

Review research paper

## PEER AND TEACHER ASSESSMENT OF ACADEMIC ESSAY WRITING: PROCEDURE AND CORRESPONDENCE

Jelena Andelković

University of Belgrade, Faculty of Organizational Sciences, Serbia

**Abstract.** *This paper reports on the assessment procedure of students' academic essays at an undergraduate course in English for Academic Purposes (EAP) at a business-oriented university during the school year 2020/21. Following the social constructivism paradigm (e.g. Vygotsky, 1962), the procedure employed peer assessment (PA) and teacher assessment (TA) of students' academic essays and was aimed at improving their essay writing performance. The paper further reports on the degree of correspondence between the grades awarded by teachers and peer assessors on a set of assessment criteria to investigate the potential of PA as a (a) learning tool and as a (b) supplementary assessment tool. Lower correlation and higher difference between mean grades awarded by teachers and peers on some of the assessment criteria may indicate the essay writing aspects students are weakest at. The results also implicate that certain adjustments in the assessment procedure need to be made in future iterations of the course, particularly with regard to assessment training, defining assessment criteria, and pairing multiple peer raters with a single teacher rater, as these changes may not only improve the benefits PA can bring, but also contribute to its validity.*

**Key words:** *peer assessment, teacher assessment, academic essay, English for academic purposes*

### 1. INTRODUCTION

In the ever-changing and diversified context of higher education, new assessment practices have been gaining prominence. Up to approximately 20 years ago, traditional assessment practices mainly focused on summative assessment, grades and marks awarded by teacher assessors only. With the emergence of constructivist paradigm, strongly influenced by complementary work of Piaget in psychology and Vygotsky in education, also came new assessment culture (Birenbaum, 2003:22). Priorities have been shifting away from summative towards formative assessment, or, in words of Havnes and McDowell (2008:7) from assessment *of* learning towards assessment *for* learning. Accordingly, students have become active participants in the assessment process, by assessing their own or the work of their peers, and thus using assessment as a “positive tool for learning” (Havnes, McDowell, 2008: 5). The greatest significance of *peer assessment (PA)* in education, according to Van Zundert et al. (2010: 270), lies in the fact

---

Submitted February 17<sup>th</sup>, 2022, accepted for publication April 5<sup>th</sup>, 2022

Corresponding author: Jelena Andelković. University of Belgrade, Faculty of Organizational Sciences, Serbia.

E-mail: jelena.plecas@fon.bg.ac.rs

that it empowers learners to become autonomous, actively participate in their own learning, and to collaborate with other learners. These three goals – empowerment, learner-centeredness, and collaboration in learning, are not only the products of PA, but also some of the major pillars of contemporary constructivism-based theories of learning.

Considering the benefits it may bring, we decided to introduce PA as a supplement to teacher's assessment (TA) of students' academic essays in an undergraduate EAP course at a business-oriented university. The procedure was carefully planned and implemented online during the summer semester of school year 2020/21. This paper outlines the assessment procedure and investigates the potential correspondence between TA and PA of students' written assignments with the aim of examining the potential of PA both as a learning tool and as a supplementary assessment tool.

The paper is structured as follows: after the Introduction, Section 2 places this study in a theoretical perspective by outlining the aspects of social constructivism-based theories of learning relevant for this paper, the concept of peer assessment, its use and reliability and validity in the context of teaching academic writing. Section 3 provides an outline of the study, including the information on the assessment procedure methodology and the results of measuring the correspondence between PA and TA. Discussion (Section 4) draws conclusions regarding the implemented assessment procedure and its advantages and limitations compared to similar studies. The last section summarizes the study, points to areas for improvement and offers suggestions for further research.

## 2. LITERATURE REVIEW

Constructivism, currently one of the leading theoretical positions in education, brought about a significant shift in the understanding of the nature of learning and teaching (Cooper 1993: 12). Constructivism-based learning theories maintain that learning is not a passive process and thus promote learner-centered instructional methods that encourage learners to be independent, autonomous, responsible and actively involved in the construction of knowledge (Hoover, 1996), Twomey Fosnot, 1989, Hein, 1991: 1–12, and Kiraly: 2005, 2012). Students are expected to be actively involved in the construction of knowledge, the role of teachers has become to support learners, promote their motivation and critical thinking rather than just to transmit knowledge. Social constructivism, drawing on the work of Lev Vygotsky (1886-1934) shares the common constructivist principles and beliefs, but also maintains that learning is a social process that occurs only when learners interact and collaborate with others, while knowledge is a social and cultural, rather than individual construct (Ernest, 1999: 4-5; Prawat & Floden, 1994: 37)

Constructivism and social constructivism have brought significant changes to the concept and practice of assessment. Since constructivist theories view learning as an “active process of sense making, instead of passive reception of knowledge, assessment tasks cannot just test reproduction, but must give space for production of knowledge” (Havnes, McDowell, 2008: 20). New modes of assessment, such as peer assessment, performance assessment, learning logs, self-assessment, etc. (Havnes, McDowell, 2008:6) have gradually gained importance, as they make learners active and responsible for their own learning, assessment criteria transparent, and the whole learning environment balanced by aligning and integrating learning, teaching and assessment (Havnes, McDowell, 2008: 6-7).

