

Peer assessment

Maryam Alqassab¹ & Ernesto Panadero²

¹ Faculty of Psychology, Chair of Education and Educational Psychology. Ludwig-Maximilians- Universität, Munich, Germany.

² Departamento de Psicología Evolutiva y de la Educación. Facultad de Psicología. Universidad Autónoma de Madrid, Spain.

Recommended citation:

Alqassab, M & Panadero, E. (2020). Peer assessment. In Brookhart, S. et al. (Eds.) *Routledge Encyclopedia of Education*. New York: Routledge

This is a pre-print of a manuscript to be published. Personal use is permitted, but it cannot be uploaded in an Open Source repository. The permission from the publisher must be obtained for any other commercial purpose. This article may not exactly replicate the published version due to editorial changes and/or formatting and corrections during the final stage of publication. Interested readers are advised to consult the official published version.

Summary

Peer assessment is an educational activity in which students judge the performance of their peers and it can take different forms depending on the characteristics of its implementation, the learners and the learning context. There are mainly two functions for peer assessment, as an assessment tool (i.e., to judge performance) and as a learning tool (i.e., to develop domain knowledge and assessment skills). While most of the research traditionally focused on peer assessment as an assessment tool, a large body of research addressing it as a learning activity has been growing for the last twenty years. Evidence from studies on both lines of research indicates: (a) the accuracy of peer assessment scores can be influenced by several characteristics of the implementation, (b) students need instructional support to be better assessors, (c) peer assessment can lead to improved performance for the assessor and, to a lesser extent, the assessee, (d) domain knowledge is a prerequisite for the development of peer assessment skills, (e) students' perspective regarding peer assessment is usually negative before the implementation and does not always become more positive afterwards, and (f) interpersonal variables (e.g., trust in peers as assessors) are adversely influenced when the main purpose of peer assessment is to produce a score. Accordingly, peer assessment should be carefully implemented taking the diversity of peer assessment practices and the purposes of assessment into account.

Keywords: Peer assessment; peer feedback; peer review; peer evaluation; peer learning; formative assessment; summative assessment; student-centred assessment.

Peer Assessment

Peer assessment is an activity in which one or more students judge the performance of their peer(s) by providing a grade/score, or qualitative comments, or both. Most researchers adopt the definition by Topping (1998) regarding peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status” (p. 250). However, various definitions of peer assessment have been proposed representing different approaches in the field. For example, while Falchikov (1995) defined peer assessment as “the process through which groups or individuals rate their peers”. (p. 175), van Gennip, Segers, and Tillema (2009) considered peer assessment as “an interpersonal process in which a performance grade exchange is being established and in which the core activity is feedback given to and received from others, aimed at enhancing the performance of an individual and/or a team or group as a whole.” (p. 42). Differences in the definitions can be attributed to the diversity of peer assessment implementation and the emphasis on different functions of peer assessment.

The variants of peer assessment

The implementation design varies depending on the characteristics of the peer assessment activity, the learners, and the learning context. Topping (1998) proposed a peer assessment typology that specifies the main seventeen variables that peer assessment in higher education can vary on, including curriculum area, objectives, focus, product/output, relation to staff assessment, official weight, directionality, privacy, contact, year, ability, constellation (assessors & assessed), place, time, requirement, and reward. Thereafter, several researchers classified, modified, and extended the typology as a result of the ongoing increase in the diversity of peer assessment practices (see Adachi, Tai, & Dawson, 2018; S. Gielen, Dochy, & Onghena, 2011; van den Berg, Admiraal, & Pilot, 2006; van Gennip et al., 2009).

Identifying the characteristics of the peer assessment design is important for teachers and researchers who use or plan to implement it. Teachers should be aware that the dimensions of peer assessment might have differential effects on the assessment activity and researchers can use these characteristics as a guide to develop stronger research designs (Gielen et al., 2011). Overall, there are two main strands of research on peer assessment: as an assessment tool and as a learning tool.

