup> bars (17 % and 33 %, respectively) show responses by the respondents of sample 1, 2<sup>nd</sup> bars (10 % and 35 %, respectively) show the responses by the respondents of sample 2**



Column Chart 11 displays the self-assessment of ESP vocabulary. 1<sup>st</sup> columns show responses by students from sample 1, and 2<sup>nd</sup> columns – from sample 2. It is evident that for the respondents of sample 2 the knowledge of vocabulary is much to be desired: only 20 % of students are positive, and 40 % assess it as negative one. To get a clear image of opinion differences between two samples of respondents, statistic processing is supposed to be beneficial.

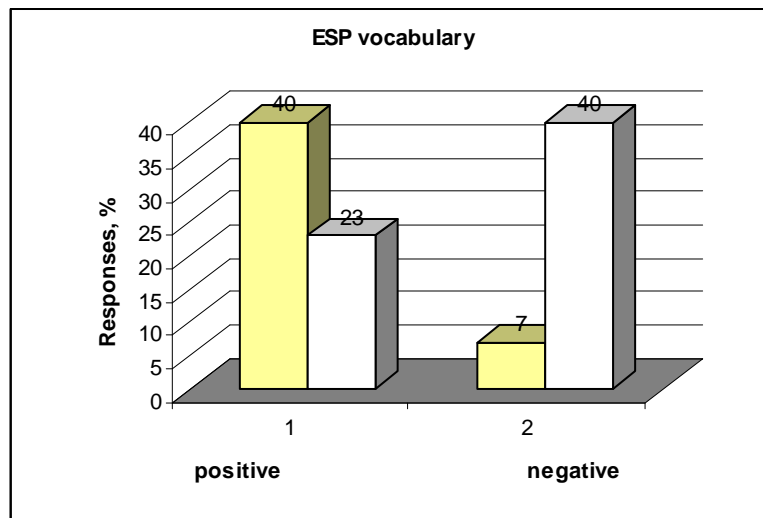


Chart 11 Self-assessment of ESP vocabulary. 1<sup>st</sup> bars (40 % and 7 %, respectively) show responses by the respondents of sample 1, 2<sup>nd</sup> bars (23 % and 40 %, respectively) show the responses by the respondents of sample 2

### Statistical processing

Statistic processing of the responses includes computation of the Means, the Standard Deviations and the Correlation coefficients. The Mean is the average score of the distribution. In a normal (Gaussian) distribution about 68 % of all scores are within one Standard Deviation of the Mean (34 % below and 34 % above). The Standard Deviation is the indicator of dispersion. It is probably the most commonly used indicator of variability (Bachman, 2005). Standard Deviation may help interpret scores when there is a perfectly normal distribution, which is seldom the case. Table 1 displays the Means and the Standard Deviations for classroom activities that have been shown in Charts 1 to 11 and described above. In Table 1, 1<sup>st</sup> column contains self-assessment statements; the 2<sup>nd</sup> column displays the Means and the Standard Deviations for sample 1, and the 3<sup>rd</sup> column – for sample 2.

Table 1 Likert's scale Means and Standard Deviations for both samples

self-assessment statements	Means & Standard Deviations, sample 1	Means & Standard Deviations, sample 2
1. Reading	4,1667 ,59209	4,3000 ,79438
2. Listening	3,7000 ,70221	4,1333 ,77608
3. Speaking	3,5667 1,00630	3,6000 ,81368
4. Writing	3,4667 ,77608	3,5333 ,86037
5. Translation L1 to L2	3,0333 ,61495	3,3000 ,87691
6. Translation L2 to L1	3,0000 ,52523	3,0000 1,01710
7. ESP vocabulary	3,3000 ,79438	2,8333 ,98553

It should be noted that the values of the Means in both samples are different except for the responses 'translation from L2 to L1'. The values of the Standard Deviations, which show the scattering of the data, also differ, in some cases quite noticeably. The pie Charts 1 to 8 seem similar visually. However this might be a misleading notion. It is essential to evaluate them statistically. Traditionally, Pearson's Correlation coefficients  $\rho$  and the Significance Levels  $Sig. p$  (2-tailed) are computed to determine if there are any correlations between the samples. Data of computations are presented in Table 2 below. It is seen that no correlations have been detected between samples for the self-assessment of reading, listening, speaking, writing, translation skills or ESP vocabulary. The large values of  $Sig. p$  (over 0.05) mean that all the probabilities of relationships are below the critical value of 95 %. For instance, for speaking skill the  $Sig. p$  is 0.076, i.e. probability is 92 %, while for other skills it is even much lower: 15 % for reading skills, 31 % for listening skills, 10 % for writing skills, etc. Moreover, the computed values of the Pearson's correlation coefficients  $\rho$  are also too small. For significant correlations, the values of the Pearson's correlation coefficients must be well over 0.5. The coefficient  $\rho$  with a 'minus' sign implies a reverse, not a direct relationship.