## 2.1. Peer assessment

Peer assessment - PA (also known as peer review, peer evaluation or peer feedback, as in Zhang et al. 2020: 1) is the process of students providing feedback (and/or grades, though not as frequently) on a product or a performance of their peers (Falchikov, 2007: 132). PA can take many different forms (it may include assessing writing, portfolios, oral presentations, test performance, etc.), be formative or summative, involve a variety of methods and outcomes, be applied under various conditions and in different places (e.g. in classroom or online, face-to-face, written, etc.), be one-way, reciprocal or mutual (Topping 2003: 65), and include assessors and assessees who are, more or less, equal in status (Topping 2009:21). The main reasons for introducing PA, according to Topping (2003: 56), are to improve assessment quality or reduce its costs, since students, though less skilled, have more time and may produce equally reliable and valid assessment.

To be successful, PA needs to be carefully planned and executed. Topping (2009: 25-26), for example, lists a number of steps in organizing PA: a) developing initiative in cooperation with colleagues rather than alone, b) specifying the purpose, rationale and expectations, c) involving students in identifying assessment criteria, d) matching students by ability, e) providing training and practice, f) providing guidelines and other scaffolding, g) specifying activities and timeline, h) monitoring and coaching, i) examining the quality of PA against TA, j) moderating the validity and reliability of PA (if more PAs of the same work are available), and k) giving feedback to students assessors so they can improve their work. Similarly, Falchinkov and Goldfinch (2000: 317), based on meta-analysis of 48 PA studies in various discipline areas, provide recommendations for implementing PA using marks and grades. When student assessors are expected to grade the work of their peers, according to these authors, teachers should: a) avoid using very large numbers of peers per assessment group, b) ask their students to use an overall global mark rather than to expect from them to rate many individual dimensions, c) involve their students determining the criteria, d) attentively design and implement PA, etc., and they will be able to implement PA successfully in any discipline and at any level. Falchikov (2007: 132-133) also emphasizes that student involvement in identifying assessment criteria makes the assessment practice both transparent and beneficial.

In recent years, constructivism-based learning theories have renewed the interest in PA, especially in formative PA, and seen its potential both as a learning tool and an assessment tool (Van Zundert et al. 2010:270).

### 2.1.1. PA as a learning tool

If used as a learning tool, PA can bring many benefits to both student assessors and assessees: it can help them identify their own strengths and weaknesses, recognize gaps in their knowledge and target areas for remedial action, make judgments about high quality work, develop meta-cognitive and other personal and professional skills (Topping 2009: 26), improve social competencies and develop intellectually (Falchinkov 2007: 133). Additionally, peer evaluation helps reduce “the power imbalance between teachers and students and can enhance the students’ status in the learning process” (Spiller, 2012:11), changes the teacher role by shifting focus from teaching to learning (Havnes, McDowell 2008: 6), etc. Despite initial anxiety assessees may feel upon acceptance of negative feedback (Topping 2003: 67), peer assessment helps peer assessees improve

their formative learning by encouraging them to clarify, review and edit their ideas during the assessment process.

The implementation of peer assessment procedures has proved to be particularly effective in evaluating students' academic writing across disciplines. Topping (2003: 70-71, 73), for example, provides an overview of papers dealing with PA of writing assignments in schools and HE institutions, mostly focusing more on its (positive) effects, substantial improvements in the effectiveness and quality of learning and subsequent learner performance, and concludes that the improvements are at least as good as the ones gained from teacher assessment (see, for example, Chaudron, 1983 and Birkeland, 1986 in Topping 2003: 77). A study similar to the one presented in this paper was conducted by Graner (1985): it compares final grades of students who rewrote their writing assignments after the initial draft was assessed by their peers, with the grades of students who did not get such feedback.