Peer assessment as an assessment tool for summative purposes

Traditionally, there was a tendency for peer assessment to take a summative focus as assessment was done on the final product, often the only feedback given was a score and entailed no further action by the assessee. A typical peer assessment activity of this type involves providing a grade on peer's work at the end of the semester, or rating the contributions by peers to a teamwork after submitting the team's project. Most of the research conducted on peer assessment focused on this type of peer assessment and investigated the reliability (i.e., agreements between multiple peers) and the accuracy of peer assessment compared to teacher's assessment (for meta-analyses see Falchikov & Goldfinch, 2000; Li et al., 2016). Some scholars claimed that peer assessment could be an alternative to teacher's assessment to reduce workload and decrease the time required for assessment especially in large classes (Sadler & Good, 2006). Conversely, others argued that the implementation of peer assessment requires teachers' time (e.g., Topping, 2009) and teachers support this position (Panadero & Brown, 2017).

The main aim of this line of research seems to be exploring whether students are well calibrated when giving a peer score in comparison to teachers; what has been called accuracy or validity (Panadero, Romero & Strijbos, 2013). Falchikov and Goldfinch (2000) reviewed 48 peer assessment studies and found that the average association between teacher and peer assessment scores was $r = 0.69$ representing a moderate agreement between teacher's and

peers' scores. A recent meta-analysis also found a similar pattern with an average value of $r = 0.63$ between teacher's and peer assessment using 69 studies (Li et al., 2016). Several factors were found to positively influence the accuracy of peer assessment among them: using paper-based instead of computer-based assessment, involving students in creating assessment criteria, conducting peer assessment in higher education instead of school education, using individual assessment object instead of group work, randomly matching assessors and assessees, using non-anonymous peer assessment, voluntary instead of compulsory assessment, and combining the scores with qualitative comments (Li et al., 2016). Thus, the accuracy of peer assessment varies depending on the design of the peer assessment activity.

The problem with the focus on summative or quantitative peer assessment is the disconnection between the assessment activity and the instructional activities that is likely to overshadow the value of involving students in assessment to promote their learning (Strijbos & Sluijsmans, 2010). Thus, a close alignment between peer assessment and learning motivated another strand of research that investigates various learning outcomes of peer assessment.

Peer assessment as a learning tool for formative purposes

Peer assessment is used as a learning activity with a formative purpose in which information about learning progress is shared with students during the learning process providing them the opportunity to improve their performance to reach their learning goals. This type of assessment is integrated in the learning activity and takes place throughout different stages of the learning activity instead of the end of the semester (Strijbos & Sluijsmans, 2010). This activity promotes self-reflection and learning gains are not restricted to the assessees, instead the assessors can learn through assessing as it involves deep cognitive activities (e.g., evaluation, questioning) that can help them consolidate their knowledge and deepen their understanding of the subject matter (Panadero et al., 2018;

Topping, 1998). The role of feedback is emphasized in this type of peer assessment and the assessment may adopt a formative focus through providing the opportunity for students to revise their work (e.g., Y. H. Cho & Cho, 2011). An example of a formative peer assessment activity is providing comments on a first draft of peer's work that subsequently gets revised and submitted.

Importantly, learning from peer assessment is not limited to the gain of domain knowledge but additionally involves the development of assessment skills (i.e. evaluate judgment, Tai, Ajjawi, Boud, Dawson, & Panadero, 2018) or higher order thinking skills (e.g., metacognition; see Chang, Wu, & Kuo, 2012; Kim & Ryu, 2013; self-regulated learning; see Panadero, Jonsson & Strijbos, 2016). In most educational settings, teaching assessment skills is not done in isolation of domain knowledge (Sluijsmans et al., 2004), hence these two skills are intertwined and are equally important. Indeed, empirical evidence suggests that domain knowledge is a prerequisite for the development of peer assessment skills (Alqassab, Strijbos, & Ufer, 2018; van Zundert, Sluijsmans, Könings, & van Merriënboer, 2012).

