LEVEL OF AGREEMENT BETWEEN TEACHER ASSESSMENT AND SELF-ASSESSMENT OF ORAL PRESENTATION SKILLS

Biljana Orovchanec-Nineska
International Balkan University, Macedonia
borovcanec@gmail.com
North Macedonia

Marija Stevkovska, MA
International Balkan University, Macedonia
m.stevkovska@ibu.edu.mk
North Macedonia

ABSTRACT

The focus of this research is agreement between self- and teacher assessment through oral presentations and it confirms the expected results already familiar in the literature i.e. students, in general, assess themselves higher than teachers. Assessment rubric of 15 questions with a five-point Likert scale was used for both self- and teacher assessment was used. For the analysis, standard statistical techniques in MS Excel were used. The study shows that although the relationship between teacher assessment and self-assessment is weak, it is still statistically relevant and it shows significant difference, but when comparing teacher assessment results with self-assessment results for each teacher separately, some unexpected outcomes emerge. Out of five, only one teacher's results confirm the well-known assumption – students assess themselves higher than teachers. However, all the others seem to be influenced by some factors and limitations that impose different results than the overall one.

Keywords: assessment, self-assessment, teacher assessment, oral presentation, agreement

1. INTRODUCTION

Oral presentations seem to be good vehicle through which self- and teacher assessment can be observed as processes directed towards acquiring new skills and ability in developing responsibility and autonomy, as well as reflection on students learning. Current trends in assessment emphasize formative assessment as an alternative and an all-inclusive process where teachers and students work together. Assessors, teachers and students, are ready to use it even as a form of summative assessment. On the other hand, it seems to be quite difficult for the managers, employers or institutions outside education, to abandon traditional assessment as it is very clear and practical. But the more they insist, the stronger the effort of the teachers to change that it is.

Formative assessment is an ongoing process of gathering information on the extent of learning, on strengths and weaknesses, which the teacher can feedback into their course planning and the actual feedback they give learners. Formative assessment is often used in a very broad sense so as to include non-quantifiable information from questionnaires and consultations. (CEFR, p.186) –"*Assessment for Learning*"; whereas, s*ummative assessment* sums up attainment at the end of the course with a grade or a quantitative mark. (CEFR, 2001, p.186) – "*Assessment of Learning".*

Similarly and Somervell (1993) suggest that formative assessment, especially self- and peer-assessment, can be used for summative purposes as part of the co-assessment by giving the teacher the power to make the final decision about a process or a product. The combination of self-, peer- and co-assessment with a

summative result enables students and teachers work together in a constructive way and achieve higher level of understanding, making all stakeholders of the educational process happy.
Self-assessment (SA) and peer-assessment (PA), as they usually go together, highlight the involvement of the students in the process of assessment and learning in general. SA is an arrangement for learners and/ or workers to consider and specify the level, value or quality of their own products (Topping, 2003, p.58). SA refers to the involvement of learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning (Boud and Felchikov, 1989, p.529). SA becomes a process for the learner through which they develop skills and abilities that would help them in many areas in the educational process. Boud and Falchikov (1989) suggest that effective learners have a realist view about their own strengths and weaknesses and they can use knowledge regarding their own learning process to direct their studying into productive directions. In addition, students' involvement into the assessment process develops other necessary lifelong learning skills such as responsibility, judgment and autonomy which have considerable importance for their professional life (Sluijsmans et al., 2001).

Studies about SA or PA through oral presentation skills are many, but still not enough. There are many studies which use different kinds of participants and instruments during the research, and make the results difficult to read and compare. The participants can vary from students with English as a second language as a major to American students with a major in science. Nevertheless, students who do self-assessment through oral presentations direct the whole process towards self-regulated learning, and via observational learning, learners compare their performance with standards of a good oral presentation. This *good* presentation is previously given as a set of instructions and explicitly written in a self-assessment grid used during the assessment processes. The oral presentation skills will evolve by achieving a better match between these standards and the current performance level (Sadler, 1989). The process itself is called *calibration* and it refers to the match between an internal evaluation and a standard (Winne, 2004). Self-assessment helps the process of calibration.