### 2.1.2. PA as an assessment tool

The benefits of using PA as an assessment tool have been somewhat debatable. Spiller (2012:2), for example, maintains that teachers, who otherwise follow the principles of constructivist learning, tend to restrain from involving students into the design and implementation of assessment tasks, but rather abide by traditional, teacher-directed forms of assessment. The competence of students to assess their peers has been investigated by a number of authors (Conway et al. 1993; Oldfield & MacAlpine, 1995; Sivan, 1996; Cheng & Warren, 1999; Kaufman & Schunn 2011). Competence is not questioned by teachers only, but by peers themselves, who challenge their own competence in relation to awarding marks, or express concerns about marking fairly and responsibly (e.g., Sluijsmans et al. 2001). Other limitations regarding students' competence in assessment involve lack of training and practice in assessing and concerns about determining criteria for assessing the work of their peers (Falchikov, 2007: 134). Peer assessment may also be potentially biased and partly influenced by social factors such as "friendship bonds, enmity or other power processes, group popularity levels of individuals, perception of criticism as socially uncomfortable or even socially rejecting and inviting reciprocation, or collusion leading to lack of differentiation" (Topping 2003:67).

Another concern about using PA seems to be the issue of its *accuracy*, i.e. its *reliability and validity*. Even though the terms reliability and validity are sometimes confused and used interchangeably (Topping, 1998: 257), Falchinkov and Goldfinch (2000: 288) maintain that *reliability* of PA is measured against the assessment of other peers, while *validity* is measured against teacher's assessment. Even though some authors see these as a concern, others claim that, if well planned and implemented, "peer assessment offers triangulation and *per se* seems likely to improve the overall reliability and validity of assessment" (Topping 2003: 68; Topping 2009:25). Since this paper outlines the use of PA as a supplement to TA, we will focus more on the issue of *validity*, i.e. "degree of correspondence between student peer assessments and the assessments made of student work by external "experts" such as professional teachers" (Topping 2003: 68), or "the extent to which students can accurately judge what they are asked to assess (...)." (Schunn et al. 2016: 2).

In most papers, it is generally assumed that the assessment made by teachers (i.e. experts) is highly valid (Topping, 2009: 25) and is therefore taken as a golden standard against which PA validity is measured using various statistical tests, such as Pearson

correlation coefficient, standard deviation, means, etc. Cho et al. 2006, however, questions the invincibility of teacher assessment, claiming that validity and reliability have to be observed from both instructors' and students' perspective, as these views may even be contradictory: i.e. while teachers consider an assessment valid and reliable, students may find the same one unreliable and invalid (Cho et al. 2006: 891-892), and list theoretical and practical reasons why peer evaluations can be just as reliable and valid as instructor ratings are.

An issue regarding the correspondence between TA and PA is the number of peer ratings against which a single teacher rating should be measured. While earlier papers (Cheng & Warren, 1999; Falchikov, 1986; Mowl & Pain, 1995; etc.) mostly establish correspondence of a single PA with a TA of the same work in order to determine the validity of PA, newer research (Cho et al 2006, Chang et al 2010, Schunn et al. 2016) compares the assessment of a single piece of work done by multiple peers to the one of a single teacher rater as it results in higher validity and reliability and may narrow down the gap between teacher and peer evaluation. Cho et al. (2006: 892) maintain that measuring one peer rating against one teacher rating is not optimal as it "confuses agreement with real ability to detect quality", and suggest that combined assessment of multiple peer raters (at least four) of a single piece of writing can be even more valid and reliable than a single teacher assessment.

Most research on the usage of PA in evaluating writing proves high or adequate validity, i.e. correspondence with TA (see, for example, Cho et al. 2006). Chang et al. 2010, however, report on using Web-based portfolio PA in high schools resulting in low level of reliability and validity and suggests that it may be improved by, among other methods, more advanced training that would provide students with more in-depth understanding of their task. In the context of assessing students' writing in foreign language learning, Zhang et al. (2020) study peer reviews of students' writing at the undergraduate level at a Chinese university over years and discover that the assessment of simpler language conventions (such as grammar or spelling) shows lower validity on the 1<sup>st</sup> year of study, while the validity of assessing higher-level dimensions of language is high regardless of the year of study.

The correspondence between PA and TA seems to depend on a number of variables, such as the context, level of the course, clarity of criteria, product or performance that is being evaluated, etc. (Topping 2003: 69). For example, correspondence is higher in higher level courses, and when assessment criteria were negotiated and discussed with students before the evaluation itself (Topping 2003: 69), and in sciences and engineering rather than in social sciences (Falchikov, Boud 1989: 424). PAs are generally more valid when supported by training, checklists, exemplification, teacher assistance, and monitoring (Topping 2009:25), and when the assessment process is structured and with clear rubrics (Cho et al. 2006: 893). Also, peer and teacher assessments tend to resemble more in the overall judgment of a product or process being assessed when they are asked to assess several individual dimensions of the product or process in question (Falchikov, Goldfinch, 2000: 287).