Table 2 Correlations data: Pearson's correlation coefficients  $\rho$  and Significance levels  $Sig. p$

Self-assessment statements	Pearson's correlation coefficients $\rho$	Significance levels $Sig. p$ (2-tailed)
1. Reading	-0.037	0.847
2. Listening	0.076	0.690
3. Speaking	0.328	0.076
4. Writing	-0.024	0.899
5. Translation L1 to L2	-0.275	0.141
6. Translation L2 to L1	0.129	0.497
7. ESP vocabulary	-0.198	0.294

The second part of the questionnaire (Appendix) refers to respondents' self-evaluation of grades and teacher's evaluation (formal testing) of their performance. It is well known that formal testing allows getting insights into the quality of students' learning. The obtained findings are summarized in Table 3 and Table 4.

**Table 3 Students' Self-evaluation vs. Teacher's Evaluation: Pearson's correlation coefficients  $\rho$  and Significance levels  $Sig. p$**

Self-assessment statements	Pearson's correlation coefficients $\rho$ Significance levels $Sig. p$ (2-tailed) Sample 1	Pearson's correlation coefficients $\rho$ Significance levels $Sig. p$ (2-tailed) Sample 2
1. Reading	0.458* 0.011 *. Correlation is significant at the 0.05 level (2-tailed).	0,667** 0.000 **. Correlation is significant at the 0.01 level (2-tailed).
2. Listening	0.873** 0.000 **. Correlation is significant at the 0.01 level (2-tailed).	0.072** 0.000 **. Correlation is significant at the 0.01 level (2-tailed).
3. Speaking	0.234 0.213 No significant Correlation is detected	0.637** 0.000 **. Correlation is significant at the 0.01 level (2-tailed).
4. Writing	0.404* 0.027 *. Correlation is significant at the 0.05 level (2-tailed).	0.301 0.106 No significant Correlation is detected
5. Translation L1 to L2	0.267 0.154 No significant Correlation is detected	0.521** 0.003 **. Correlation is significant at the 0.01 level (2-tailed).
6. Translation L2 to L1	0.113 0.553 No significant Correlation is detected	0.508** 0.004 **. Correlation is significant at the 0.01 level (2-tailed).
7. ESP vocabulary	0.346 0.061 No significant Correlation is detected	0.713** 0.000 **. Correlation is significant at the 0.01 level (2-tailed).

Students' self-evaluations of class activities versus teacher's evaluations are computed and presented in Table 3. The 1<sup>st</sup> column shows the list of class activities, the 2<sup>nd</sup> and the 3<sup>rd</sup> columns the computations for sample 1 and sample 2, respectively. The first numbers in these columns display the values of the Pearson's Correlation Coefficients  $\rho$ , and the second numbers show the Significance levels  $Sig. p$ . It may be seen that for sample 1 good correlations are obtained for reading activities (probability 95 %), for listening activities (probability 99 %) and writing activities (probability 99 %). For sample 2, good correlations (99 %) are obtained for all activities except writing. The findings might be interpreted as follows: respondents of sample 2 have better English language skills and are more realistic about their performance.