Studies investigating the effect of peer assessment on improving domain knowledge and performance are still limited (for reviews see Panadero et al., 2018; van Gennip et al., 2009; M. van Zundert et al., 2010). Findings on improved performance appear to be mixed with some studies showing an effect (K. Cho & MacArthur, 2011; Gielen & De Wever, 2015; Gielen & De Wever, 2012) and others not showing an impact on performance (e.g., Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Sluijsmans et al., 2004). Regarding improvements in peer assessment skills, particularly peer feedback content, research shows that training and structuring the peer assessment activity through rubrics and scaffolds results in better peer feedback skills (Gan & Hattie, 2014; Sluijsmans et al., 2004). However,

domain knowledge appears to influence the effects of training (Alqassab et al., 2018b).

Many peer assessment studies measure students' perspective (e.g., beliefs, attitudes, perceptions) either as a main outcome variable (e.g., Cheng & Warren, 1997; Wang, 2014) or as an additional outcome variable (Alqassab et al., 2018b; Sluijsmans et al., 2004). The findings are also mixed as some studies indicated positive effects of peer assessment on students' perspective (Cheng & Warren, 1997; Sluijsmans et al., 2004) and others showed negative effects (Alqassab et al., 2018b; Wang, 2014). The problem with relying on students' perspective as a main outcome variable of peer assessment is that it does not provide an objective measure of learning and thus might not reflect a real effect (Panadero et al., 2018).

Some researchers combine the assessment (i.e., summative) and the learning (i.e., formative) approaches of peer assessment to get the most out of it. For example, through implementing peer assessment during the learning process but also using the outcomes of peer assessment (grades or comments) to make summative judgements (e.g., Chinn, 2005). This usually—although not necessarily—takes the form of combining scores with comments. Some researchers argue that combining a grade with qualitative comments can dilute the learning benefits of assessment (Black & Wiliam, 1998; Boud, 2000). Yet, no study on peer assessment—to our knowledge—has investigated whether combining comments with grades adversely influences learning.

Interpersonal variables in peer assessment

The social nature of peer assessment necessitates the need to investigate the role of variables related to interpersonal communication between peers during the peer assessment activity (i.e., interpersonal variables; Panadero et al., 2018). Despite the importance of interpersonal variables in shaping the processes and outcomes of peer assessment, research investigating their roles in relation to learning is still limited (Panadero, 2016; Strijbos & Sluijsmans, 2010). In a review of studies that explored social and interpersonal factors in peer

assessment, Panadero (2016) identified several interpersonal variables including

psychological safety (beliefs about the safeness to make critical judgments in a group; e.g., van Gennip et al., 2010), friendship marking (awarding friends higher grades; e.g., Panadero et al., 2013), trust in self and others as assessors (e.g., Cheng & Tsai, 2012), value diversity (differences in members' perceptions of their learning goals or outcomes; e.g., van Gennip et al., 2010), and interdependence between assessors and assesses (e.g., van Gennip et al., 2010). The review suggested that adopting a summative purpose of peer assessment might have adverse effects on interpersonal variables (Panadero, 2016).

Implications and conclusions

The current findings of peer assessment research suggest that in order to ensure successful peer assessment outcomes, students need instructional support through providing training (e.g., Sluijsmans et al., 2004), scaffolds, and assessment criteria (e.g., Gan & Hattie, 2014; S. Gielen et al., 2010). All of these activities require additional preparation and involvement by the teacher and thus peer assessment is unlikely to save time in the short term. Yet, still no clear evidence supports a positive effect on performance perhaps due to the complexity of peer assessment and the relatively short interventions in the literature.

Feedback seems to be equally useful for peer assessment activities with summative purposes as it can result in more accurate peer assessment scores (Li et al., 2016).