### 3. THIS STUDY

### 3.1. Context

English for academic purposes (EAP) has been a part of the curriculum at our business-oriented university since the school year 2017/18 as an elective course in the 8<sup>th</sup> semester of undergraduate studies. Even though titled English for Specific Purposes 3 (ESP 3, in addition to mandatory ESP 1 and ESP 2 offered early during undergraduate studies), its focus has been academic English or, more specifically, writing for general and specific academic purposes, and it was aimed at students who intended to pursue their academic careers either in the country or abroad. During the four iterations of the course, the syllabus and teaching methods have been adapted several times to suit the students' needs discovered in needs analyses and course satisfaction surveys (Meršnik & Anđelković 2019).

The initial conception to include several academic genres in the academic writing course was abandoned as it proved to be too overwhelming and time-consuming for students, while the actual achievement was not completely satisfactory. Consequently, during the last two iterations of the course (in school years 2019/20 and 2020/21) the focus was transferred to academic essay writing only. In both 2019/20 and 2020/21, during the online classes (introduced due to COVID-19 pandemic) students were exposed to various samples of academic reading texts belonging to diverse genres (research articles, reports, excerpts from academic course books, etc.) on various general academic topics and were guided, through scaffolded instruction, to produce segments (introductions, bodies, conclusions) of argumentative, descriptive, and narrative academic essays related to these topics. In 2019/20, the final exam consisted of a general academic vocabulary test (40% of the grade) and an academic essay writing assignment (60% of the grade) related to general academic topics covered during classes. Both the test and the written assignment were graded by EAP teachers; the assessment was therefore summative rather than formative and performed by the EAP teachers only.

Several changes in the type and the assessment of students' written assignments were made in the school year 2020/21. While the course syllabus, teaching materials, methods, and grading (60% for the essay and 40% for the test) remained the same, students were now able to submit their academic essay assignments during the semester rather than to write them during the exam session (the academic vocabulary test was still to be done in the exam period). This enabled students to reflect more on the topic they selected, carefully choose references, invest more time in writing, and consequently produce a higher quality work. Additionally, the topics students could choose from were more specifically related to students' fields of studies rather than general academic ones, and therefore expected to be more relevant for students' interests and future academic career.

In addition to the abovementioned, another significant change was the introduction of peer assessment of students' academic essays, alongside teachers' assessment, and the opportunity to submit corrected versions of the essays based on the first draft assessment if the author considered it necessary. Students – peers were not involved in the grading of the essays; peer assessment was rather used as a learning tool for both peer assessors and peer assessees.

The following sub-section will provide more details regarding the procedure of academic essay assignment, submission, assessment and resubmission in our EAP course.

### 3.2. Procedure

In the spring semester of 2020/21, the total of 53 fourth-year undergraduate students chose the course in EAP, while 36 out of them (68%) decided to submit the essay assignment during the semester rather than to write it in the exam session.

The 36 students were sent a link to a Google drive document containing a list of 71 topics for narrative, argumentative, and descriptive academic essays related to their field of studies and interests expressed in the needs analysis questionnaire conducted during the first class in the semester. To make the list more relevant to our students, the essay topics were designed with the assistance of content teachers at our university. Apart from choosing a topic of interest, each student was allowed to suggest changes and adapt the topic even more to his/her interests and knowledge; none of the students, however, used this opportunity. Students were also informed that the essays will be assessed by their peers anonymously, but that their final grade will not be dependent upon it (as also suggested by Topping 2009: 24).

The students were also presented with *Essay writing and assessment timeline* (Table 1) and sent a short *Essay writing checklist* (Fig. 1) to guide them through the writing process. The checklist only summarized the essay writing guidelines and instructions taught throughout the semester using various types of reading and writing exercises, assignments and numerous examples. There were no detailed technical essay writing guidelines as this was not considered important; students were only asked to submit the file in *.pdf* format, rename the document (*NameSurnameIDnumber*), and not to put their personal information inside the document itself, as this facilitates the handling of the document and anonymization prior to sending it for peer assessment.

Table 1 Essay writing and assessment timeline

<i>Week 1 - 2</i>	Teachers provide general information about the essay assignment
<i>Week 3</i>	Students choose a topic from the list provided
<i>Week 3</i>	Teachers send the Essay writing checklist to students
<i>Week 4 - 6</i>	Students write and submit the 1 <sup>st</sup> version of the essay
<i>Week 7</i>	Teachers check students' work for plagiarism, make it anonymous, randomly match and send to students for PA together with the assessment form
<i>Weeks 8 - 10</i>	Both students and teachers work on the assessment and send the TA and PA forms to assesses / authors
<i>Week 11 - 12</i>	Assesseees / authors rewrite and resubmit their essays based on TA and PA received
<i>Week 13</i>	Teachers hold focus group interviews with participants in the essay writing and assessment activity