There are a lot of variables that affect the reliability and quality of the research. Student success and level are important variables that affect self-assessment. There is a tendency that more able students under-rate themselves and vice versa, weaker students over-rate themselves (Dochy et al 1999). He also reports that advanced students evaluate their performance more accurate than novices. Boud and Falchikov (1989) say that self-assessment results get more accurate over time with experience, maturity and practice. There are also personal differences in standards and rating styles that affect the assessment for both self- and teacher assessment, but clear instructions and training improve assessment skills. On the part of the teachers, there is the everlasting question which is little researched in comparison to self-assessment: Is this "expert" or teacher assessment so undoubtedly reliable? It turns out that the results of the inter-rater reliability among the teachers raise a lot of questions that should be further answered.

The research questions which are covered in this study are as follows:

- What is the overall level of agreement between self-assessment and teacher assessment?
- What are the individual levels of agreements (teacher – student)?
- What is the inter-rater reliability of the teachers?

2. METHOD

2.1 Participants and procedure

The participants that took part in this study made a short audio-visual power point presentation about a famous person from an English-speaking country. Each of the presentation lasted approximately 5 minutes and there were four presentations in a day. The whole process took about two weeks to be finished. There were 31 students of English as a foreign language in the preparatory year at International Balkan University

in Skopje, Macedonia. They were at the level of B1 or B1+ according to CEFR. The balance between female and male was kept and there were approximately 50% of each. The students' average age was about 19. Each student made one self-assessment and was asked to do it immediately after the presentation. There were also six experienced teachers who did the teacher assessment. They did 6.2 assessments in average. Both students and teachers were previously familiar with the assessment rubric and the meaning of the questions. Students received a detailed training with explicit instructions and video samples of what is expected from them in terms of preparing and presenting the presentation. The teachers participated in creating the assessment rubric which was based on a rubric previously used through the years in preparatory year as an assessment tool. They all knew the rubric well. The rubric was consisted of 15 oral presentation evaluation criteria divided into three categories: seven criteria about *the content and organization*, four criteria about the nature of the *delivery* and four criteria about the *language use* (see Appendix 1) A 5-point Likert scale was used to quantify the results.

This is an example of a question with a five-point Likert scale.

Was the beginning/ opening interesting?
1 -------- 2 ------- 3 ------- 4 ------- 5
Very bad average very good

The same assessment rubric was used by both students and teachers.

Assessors	Total number of assessments	Average number of assessed presentations for one assessor
Teachers	31	6.2
Students	31	1

Table 1. Summary of the assessment procedure

2.2 Analysis

The obtained data were entered in MS Excel for data analysis. They were analyzed using three statistical analyses. First, to measure the relationship between teacher assessment and self-assessment Pearson-Product Moment correlation analysis was performed. Secondly, to compare the total rubric scores of teacher assessment with that of self-assessment an independent-samples t-test was conducted. Mean scores of teacher assessment and self-assessment were calculated and compared. Finally, to compare the scores of each teacher assessment with the corresponding self-assessment independent-samples t-tests were done for each teacher separately.

Reliability is a key point with different assessor. There are several ways to calculate and interpret the correlation between assessors. A Pearson product-moment correlation coefficient was computed to assess the relationship between teacher assessment scores and self-assessment scores. Based on the results of the study, teacher assessment is weakly related to self-assessment (r =.33, p <.001). A scatterplot summarizes the results (Figure 1). Overall, there was a weak, not very positive correlation between TA and SA. Increases in scores of SA were not always correlated with increases in scores of TA.