**Table 4 Grading performance: ratio of students' self-evaluation to teacher's (formal) evaluation**

<b>Self-evaluation grades of class activities</b>	<b>Sample 1 Ratio of self-evaluation to teacher's evaluation (%)</b>	<b>Sample 2 Ratio of self-evaluation to teacher's evaluation data (%)</b>
1. Reading	50%	57%
2. Listening	17%	53%
3. Speaking	17%	63%
4. Writing	37%	30%
5. Translation L1 to L2	27%	47%
6. Translation L2 to L1	27%	30%
7. ESP vocabulary	43%	50%

Table 4 allows getting insights into the comparative analysis of students' self-evaluation of grades in each activity and teacher's evaluation (formal testing) of their performance. Similarly as in Table 3, the 1<sup>st</sup> column presents the list of activities. The 2<sup>nd</sup> column and the 3<sup>rd</sup> column display the ratios of students' self-evaluation to teacher's evaluation in percentage for the 1<sup>st</sup> sample and the 2<sup>nd</sup> sample, respectively. It proves the data of computations shown in Table 3: performance of students in the 2<sup>nd</sup> sample is much better than in the 1<sup>st</sup> sample in all activities. The ratio of 50 % means that self-evaluation of grades by the half of the respondents coincides with the teacher's evaluation. Lower values of percentage, i.e. 17 %, imply that respondents' self-assigned grades are either lower or higher than assigned by their teacher. The interpretation of the differences between the samples might be due to the common phenomena of students either over-estimating or under-estimating their abilities (Dudley-Evans & Jo St John, 2000). However, overall it may be concluded that some of the learners are realistic about their language abilities and performance in linguistic tasks.

### **Conclusions**

The following conclusions have been drawn. Students' self-assessments of language skills in class activities demonstrate their perceptions of difficulties they usually face in the ESP classes. There is a significant difference between self-assessment responses and self-evaluation results. In the self-assessment responses, students estimate their language skills optimistically, while grading one's performance might be quite problematic. This point is clearly demonstrated by comparing students' self-evaluation data with teacher's evaluation: there is no total congruence between them - teacher's evaluation may differ from students'

self-evaluation. However, the respondents of sample 2 are more realistic in their self-evaluation than the respondents of sample 1, which is shown by the findings in Table 4. The interpretation of the differences might be due to the fact that some students tend to under-estimate or over-estimate their knowledge and linguistic ability. Nevertheless, seeking to improve the learning it is essential to raise students' awareness of their achievements by encouraging them to self-assess their language skills and self-evaluate their performance in class activities.

### References

- Bachman, L. F. (2005). *Statistical Analyses for Language Assessment*. Cambridge University Press.
- Bachman, L. F. (2007). *What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment*. Cambridge University Press.
- Beckett, G. H., & Slater, T. (2005). The Project Framework: a Tool for Language, Content, and Skills Integration. *ELT Journal* 39/2: 108-116.
- Douglas, D. (2000). *Assessing Language for Specific Purposes*. Cambridge: Cambridge University Press.
- Douglas Brown, H. (2004). *Language Assessment. Principles and Classroom practices*. Pearson Education, Inc.
- Dudley-Evans, T., & Jo St John, M. *Developments in ESP*. Cambridge University Press.
- Frank, J. (2012). The Roles of Assessment in Language Teaching. *English Teaching Forum*. No 3. p.32.
- Dornyei, Z. (2003). *Questionnaires in Second Language Research*. Lawrence Erlbaum Associates, Inc.
- Kavaliauskienė, G. Students Reflections on Learning English for Specific Purposes. [http://www.esp-world.info/Articles\\_15/ESP\\_REFLECTIONS\\_ON\\_LEARNING.htm](http://www.esp-world.info/Articles_15/ESP_REFLECTIONS_ON_LEARNING.htm)
- Nunan, D. (1988). *"The Learner-Centered Curriculum"*. Cambridge University Press.
- Nunes, A. (2004). Portfolios in the EFL Classroom: Disclosing an Informed Practice. *ELT Journal* 58/4: 327-335.
- Oxford Advanced Learner's Dictionary. (2012). Oxford University Press.
- Stoyhoff, S. (2012). Looking Backward and Forward at Classroom-based Language Assessment. *ELT Journal*, 66 (4), pp. 523-532.

### Appendix. Self-assessment of language skills.

#### Circle the right answer.

- Reading is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 Listening is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 Speaking is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 Writing is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 Translation from L1 to L2 is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 Translation from L2 to L1 is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.  
 ESP vocabulary is 1) very difficult, 2) difficult, 3) not sure, 4) easy, 5) very easy.

**Self-evaluation of performance and formal evaluation**

**Write the self-evaluation grade**

Class activities	Self-evaluation grade	Teacher's evaluation grade
Reading		
Listening		
Speaking		
Writing		
Translation from L1 to L2		
Translation from L2 to L1		
ESP vocabulary		