<i>BEFORE SUBMITTING THE ESSAY, MAKE SURE YOU:</i>	
<ul style="list-style-type: none"> <li>▪ Choose at least two relevant sources pertaining to the topic and use it for your initial research</li> <li>▪ Use in-text references or paraphrases of your essay. <b>DO NOT</b> plagiarize!</li> <li>▪ List the references in the reference list following the essay</li> <li>▪ Brainstorm and draft an outline of your essay, reread and redraft the first version</li> <li>▪ Stick to the essay type printed in the brackets next to the topic (defend, describe, discuss) and its structure</li> <li>▪ Use adequate academic vocabulary and formal language</li> <li>▪ Be careful of the structure of each paragraph – topic sentence, supporting evidence, concluding sentence.</li> </ul>	

Fig.1 Essay writing checklist

Following the submission of the first version of the essay (approximately three weeks after the topics had been chosen), the submitted documents were checked for plagiarism using *Turnitin* plagiarism detection software; a third (12 out of 36) was returned to the authors for revision as these essays displayed more than 10% similarity with various sources (excluding direct quotes). They were asked to resubmit the corrected essay versions before they are forwarded it for assessment.

### 3.3. PA and TA methodology

Upon collection of all the 36 essays (including the ones resubmitted after plagiarism check), students' names and other personal information were made anonymous and each student and his/her essay was assigned a number (01-36).

An assessment form, previously created in Microsoft forms by adapting the one devised in Mowl & Pain (1995: 329) was used for both teacher assessment and peer assessment of students' work. The assessment form consisted of 13 statements, or 13 assessment criteria, which both peer and teacher assessors were expected to mark on a 5-point Likert scale (1-lowest, 5 – highest grade). The statements were divided into three categories: structure and content (statements 1-8), language (9, 10, 11), and references (12, 13).

Students were randomly matched for peer assessment and not familiar with the grades teacher assessors awarded for the same work. The two EAP teachers divided the essays into two groups, so each teacher assessed 18 essays. Each student assessor was sent an anonymized essay and a link to the assessment form, and the information on the peer assessment deadline. Students were informed that the submission of the filled-out form does not bring them additional points, but is a precondition for being graded in the course. No additional instructions about the form were provided, but students were encouraged to approach teachers with questions if needed.

Upon deadline expiry, the two teachers collected the assessment forms, sent each student both teacher and peer feedback on his/her work, and informed the students about the possibility of resubmitting their work after making alterations based on the two assessment forms. Thirty-two out of 36 students decided to submit the second version of the essay for grading, while only four students did not find this necessary.

At the end of the course, two focus group interview sessions were organized to learn about students' views on this kind of assessment procedure, its effectiveness, and ways of adapting it for future generations.

### 3.4. Results

The same assessment form was used by both teachers and peers. The average time teachers took to complete the form for each essay was 28 minutes and 36 seconds, while peers took 21 minutes and 21 seconds. The total of PAs and TAs submitted was 36, corresponding to the number of students who submitted their essays.

To measure the correspondence between PA and TA, the replies to the closed questions in the assessment form were compared using means ( $\bar{x}_1$  and  $\bar{x}_2$ ), standard deviations (SD1 and SD2) and Pearson correlation coefficient ( $r$ ) (Table 2).



Table 2 PA and TA correspondence

No	Criterion	N	PA		TA		PA - TA correlation	
			PA mean ( $\bar{x}_1$ )	PA standard deviation (SD1)	TA mean ( $\bar{x}_2$ )	TA standard deviation (SD2)	PA-TA means difference ( $\bar{x}_1 - \bar{x}_2$ )	Pearson correlation coefficient (r)
1	The essay has a clear and logical structure (there is an introduction, middle section(s), and a conclusion)	36	4.63	0.76	4.36	0.79	0.27	0.18
2	The introduction provides a clear outline of what the essay is all about.	36	4.53	1.25	3.36	0.73	1.17	-0.11
3	The conclusion draws together the important points made in the middle sections of the essay.	36	3.95	1.48	2.97	1.11	0.98	0.36
4	The essay clearly answers the essay title.	36	4.5	1.16	4.17	0.73	0.33	0.25
5	The essay structure is appropriate for the essay type in question (describe, discuss, or defend).	36	4.55	1.36	3.53	0.69	1.02	0.23
6	Each paragraph represents one idea / point of view.	36	4.34	1.14	4.11	0.88	0.23	0.32
7	Each paragraph contains a topic sentence.	36	4.39	1.27	3.33	0.86	1.06	0.24
8	Each paragraph contains a concluding sentence.	36	4.16	1.29	2.78	1.05	1.38	-0.02
1-8	<b>STRUCTURE AND CONTENT</b>	36	4.31	1.21	3.57	0.85	0.74	0.18
9	The essay is generally readable, interesting and well presented.	36	4.18	0.89	4.19	0.80	-0.01	0.23
10	Grammar and spelling are used properly.	36	4.32	0.66	4.28	0.74	0.04	0.28
11	The author uses appropriate (academic) vocabulary.	36	4.45	0.93	3.78	0.80	0.67	0.42
9-11	<b>LANGUAGE</b>	36	4.32	0.83	4.08	0.78	0.24	0.31
12	The essay draws upon relevant literature.	36	4.63	1.43	2.67	0.83	1.96	0.30
13	The sources used in the essay are acknowledged by properly and consistently using one style of referencing.	36	4.34	1.29	2.33	1.10	2.01	0.51
12-13	<b>REFERENCES</b>	36	4.49	1.36	2.5	0.96	1.99	0.41
<b>TOTAL</b>		36	4.37	1.13	3.38	0.86	0.99	0.30