Teacher mean	Self mean	Teacher SD	Self SD	Pearson *r*
48.09	**56.35**	**10.4**	**7.74**	**0.33**

Table 2. Teacher scores versus self-assessment scores (n=31)

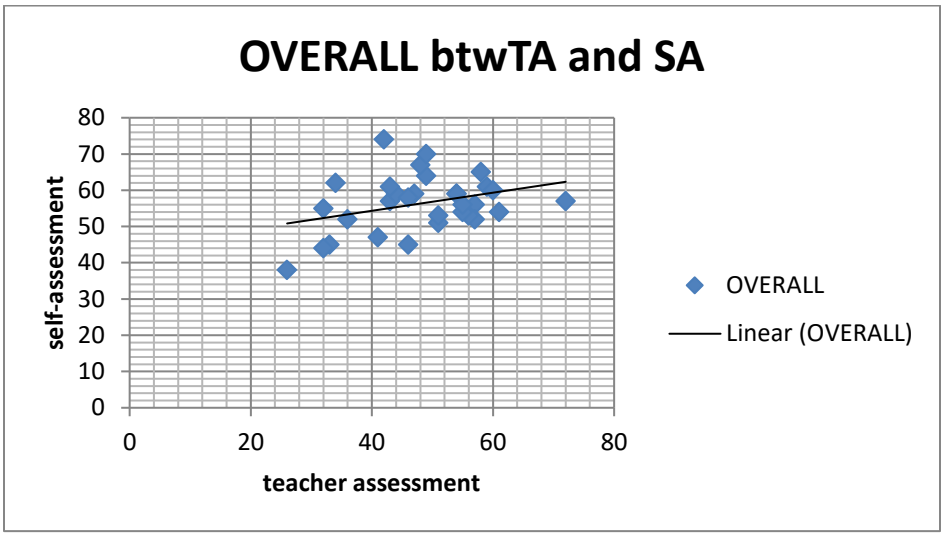


Figure 1. The correlation between teacher assessment and self-assessment is not strong, but it is still statistically relevant.

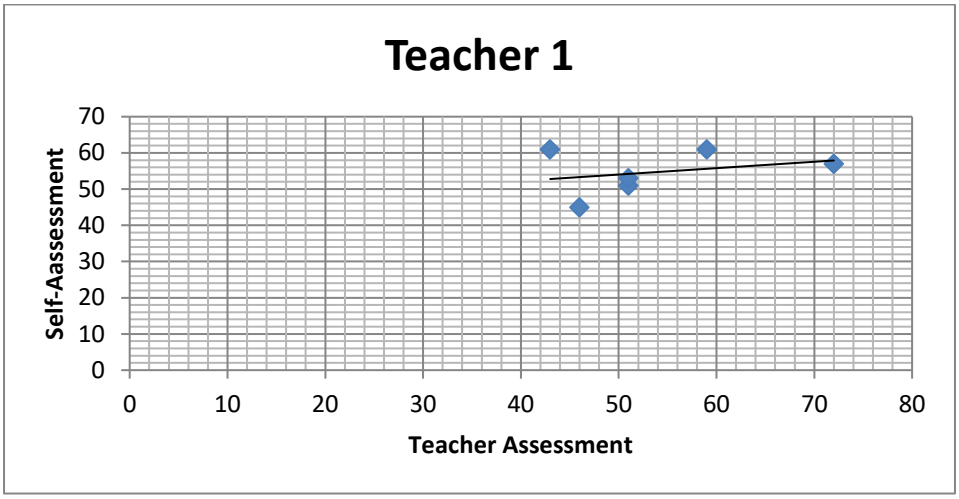
On the other hand, an independent-samples t-test was conducted to compare the total rubric score of teacher assessment scores and the total rubric score of self-assessment scores. There was a significant difference in the scores of teacher assessment (*M*=48.1, *SD*=10.4) and self-assessment(*M*=56.4, *SD*=7.7; *t* (60)=3.54, *p* <.001. These results suggest that teacher assessment and self-assessment are different and generally students assess their own work higher than teachers."

The correlations between the scores of selfs and teachers show a considerable variation in the marks. Independent-samples t-tests were conducted to compare the separate rubric scores of teacher assessments and the separate rubric scores of self-assessments. Pearson product-moment correlation coefficients were also computed and the individual relationships between teacher assessments and self-assessments show that there is a big discrepancy and that the range of correlation is big.

