2023-10-10

Self-Assessment Or Peer Assessment? Which Is Better Predictor Of Test Results

Nárcisz KULCSÁR
Széchenyi István University, Hungary, kulcsar.narcisz@math.sze.hu

SELF-ASSESSMENT OR PEER ASSESSMENT? WHICH IS BETTER PREDICTOR OF TEST RESULTS

Nárcisz Kulcsár [1]
Széchenyi István University
Győr, Hungary
0000-0001-8525-5851

Conference Key Areas*: Fundamentals of Engineering: Mathematics and the Sciences*
Keywords*: self-assessment, peer assessment*

ABSTRACT

Internationally accredited engineering programmes are becoming increasingly important in the internationalisation agenda of universities. ABET has highlighted transversal skills in its accreditation criteria for engineering degrees. Preferred transferable skills include the ability of students to reflect on their own performance, the ability to give constructive feedback and the ability to make judgements.

Students' self- and peer-assessment was examined in the context of a basic mathematics course. During the maths midterm tests, students self-assessed on each task, and assessed another student's test. These assessments were compared with the points given by the teacher. 84% of students overestimated their actual performance and more than 60% of them overestimated their peer's performance, and both overestimations were low. According to students' opinion, peer assessment is as easy as self-assessment, it is not easier for them to spot mistakes in other people's work than in their own. The research results showed significant difference in the accuracy of peer and self-assessment, peer assessment is closer to teacher evaluation than self-assessment. Contrary to our previous research, now we did not find a significant correlation between students' performance and assessment accuracy in the first test. One reason for this may be that these students have failed this subject at least once.

As further learning is only possible once we have identified what needs to be learned, the ability to assess the gained knowledge as accurately as possible is appreciating. In addition to meeting accreditation requirements, the different type of assessments' cognitive and affective effects on learning outcomes make it a good choice for classroom use.

[1] *Nárcisz Kulcsár*

kulcsar.narcisz@math.sze.hu

1 INTRODUCTION

In the last decade, the rapid development of engineering industry and the acceleration of robotisation have had an impact on employment. Although education is a slowly changing system, it has to respond to these changes. In addition to technical knowledge, transversal skills have become increasingly valued and are being developed with increasing emphasis by higher education institutions. International accreditation requirements also address the development of these competences. There is a huge literature on the study of transversal skills, but there is no single agreed definition between academic and non-academic organisations, which makes it difficult to measure [1]. Despite the diversity of the list of transversal skills, the ability to self- and peer-assess is one of them [2].

Because learning is more than simply acquiring knowledge, it involves students' active participation in judging their own work and proactively seeking and using inputs from others. Self-assessment (SA) and peer assessment (PA) require students to take an active and reflective role, to understand and apply assessment criteria, to seek and use feedback and to evaluate their own and others' work [3]

1.1 Self-assessment

First, it is important to clarify that self-assessment is an umbrella term that encompasses a range of self assessment. Panadero et al. [4] identified 20 categories of SA implementations, varying from a simple form of awarding a grade for their own work (i.e., self-grading or self-marking) to a more complex form that evaluate their own work based on predetermined criteria, capturing the strengths and weaknesses of their own work.

1.2 Peer assessment

Secondly, peer assessment, like self-assessment, is also an umbrella concept that encompasses a range of peer assessment. Van Helden et al. [5] distinguished three types of PA according to their function in educational output.

1. Peer review: students review each other's (written) output and give feedback to each other. The recipient of the feedback is not obliged to reply to the feedback and change their output based on the feedback. Examples of outputs : essays, reports, computer code.
2. Peer grading: students grade each other's work in a formative or summative way based on a pre-defined set of criteria. It is not a detailed feedback, rather it is limited to the answer is correct or to what extent the student has delivered what was asked based on the given criteria.
3. Peer evaluation: students evaluate each other during the learning process and reflect on for instance transversal skills within this process.