The comparison between grades awarded by teachers and peers shows that the mean grade awarded by peers ( $\bar{x}_1$ ) is consistently higher than the one awarded by the teachers ( $\bar{x}_2$ ); the overall difference between the means is ( $\bar{x}_1 - \bar{x}_2 = 0.99$ ). Individual criteria, or sets of criteria, however, differ in TA-PA means difference, with the criteria set C, related to

the relevance of used literature and proper referencing, showing the greatest difference ( $\bar{x}_1 - \bar{x}_2 = 1.99$ ), and set B, related to language use), showing the biggest similarity ( $\bar{x}_1 - \bar{x}_2 = 0.24$ ). When we look at the criteria individually, the mean grades seem to be almost equal for criteria that refer to the overall judgment of the essay (1 - the structure of the essay, 4 - the connection between the essay title and the essay content, 6 - the coherence and consistency of each paragraph, 9 - general impression of the essay's readability, and 10 - the use of grammar and spelling). There is approximately one grade difference in the assessment of the contents of introduction and conclusion (criteria 2 and 3), the appropriateness of the essay type chosen (5), the existence of topic sentence (7), and slightly more similar mean grade for the use of academic vocabulary (11).

In order to look at the correspondence between the PA and the TA of academic essays more closely, we have calculated the Pearson correlation coefficient ( $r$ ) for each of the criteria and set of criteria in the assessment form. On the average, there is a low positive correlation ( $r = 0.3$ ) between TAs and PAs. When we look at sets of criteria, the highest correlation (though still low) is shown with set C ( $r = 0.41$ ), and slightly lower with sets B ( $r = 0.31$ ) and set A ( $r = 0.18$ ). If we observe individual criteria separately, a very low negative correlation is detected with teacher's and students' grades regarding the structure of the introduction (no. 2,  $r = -0.11$ ) and the existence of concluding sentences in each paragraph (no. 8,  $r = -0.02$ ), while the highest correlation between TAs and PAs is related to the consistent use of a referencing style (no. 13,  $r = 0.51$ ).

With regard to SD values, teachers' grades generally show higher values for all criteria, apart from no. 1 and no. 10, where SDs are almost equal. When we observe sets of criteria, both TAs and PAs show the lowest SD with regard to set B - the language of the essay, and the highest with set C - proper referencing. In terms of individual criteria, the lowest standard deviation in TA is the one regarding the use of grammar and vocabulary (criterion 10,  $SD_2 = 0.659$ ), while for PA it is the one regarding the essay type (criterion 5,  $SD_1 = 0.686$ ). The highest dispersion of teacher's marks regards the structure of the essay conclusion (no. 3,  $SD_2 = 1.483$ ), and for students it is the way sources are acknowledged in the essay (no. 13,  $SD_1 = 1.097$ ). The highest similarity between SDs in teachers' grades and peers' grades was for criteria related to the proper use of language (criteria 9, 10, 11), essay structure (criterion 1), and proper and consistent referencing (criterion 13). The lowest similarity between SDs is observed in grades related to the content of the introduction (no. 2), essay type (no. 5), the relevance of literature (no. 12).

#### 4. DISCUSSION

Generally speaking, the results presented above indicate the students tend to give their peers higher grades than the teachers, as indicated by the overall means and SD values above. High grades may be the result of students feeling less competent (as also indicated by Sluijsmans et al. 2001, Falchikov 2007: 134), or being biased and hesitant to criticize (Topping 2003: 67), so they tend to grade their peers highly. Students' grades are also less diversified, which corresponds to Topping's study (2009:24) which proved that TAs show a greater range of grades than PAs.

The comparison between mean grades in TAs and PAs for the three sets of criteria shows that the most significant grade difference is for set C (Referencing), and the lowest is for set B (Language). In terms of individual criteria, grades tend to be more similar

when the criterion is more general and refers to the overall judgment of the essay (for example, its structure and readability) or language conventions (e.g. proper use of grammar and spelling), as also indicated by Zhang et al. 2020. On the other hand, the more significant mean grade difference is for the criteria are related to more specific aspects of the essay (e.g. the appropriateness of vocabulary and literature, proper referencing) and/or to the ones that require assessors to perform more cognitively demanding tasks (such as deciding whether the introduction and conclusion draw together relevant information or whether the essay content corresponds to the required essay type). This conclusion is in line with the one made by Falchikov & Goldfinch (2000: 287), who maintain that TAs and PAs “tend to resemble more in the overall judgment of a product or process being assessed, then when they are asked to assess several individual dimensions of the product or process in question”.