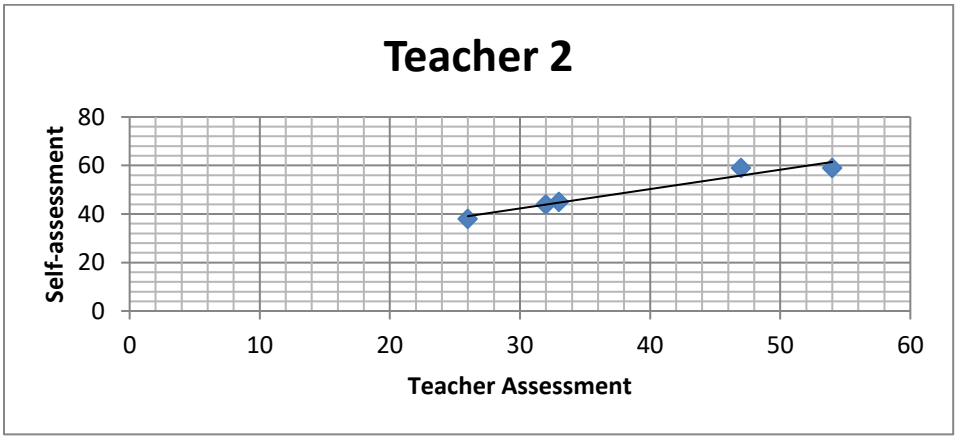
	Teacher mean	Self mean	Teacher SD	Self SD	Pearson *r*
Teacher 1	53.66	54.66	10.50	6.25	.30
Teacher 2	38.4	49	11.63	9.51	.97
Teacher 3	44.66	58.33	11.6	3.21	-0.88
Teacher 4	49.44	61.8	5.24	7.51	-0.69
Teacher 5	49.75	55.37	11.56	5.34	.54

Table 3. Separate teacher assessment scores compared to self-assessment scores

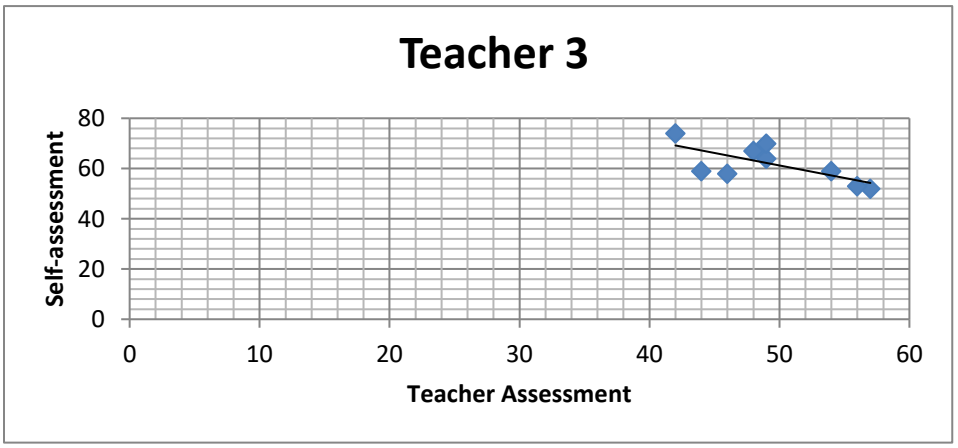
Based on these results we can clearly see that the score assessment of some teachers, for example Teacher 1 (r =.30, n=6, p = .84), is weakly related to the self-assessment score. It means that there is almost no difference in the final results, although they do not always increase simultaneously together. However, Teacher 2 assessment score (r =.97, n=5, p =.15) is strongly related to self-assessment, but still statistically insignificant or there is no difference in the final results. The results concerning self-assessment scores showed that students had realistic perceptions of their own abilities and made rational judgements or the teacher explained the evaluation criteria in a manner which was very well comprehended. The only relevant score that shows difference in scores and a kind of good correlation is Teacher 3 score (r =-0.69, n=9, p <.001), but the correlation is negative and as teacher's mark grow, as student's marks drop. The scatterplots summarize the results (Figure 2).



