Peer and self-assessment to improve mathematics competence in pre-service middle-school teachers

Antoine Julien[1,†], Elisabeth Romijn[1,2,†] and Alexander Schmeding[1,†]

[1]Nord University, Norway; antoine.julien@nord.no, alexander.schmeding@nord.no

[2]Current address [from Aug. 2022]: NTNU, Norway; elisabeth.romijn@ntnu.no

[†]All authors contributed equally and are listed in alphabetical order

Abstract: We report on an exploratory study in which we used self-assessment and peer assessment in a mathematics class for pre-service middle-school teachers.

Keywords: assessment, mathematics, teacher education, computer aided instruction.

Assessment and feedback are amongst the most effective tools teachers can use to promote students' learning (Hattie, 2008). Assessment, however, can be resource-intensive, especially when it comes to giving written feedback on students' assignments. Hence, peer-assessment can be a tool by which students can achieve better learning without [the university spending] extra resources, as described by Maugesten (2005). Generally, peer assessment and self-assessment can be performed in various manners with different objectives (Topping, 1998). Following Panadero et al. (2016, p. 804), we define self-evaluation as any activity by which students describe or evaluate the qualities of their own learning processes and products. We use Topping's (1998, p. 250) definition of peer assessment as being an arrangement in which students evaluate the work of peers of similar status. Our interest in self- and peer assessment stems from the goal of enhancing both students mathematical and assessment skills without increasing the workload of university staff.

Falchikov and Goldfinch (2000) reviewed peer assessment studies in higher education, though not specifically in mathematics. They highlight the importance of the grading criteria and instructions which are given to the students. See also Li et al. (2019), in which moderate correlation between student-given grades and teacher-given grades is found (in line with our findings); this correlation was however higher when using paper-based rather than computer-assisted assessment. This is interesting as ICT tools are crucial to facilitate and scale our research into teaching practice.

In the present article, we report on an exploratory study in which we used self-assessment and peer assessment in a mathematics class for pre-service middle-school teachers. Although we gathered a variety of data, including self- and peer assessment data on a mathematical task, mathematical confidence score and interviews, only some data is relevant to this paper (see Methodology section). Our study is similar to Zevenbergen (2001) but differs in important details (our students only graded one delivery not several and the scores they awarded were inconsequential for student grades).

Our main goal was to investigate how pre-service teachers assess their own and each other's mathematical work, and how these activities of assessing benefit them. Through collection of assessment data and semi-structured interviews, we aim at answering the following question: How accurately do pre-service teacher's assessment of own and peer's work agree with the educator's assessment?

In planning and executing this peer-assessment experiment, we were crucially assisted by ICT tools: the online learning platform and questionnaires facilitated the distribution of assignments to peers for grading and allowed students to give feedback effectively to each other. This level of automatization enabled peer assessment beyond an experimental setting as a routine teaching practice.

Methodology and research setting

The group from which we collected data consisted of three different classes of pre-service teachers in a Norwegian university. The first two groups were two sections of the same mathematics course, given at two different campuses of the university. These students ($n = 36$) were in their second year of a 5-year integrated master's program of middle-school teacher education. The third group consisted of one class of fourth year students ($n = 11$) participating in the final mathematics course in a five-year integrated high-school teacher education program. All three groups were studying probability and statistics, which allowed for similar assignments.

In each of these classes, the number of participants was rather low (47 student across all three groups), so we cannot, in general, expect our results to be statistically significant. Therefore, all of our findings should be interpreted as a preliminary evaluation. However, see Schönbrodt and Perugini (2013) for an overview of (small) sample size versus stability of correlations.

The aspects of the study's protocol relevant to this paper are described below:

1. Measure students' mathematical confidence through an online questionnaire inspired by Pierce (2007). The score ranges from 1 (low) to 5 (high).
2. Develop a homework assignment with several mathematics exercises. We purposefully included exercises of different nature: from computation-based to very open.
3. Immediately after completing the problems and delivering their work, the students were asked to complete a questionnaire in which they assessed their own performance out of 10 for each task (*self-assessment*).
4. Develop a grading guide. The content of the guide was carefully considered such that it provided a framework without dictating how the grading and feedback should be executed.
5. The participants were randomly assigned a peer's work and asked to give written feedback on each task, as well as give a grade out of 10 for each task using the grading guide (*peer-assessment*).
6. Independently, we graded the students' papers, with the same criteria (*educator evaluation*).
7. We interviewed three students, and transcribed each interview.

The self-assessment in step 3 was performed by the participants without access to the grading guide used in steps 5 and 6. This was done to obtain expectation of performance directly without measuring against an external corrective. Our research design included the following aspects. All scores (self-assessment, peer assessment or educator evaluation) were converted to percentages in the data-analysis. Next, even though we asked the students to give formative assessment to their peers, we will restrict our analysis on the numerical grades. It is also worth noting that in the self-evaluation, students were asked to estimate how well they performed, and not which grade they think they would get, or which grade they would give themselves. How important this request is to the numerical results is unclear and may require further investigation.