The Pearson correlation coefficient indicated low positive correlation between the mean grade awarded by teachers and the one awarded by peers for 11 out of 13 criteria for assessment. Even though such results indicate that the method of measuring validity of peer assessment is not completely optimal (as also suggested by Cho et al. 2006: 892), the detected correlation indicates that the use of PA in our EAP course has potential to be used in the future if certain deficiencies are corrected. It is interesting, though, that the highest detected correlation refers the consistent use of a referencing style, having in mind that the mean grade awarded by teachers and peers for this criterion (no. 13) is the highest of all.

## 5. CONCLUSION

The present study reports on the use of peer assessment of academic essays as a supplementary activity in an EAP course at a business-oriented university during the school year 2020/21. The paper outlines the assessment procedure and reports on the correspondence between teachers' and peers' grades on a set of criteria comprising an online assessment form using several statistical measures, with the aim of investigating the potential of PA as a learning tool and a supplementary or substitute assessment tool. We also hoped that the procedure would help us discover the aspects of essay writing the students are weakest at, so these aspects can be addressed in the future iterations of our EAP course.

Based on the comparison between data gained from TA and PA, we can indirectly infer what students' weak points in essay writing and grading are, or what instructions teachers failed to provide prior to or during the essay writing assignment. For example, significant differences between mean grades awarded for the last set of criteria (proper referencing, the choice and the acknowledgement of used sources) may indicate that these aspects of essay writing need to be addressed more thoroughly in future iterations of the course. Secondly, low positive correlation between peers' and teachers' ratings may indicate that students are either insecure about their assessment competencies or they do not understand them completely, which reduces the validity of their assessment and further indicates to deficiencies in the assessment procedure and the preparation of students for their assessment task.

All the above points to limitations that need to be addressed in future iterations of the essay assessment procedure. Even though the procedure itself was carefully planned following the guidelines in Topping (2009: 25), several steps in the procedure need to be improved. Firstly, students were not included in determining criteria for assessment as they were

experienced neither with PA nor with academic essay writing before. Considering low positive correlation between PAs and TAs, we believe that involving students in discussions regarding essay evaluation criteria (as Falchikov 2007: 132-133, suggests) and providing more training (as suggested by Chang et al. 2010) may contribute to better achievement, higher correlation with teachers' grades and higher validity of peer assessment. Particular attention needs to be paid to the criteria which show most significant differences between teacher's and peer's grades, and the ones that have the lowest correlation.

Further research regarding this topic may include measuring the reliability and readdressing the validity of PA by including multiple peer assessors per one essay, as suggested by Cho et al. (2006: 892) and Schunn at al. (2016: 4). In the assessment procedure explained in this paper, reliability was not measured since each essay was assessed by a single peer assessor. In other words, the corrections made in students' essays may have been based on teachers' assessment only because students find the assessment made by their peers unreliable or invalid. Additionally, validity and reliability need to be looked at not only from teachers', but from students' perspective as well, by inviting students to fill in an assessment helpfulness scale, as suggested by Cho et al. (2006: 894).

#### REFERENCES

- Birenbaum, M. 2003. "New insights into learning and teaching and their implications for assessment." In *Optimising new modes of assessment: In search of qualities and standards*, pp. 13-36. Springer, Dordrecht. [https://doi.org/10.1007/0-306-48125-1\\_2](https://doi.org/10.1007/0-306-48125-1_2)
- Cho, K., Schunn, C. D., & Wilson, R. W. 2006. Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891. <https://doi.org/10.1037/0022-0663.98.4.891>
- Birenbaum, M. 2003. New insights into learning and teaching and their implications for assessment. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 13-36). Springer, Dordrecht. DOI: 10.1007/0-306-48125-1\_2 DOI: 10.1007/0-306-48125-1\_2
- Chang, C. C., Tseng, K. H., Chou, P. N., & Chen, Y. H. 2011. Reliability and validity of Web-based portfolio peer assessment: A case study for a senior high school's students taking computer course. *Computers & Education*, 57(1), 1306-1316. <https://doi.org/10.1016/j.compedu.2011.01.014>
- Cheng, W., & Warren, M. 1999. Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in higher education*, 5(2), 243-255. <https://doi.org/10.1080/135625100114885>
- Conway, R., Kember, D., Sivan, A., & Wu, M. 1993. Peer assessment of an individual's contribution to a group project. *Assessment & Evaluation in Higher Education*, 18(1), 45-56. <https://doi.org/10.1080/0260293930180104>
- Cooper, P. A. 1993. Paradigm shifts in designed instruction: From behaviorism to cognitivism to constructivism. *Educational technology*, 33(5), 12-19.
- Ernest, P. 1999. Social constructivism as a philosophy of mathematics: Radical constructivism rehabilitated? Retrieved from <https://systemika.g-i.cz/record/1595/files/Ernest,%20Paul%20-%20Social%20Constructivism%20as%20Philosophy%20of%20Mathematics.pdf>