Teacher 1 (r =.30, n=6, p = .84)



Teacher 2 (r =.97, n=5, p =.15)



Teacher 3 (r =-0.69, n=9, p <.001)

Figure 2. Separate correlations between teachers and students

3. DISCUSSION AND CONCLUSION

In this research we get closer to the alternative assessment or formative assessment. Self- and teacher assessment are becoming part of the co-assessment which helps both students and teachers. Students become more independent, reliable and responsible, and teachers share the burden and the knowledge with the students preparing them for a life-learning journey. This process is also part of the self-regulating learning which uses oral presentation skills as vehicle towards its means.

In general, the comparison of the overall rubric scores of self- and teacher assessment concluded that the level of agreement is positive but low, or with other words, there was a significant difference of their scores which is similar to those reported in other studies. Students, in general, assess themselves higher than teachers.

Comparison of self- and teacher assessment rubric scores emphasizes the positive relationship, although with lots of discrepancy on individual level. Low correlation between the scores shown above tells us that teachers and students interpret the criteria differently. High correlation shows that the comprehension of the evaluation criteria is the same between teachers and students. This can be a result of many reasons. One of them is the experience - teachers' experience in teaching, assessing and conducting self-assessment. Teachers have bigger and much longer experience than students. They recall larger sets of models from the past. The criteria that they build are not confined to that one classroom in the present. They stretch over the years before and over the years to come. No matter how fantastic it sounds, teachers do project their work in future too.

However, experience can have negative impact on teacher's work as well. In regard to the inter-rater reliability of the teachers, when some informal interviews were made with the teachers after these diverse individual teachers' results, they said that being aware of the social and personal background of the students from the previous generations, they did not expect much from the new generations as well. They seem to be more lenient as it can be seen from Teacher 1, where there is almost no difference in the means and the overall rubric score. This negative attitude towards the newcomers affected students badly, and it should be taken into consideration. These kinds of studies can serve as a wake-up call.

Other important factor that may reflect on the research is the number of presentations assessed. The teacher with the highest number of assessed presentations (*n=9)* got most reliable results. A decision must be made beforehand upon the lowest and highest possible number of presentations assessed by one assessor in order to get reliable results.

There might be other reasons for such deference. In creating the rubrics, not all teachers seemed to take a serious attitude towards its creation. Some of the teachers were not aware of some basic notions of assessment, which mean that they should also be offered pre-training sessions just like the students. The assessment pre-training should be offered to both teachers and students or the assessors in general. Students should also be involved into the process of rubrics creation and defining of the assessment criteria. Falchikov (2005) suggests developing evaluation criteria in close collaboration with students. Low reliability level suggests that training of assessors is very important not just for the study itself, but also for the quality of educational process. More experienced students tend to be more accurate in their self-assessment than less experienced students (Lejk & Wyvill, 2001).

The length of the rubric might be another reason that affects this discrepancy. As reported by Sluijsmans (2002) seven-item questionnaire was suggested, which makes my fifteen-item rubric complicated and difficult. De Grez et al. (2009b) distinguish nine criteria rubric with descriptors and indicators provided to support the assessment process.

Providing learners with a good quality feedback about their oral presentation skills is also important for the acquisition process. Winne (2004) stress the importance of the feedback and its accuracy, and York (2003) state not only the quality of the feedback, but also what students do with that feedback. Students should take right actions in future based on good quality feedback. The combination of the two is important.

REFERENCES

Boud D and Falchikov N (2006) Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education* 31(4): 399–413.

De Grez L, Valcke M and Roozen I (2012) How effective are self- and peer assessment of oral presentation skills compared with teachers' assessment? *Active Learning in Higher Education* 13(2) 129-142.

De Grez L, Valcke M and Roozen I (2009a) The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers and Education* 53: 112–20.