Results of empirical research

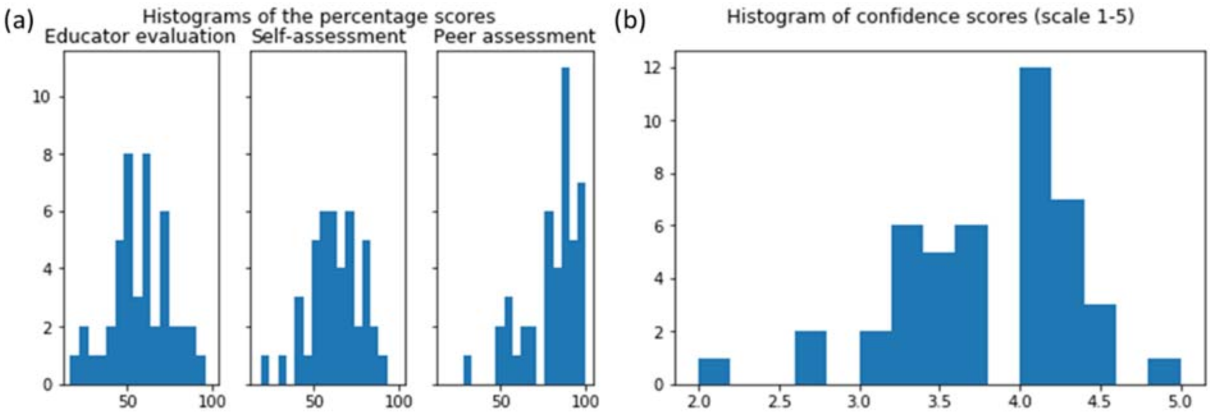

Figure 1: Distribution of students according to assessment scores in percentages (a) and the confidence score calculated based on the self-assessment questionnaire (b).

The full set of participants consisted of 47 students, some of whom did not participate in all activities. Non-answers pertaining to these participants have been removed from the statistics of the items they were not included in. Thus, the data set fluctuates depending on the item, but always contains at least 42 students in every item. The statistical methods used in this section are standard and can be found in introductory statistics texts (e.g., Ross, 2010). Computations were carried out using Python's Pandas and SciPy packages. We evaluated the point scores (in percent) achieved by participants on the assignment in three categories: Educator evaluation (given by the authors), self-assessment (awarded by the students themselves on delivery), peer assessment (awarded by another student). The histogram in Figure 1a shows the percentage scores divided into 7.5% bins each.

The histogram gives a visual indication that the educator evaluation and self-assessment are already (asymptotically) normally distributed. Note that there is a large deviation between the self-assessments and the peer assessment item. On average the students have evaluated each other much more positively (mean = 80.3%) than the educator has evaluated them (mean = 58.8%). This might be a consequence of the process not being anonymized (students knew whom they assessed and by whom they were assessed).

The mean of the self-assessment score of 63.1% is similar to the mean of scores awarded by the educators. Computing the (Pearson) correlation for both items yields a coefficient r= 0.31 which is significant on the 5% scale (p=0.04). Interestingly though, the correlation coefficient for educator evaluation vs. peer assessment score (a student received from another) turns out to be r= 0.34 with p=0.03. In future work, we will investigate the underlying mechanisms of these findings. Beyond the summative assessment, we collected items tracking the student's confidence in mathematics on a scale from 1 (lowest) to 5 (highest). We called their mean the confidence score (cf. Figure 1b).

In the complete group there is a weak correlation between confidence score and points awarded in self-assessment of the exercise (r= 0.24, p=0.14). The result is not significant, which is not surprising considering the size of the group considered (cf. Schönbrodt & Perugini, 2013). However, the confidence score is also weakly correlated to educator evaluation (r=0.22, p=0.07) with a surprisingly low *p*-value (though it doesn't meet the 0.05 conventional threshold for significance). Therefore, this lends some credibility to the items evaluated for the confidence score as a useful albeit weak predictor for success in mathematical exercises. We plan to expand the collection of items to make the confidence score more robust and useful as a predictor.

Conclusion and outlook

Preliminary analysis of the interview data suggests that the activity of assessing gave students a deeper understanding of the material, maybe more so than receiving feedback from other students. In future papers, we will explore this point, and investigate the role of the grading aid. We also plan to investigate to which extent the validity of the peer-evaluation can be improved by having the student's grade (and therefore compare) several of their peers' works. Iterating on these findings, we will adapt peer-review practices in future courses via a design-based research process.

References

Brignell, C., Wicks, T., Tomas, C., & Halls, J. (2019). The impact of peer assessment on mathematics students' understanding of marking criteria and their ability to self-regulate learning. *MSOR Connections*, *18*(1). https://doi.org/10.21100/msor.v18i1.1019

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

Hattie, J. (2008). *Visible learning.* Routledge. https://doi.org/10.4324/9780203887332

Li, H., Xiong, Y., Hunter C.V., Guo, X., & Tywoniw, R. (2019). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45:2,* 193-211. https://doi.org/10.1080/02602938.2019.1620679

Maugesten, M. (2005). Bedre læring av matematikk uten bruk av ekstra ressurser? *Uniped*, *28*(2), 28–39.

Panadero, E., Brown, G. L., & Strijbos, J.-W. (2016). The future of student self-assessment: a review of known unknowns and potential directions. *Educational Psychology Review* 28, 803–830. https://doi.org/10.1007/s10648-015-9350-2

Pierce, R., Stacey, K., & Barkatsas, A. (2007). A scale for monitoring students' attitudes to learning mathematics with technology. *Computers and Education*, *48*(2), 285–300. https://doi.org/10.1016/j.compedu.2005.01.006

Ross, S. M. (2010). *Introductory statistics.* (3rd. ed.). Elsevier/Academic Press. https://doi.org/10.1016/C2009-0-28078-0

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in preservice mathematics education. *Journal of Mathematics Teacher Education,* 4, 95–113. https://doi.org/10.1023/A:1011401532410