Different peer assessment methods are used depending on the content of the subjects and the skills developed by the subject. In mathematics education, peer review and peer grading is the most common form. Pick et al. [6] used self assessment and peer review in a matematics course for first year engineering students. Students assessed each other on 4 criteria and they had to reflect weekly on comments received from peers:

1. Effort (Clear evidence of effort in answering - even if not correct)
2. Correctness (All correct)
3. Coherence (Method can be followed very clearly (even if answer not correct). Excellent annotation, notation and clear steps)
4. Conciseness (Method used is appropriate and very efficient)

This research is a good example of the many ways in which maths performance can and should be assessed. Until now, mathematics education has focused primarily on solving problems correctly and evaluating only this.

1.3 Objective

In a previous study I examined the accuracy of self-assesment of engineering students, some relationships between self-assesment and performance, and the impact of feedback on self-assesment [7]. In addition to self-assessment, peer assessment also plays a role in this research. Therefore, the present study aimed to answer the following questions:

1. To what extent are engineering students overestimating and underestimating their performance and their peer's performance in a basic mathematics course?
2. Is there a significant interrelationship among accuracy scores and performance?
3. Is there a significant difference in the accuracy of self-assessment/peer assessment between students who fulfilled mid-term requirements and those who did not?
4. Is peer or self-assessment closer to teacher assessment?
5. How easy do students find it to evaluate their own and others' work?

2 METHODOLOGY

2.1 Measure of self-assesment

Several indices of self-assessment can be distinguished, e.g. the accuracy (reliability) and the direction of the bias (validity). Based on the literature [8] the accuracy and direction of students' self-assessment was measured using two indicators: the realism/bias score and the accuracy score.

Realism/bias score = (Average self-assesment score over all items in the test) – (Average performance score over all items in the test)

Accuracy score = the absolute value of the difference between the self-assesment score and performance score for each test item, summed over all items on a test, and divided by the total number of items

During the semester, students wrote two midterm tests and an exam. To take the exam, students must achive a score of 50% in the two tests together. Those who did not meet this requirement could take a make-up test. Each test consisted of 6 tasks for 2 points per task. Before tests, students were given the opportunity (extra lessons) to take more mock tests and learn the scoring rules for each task. Students graded each task scoring 0, 1 or 2 points. Teacher assessment could also be 0, 1 or 2 points.

Based on this, the bias value could take a value between -2 and 2, where a positive value indicates that the student overestimated his performance, while a negative value indicates underestimation. Values close to 0 indicate a lack of bias. The accuracy score could take a value between 0 and 2, where 0 indicates complete accuracy and 2 indicates complete inaccuracy.

2.2 Participants

142 engineering students took the course Mathematics 2, 124 students wrote the first midterm test, 99 the second midterm test. All students have registered for the course at least once, but have not fulfilled the basic requirements of the course.

3 RESULTS

Self- and peer-realism scores were calculated from the results of self- and peer-assessment following the midterm tests and from the teacher's assessment. 84% of students overestimated their actual performance in both midterms. Peers scored the tests more critically, 79% of them overestimated the other's performance on the first test and 67% of them on the second (Table 1).

Table 1. Distribution of self- and peer-realism scores

	Self-realism score midterm test1			Self-realism score midterm test2			Peer-realism score midterm test1			Peer-realism score midterm test2		
	<0	0	>0	<0	0	>0	<0	0	>0	<0	0	>0
Number of students	7	11	95	7	5	62	11	13	92	10	14	49