De Grez L, Valcke M and Roozen I (2009b) The impact of goal orientation, self-reflection and personal characteristics on the acquisition of oral presentations skills. *European Journal of Psychology of Education* 24(3): 293–306.

*De Grez et al.*141Kerby D and Romine J (2009) Develop oral presentation skills through accounting curriculum design and course-embedded assessment. *Journal of Education for Business* 85: 172–9.

Dochy F and Cacallar E (eds) *Optimising New Modes of Assessment: In Search of Qualities and Standards.*Dordrecht: Kluwer Academic, pp. 37–54.

Falchikov N (2005) *Improving Assessment Through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education.* New York: RoutledgeFalmer.

Freeman M (1995) Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education* 20(3): 289–301.

Gibbs G (2006) How assessment frames student learning. In: Bryan C and Clegg K (eds) *Innovative Assessment in Higher Education*. London: Routledge, pp. 23–36.

Hafner J and Hafner P (2003) Quantitative analysis of the rubric as an assessment tool: An empirical study of peer-group rating. *International Journal of Science Education*25(12): 1509–28.

Hughes I and Large B (1993) Staff and peer-group assessment of oral communication skills. *Studies in Higher Education* 18(3): 379–85.

Kilic D (2016) An examination of using self-, peer-, and teacher- assessment in higher education: a case study in teacher education. *Higher Education Studies* 6(1): 136-144.

Langan A. Mark et al. (2008) Relationship between students characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment and Evaluation in Higher Education* 33(2): 179-190.

Miller P (2003) The effect of scoring criteria specificity on peer and self-assessment. *Assessment and Evaluation in Higher Education* 28(4): 383–94.

Sluijsmans D (2002) Student involvement in assessment: The training of peer assessment skills. Unpublished doctoral thesis, Open University of the Netherlands, Heerlen.

Sluijsmans D (2008) Towards (quasi-) experimental research on the design of peer assessment. In: van den Heuvel-Panhuizen M and Lacher M (eds) *Challenging Assessment: Book of Abstracts of the Fourth Biennial Earli/Northumbria Assessment Conference.* Berlin: Humboldt-Universität, p. 46.

Sluijsmans D, Moerkerke G, Van Merriënboer J and Dochy F (2001) Peer assessment in problem based learning. *Studies in Educational Evaluation* 27: 153–73.

Topping K (1998) Peer assessment between students in colleges and universities. *Review of Educational Research* 68(3): 249–76.

Topping K (2003) Self- and peer assessment in school and university: Reliability, validity and utility. In: Segers M, Dochy F and Cascallar E (eds) *Optimising New Modes of Assessment: In Search of Qualities and Standards.*Dordrecht: Kluwer Academic, pp. 55–87.

Topping K (2009) Peer assessment. *Theory Into Practice* 48: 20–7.

Winne P (2004) Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research* 41: 466–88.
Yorke M (2003) Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education* 45: 47–501.

APPENDIX 1

Oral Presentation

Teachers/ Self-assessment evaluation form
Topic: …… Class: ……

Organization and content:

1. Was the beginning/ opening interesting?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
2. Was the contents list good?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
3. Were there any pictures?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too many	
4. Was the text on the slides easy to read?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very difficult average very easy	
5. Was there enough text?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too much	
6. Did you use details/ examples/ facts to support the main points?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too many	
7. Was there clear conclusion in the end?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	

Delivery:

8. Did you prepare yourself well?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
9. Did you read the presentation?	1 -------- 2 ------- 3 ------- 4 ------- 5 All of it average only the notes	
10. Did you make eye contact?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too much	
11. Did you use your hands?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too much	

Language:

12. Was grammar correct?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
13. Was vocabulary appropriate for the level?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
14. Was pronunciation good?	1 -------- 2 ------- 3 ------- 4 ------- 5 Very bad average very good	
15. Did you use transition words?	1 -------- 2 ------- 3 ------- 4 ------- 5 Not enough average too many	