- Falchikov, N., & Boud, D. 1989. Student self-assessment in higher education: A meta-analysis. *Review of educational research*, 59(4), 395-430. <https://doi.org/10.3102/00346543059004395>
- Falchikov, N., & Goldfinch, J. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research*, 70(3), 287-322. <https://doi.org/10.3102/00346543070003287>
- Falchikov, N. 2007. The place of peers in learning and assessment. In: Boud, D., & Falchikov, N. (Eds.). *Rethinking assessment in higher education* (pp. 138-153). Routledge.
- Graner, M. H. 1985. Revision techniques: Peer editing and the revision workshop. *Dissertation Abstracts International*, 47, 109.
- Havnes, A., & McDowell, L. (Eds.). 2008. *Balancing dilemmas in assessment and learning in contemporary education*. Routledge.
- Hein, G. E. 1991, October. *Constructivist learning theory*. Paper presented at the CECA (International Committee of Museum Educators) Conference, Jerusalem, Israel.
- Hoover, W. A. 1996. The practice implications of constructivism. *SEDL Letter*, 9(3), 1-2.
- Kiraly, D. 2005. Project-based learning: A case for situated translation. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 50(4), 1098-1111. <https://doi.org/10.7202/012063ar>
- Kaufman, J. H., & Schunn, C. D. 2011. Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*, 39(3), 387-406. <https://doi.org/10.1007/s11251-010-9133-6>
- Meršnik, M., Anđelković, J. 2020. Teaching academic writing at a business-oriented university. *Conference Proceedings of the IV International Conference: From Theory to Practice in Language for Specific Purposes*, pp. 158-173, ISSN: 1849-9279
- Oldfield, K. A., & Macalpine, J. M. K. 1995. Peer and Self-assessment at Tertiary Level--an experiential report. *Assessment in Higher Education*, 20(1), 125-132. <https://doi.org/10.1080/0260293950200113>
- Mowl, G., & Pain, R. 1995. Using self and peer assessment to improve students' essay writing: A case study from geography. *Innovations in Education and Training International*, 32(4), 324-335. <https://doi.org/10.1080/1355800950320404>
- Prawat, R. S., & Floden, R. E. 1994. Philosophical perspectives on constructivist views of learning. *Educational Psychologist*, 29(1), 37-48. [https://doi.org/10.1207/s15326985ep2901\\_4](https://doi.org/10.1207/s15326985ep2901_4)
- Schunn, C., Godley, A., DeMartino, S. 2016. *The Reliability and Validity of Peer Review of Writing in High School AP English Classes*. *Journal of Adolescent & Adult Literacy*, 60(1), 13-23. doi:10.1002/jaal.525
- Sivan, A. 2010. Implementing peer assessment to improve teaching and learning. *Hong Kong Baptist University*.
- Spiller, D. 2012. Assessment matters: Self-assessment and peer assessment. *The University of Waikato*, 13.
- Sluijsmans, D. M., Moerkerke, G., Van Merriënboer, J. J., & Dochy, F. J. 2001. Peer assessment in problem based learning. *Studies in educational evaluation*, 27(2), 153-173.
- Topping, K. 2003. Self and peer assessment in school and university: Reliability, validity and utility. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Springer, Dordrecht. DOI: 10.1007/0-306-48125-1\_4
- Topping, K. J. 2009. Peer assessment. *Theory into practice*, 48(1), 20-27. <https://doi.org/10.1080/00405840802577569>

- Twomey Fosnot, C. 1989. *Enquiring teachers, enquiring learners: A constructivist approach for teaching*. Teachers College Press. Retrieved from <http://www.exploratorium.edu/ifi/resources/constructivistlearning.html>
- Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. 2010. Effective peer assessment processes: Research findings and future directions. *Learning and instruction*, 20(4), 270-279. <https://doi.org/10.1016/j.learninstruc.2009.08.004>
- Vygotsky, L. S. 1978. *Mind in society*. Harvard University Press.
- Zhang, F., Schunn, C., Li, W., & Long, M. 2020. Changes in the reliability and validity of peer assessment across the college years. *Assessment & evaluation in higher education*, 45(8), 1073-1087. <https://doi.org/10.1080/02602938.2020.1724260>