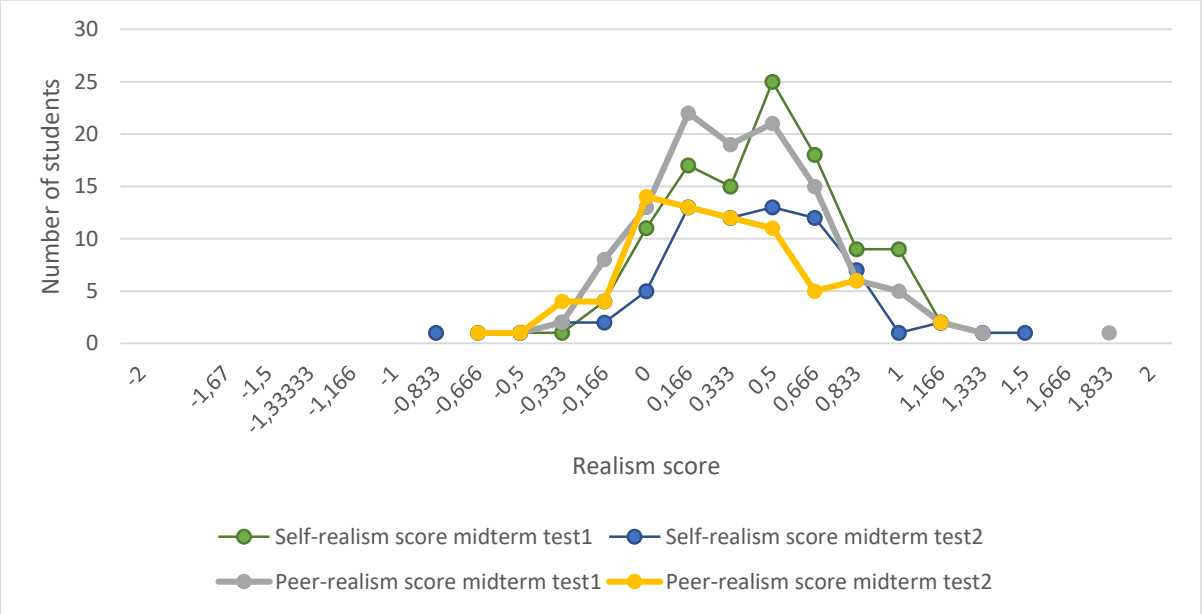

Fig. 1. Realism scores in the tests

In the interval 0-0.5 we talk about low overestimation, between 0.5-1 moderate overestimation, between 1-1.5 high overestimation, between 1.5-2 very high overestimation. Peer assessment is closer to teacher assessment than self-assessment, namely peer assessment shows lower overestimation than self-

assessment (Self-realism score1: mean 0.43, Self-realism score2: mean 0.42, Peer-realism score1: mean 0.37, Peer-realism score2: mean 0.27) (Fig 1).

When comparing peer and self-assessment, it can be seen that peer assessment is closer to teacher's assessment than self-assessment in both tests. Using a paired samples t-test, this difference is significant (Table 2). Thus, it can be said that students score their peers' tests more strictly than their own, they notice errors in their peers' tests more easily. One reason for this may be that it is easier for students to check the sub-steps of an existing thought process than to create and construct a new one.

Table 2. Paired samples test between self- and peer-realism scores

		N	Mean	SD	Sig.
Realism score midterm test1	Self-realism score	110	0,432	0,356	0,040
	Peer-realism score	110	0,368	0,374	
Realism score midterm test2	Self-realism score	72	0,421	0,4	0,000
	Peer-realism score	72	0,271	0,364	

Accuracy scores were used to find correlation between self-assessment accuracy and test results. The results of the correlation calculation are shown in Table 3, which does not show correlation between the accuracy scores and the first test scores. In contrast, for the second test we found a negative correlation between the accuracy scores and the test scores. Negative correlation means that students with better results in tests have an accuracy score close to 0, i.e. they give a more accurate self-assessment of their own performance than students with weaker results.

Table 3. Correlation between accuracy scores and test scores

	Accuracy score midterm test1	Accuracy score midterm test2
Midterm test1 score	-0,078	
Midterm test2 score		-0,383**
Total score of tests	0,29	-0,297*

*p<0,05 **p<0,01

In a previous research [6], there was a significant correlation between test scores and accuracy score for both tests, which is partly in contrast to the current results. One reason for this may be that while in the previous research students took the subject for the first time, in the current research students took it for at least the second time which means underperforming students were only in the sample.

An influencing factor behind the change in the significance of the correlation in the current study may be that those students did not come to take the second test who had little chance of completing the subject based on their poor first result. 25 fewer students wrote the second test, and 19 of them got 4 or less scores on the first test.

If students' self-assessment is further examined in terms of their performance, then while there is no significant difference between students' accuracy scores who meet the requirements of the subject and those who do not in the first mid-term, there is a significant differencs in the second (Table 4). Thus, the accuracy scores of students who finally fulfilled the requirements of the course improved significantly compared to students who did not complete the course requirements.

Table 4. Difference between accuracy scores based on the fulfillment of requirements

		N	Mean	SD	Sig.
Accuracy score midterm test1	Not fulfilled the requirements	59	0,590	0,331	0,667
	Fulfilled the requirements	56	0,565	0,285	
Accuracy score midterm test2	Not fulfilled the requirements	27	0,716	0,351	0,01
	Fulfilled the requirements	46	0,514	0,216	

3.1 Students feedback about self and peer assessment

Students were asked how easy it was to assess their own work and others. They rated the difficulty of self- and peer-assessment on a 5-point Likert scale. Students found self-assessment and peer assessment almost equally easy (mean of self-assessment 2.69, mean of peer assessment 2.80).

"Evaluating our own work is difficult because it is difficult not to be biased against ourselves. However, evaluating a student's work can be easier because we can discover solutions that we hadn't thought of, or we can be reassured that he or she has carried out the task in the same way and that his or her results are the same as ours. "

"It's very difficult to evaluate until there is only one answer for another subject, there are many steps here everyone thinks differently. But after some practice I could assess with more confidence. "

"I think that as difficult and sensitive as the topic is, it is also useful because we can see and learn from the solutions of our fellow students. "

4 CONCLUSIONS

Overall, the data suggest that the accuracy of self-assessment varies significantly during the semester, especially for students who meet the requirements of the subject. Factors that may affect the improvement in self-assessment accuracy include checking the mistakes in the first test, practicing and scoring the mock tests, and practicing with midterm quizzes. Accuracy of peer assessment showed a significant difference from self-assessment, even though students perceive self-assessment and

peer assessment to be equally easy. The students' evaluation is closer to the teacher's when it comes to evaluating their peers' work than their own.

As engineers work in teams, they need to evaluate their own work and understand and evaluate the work of others as well. Therefore, the development of these competences should also be emphasised during their university studies.

REFERENCES

[1] Leandro Cruz, M. (2021), Measurement and Practice of Transversal Competencies in Engineering Education: Evaluation of Perceptions and Stimulation of Reflections of industry, lecturers and students. https://doi.org/10.4233/uuid:730e80b5-e567-494d-8bb2-df8c71e6de69

[2] Jalali, Y., Kovaks, H., Isaac, S., & Dehler Zufferey, J. (2022), Bringing Visibility to Transversal Skills in Engineering Education: Towards an Organizing Framework. Proceedings of the 50th Annual Conference of the European Society or Engineering Education, pp. 1252-1259.

[3] Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022), Effects of self-assessment and peer-assessment interventions on academic performance: A pairwise and network meta-analysis. Educational Research Review, 100484.

[4] Panadero, E., Brown, G. T., & Strijbos, J. W. (2016), The future of student self-assessment: A review of known unknowns and potential directions. Educational psychology review, 28, pp. 803-830.

[5] Van Helden, G., Van Der Werf, V., Saunders-Smits, G. N., & Specht, M. M. (2023), The Use of Digital Peer Assessment in Higher Education–an Umbrella Review of Literature, IEEE Access.

[6] Pick, L., Manda, K., & Cole, J. (2022), A Mathematics Peer Assessment Process for Flexible Modes of Delivery, Proceedings of the 50th Annual Conference of the European Society or Engineering Education, pp. 1436-1444.

[7] Kulcsár, N. (2022), Is the engineering students' self- assessment accurate? Analysing midterm tests in terms of self-assessment accuracy, Proceedings of the 50th Annual Conference of the European Society or Engineering Education, pp. 2028–2033.

[8] González-Betancor, S. M., Bolívar-Cruz, A., & Verano-Tacoronte, D. (2019), Self-assessment accuracy in higher education: The influence of gender and performance of university student, *Active learning in higher education*, Vol. 20, No. 2, pp. 101-114.