Additionally, summative peer assessment is likely to have negative effects on interpersonal variables (e.g., trust or psychological safety) because limited feedback is communicated to the students and the awarded scores are not clearly explained which can increase students' perceptions of discomfort and unfairness of peer assessment (Panadero, 2016).

Overall, the implementation of peer assessment can be challenging for teachers and researchers due to the interplay between three main factors. First, students' reluctance to use or trust peer assessment that does not always decrease with more exposure to peer assessment

(e.g., Wang, 2014). Second, teachers often lack relevant training to design and implement peer

assessment activities (Panadero & Brown, 2017). Third, the complexity of peer assessment design, due to the possibility of conducting peer assessment in various ways, is often taken for granted by researchers and practitioners which posits a challenge to draw solid conclusions. Accordingly, there is still a need for empirical studies to investigate the dynamics of peer assessment, and this should be guided by clear specifications of the characteristics of the implemented peer assessment activity.

References

- Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development, 37*(3), 453–467. <https://doi.org/10.1080/07294360.2017.1405913>
- Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018a). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. *Learning and Instruction, 58*, 182–192. <https://doi.org/10.1016/j.learninstruc.2018.07.003>
- Alqassab, M., Strijbos, J.-W., & Ufer, S. (2018b). Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11–30. <https://doi.org/10.1007/s10212-017-0342-0>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. <https://doi.org/10.1080/713695728>

- Chang, S.-H., Wu, T.-C., & Kuo, Y.-K. (2012). Project-based learning with an online peer assessment system in a photonics instruction for enhancing led design skills. *The Turkish Online Journal of Educational Technology*, 11(4), 12.
- Cheng, K.-H., & Tsai, C.-C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, 28(4). <https://doi.org/10.14742/ajet.830>
- Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233–239. <https://doi.org/10.1080/03075079712331381064>
- Chinn, D. (2005). Peer Assessment in the Algorithms Course. *ITiCSE '05 Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education*. Portugal, 37 (3), 69–73. <https://doi.org/10.1145/1151954.1067468>
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–84. <https://doi.org/10.1037/a0021950>
- Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, 39(5), 629–643. <https://doi.org/10.1007/s11251-010-9146-1>
- Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International*, 32(2), 175–187. <https://doi.org/10.1080/1355800950320212>
- 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>
- Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6), 861–878. <https://doi.org/10.1007/s11251-014-9319-4>

- Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality: Structuring the peer assessment process. *Journal of Computer Assisted Learning*, 31(5), 435–449. <https://doi.org/10.1111/jcal.12096>
- Gielen, M., & De Wever, B. (2012). Peer Assessment in a Wiki: Product Improvement, Students' Learning And Perception Regarding Peer Feedback. *Procedia - Social and Behavioral Sciences*, 69, 585–594. <https://doi.org/10.1016/j.sbspro.2012.11.450>
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, 36(2), 137–155. <https://doi.org/10.1080/02602930903221444>
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304–315. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks. *Educational Technology Research and Development*, 61(4), 549–561. <https://doi.org/10.1007/s11423-012-9266-1>
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S., & K. Suen, H. (2016). Peer assessment in the digital age: a meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245–264. <https://doi.org/10.1080/02602938.2014.999746>
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.). *Handbook of Social and Human Conditions in Assessment* (pp. 247-266). New York: Routledge.

- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156. doi:10.1007/s10212-015-0282-5
- Panadero, E., Jonsson, A. & Alqassab, M. (2018). Peer feedback used for formative purposes: Review of findings. In A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback*. Cambridge University Press.
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies In Educational Evaluation*, 39(4), 195-203.
doi:10.1016/j.stueduc.2013.10.005
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1 - 31. doi:10.1207/s15326977ea1101_1
- Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: effects on performance and perceptions. *Innovations in Education and Teaching International*, 41(1), 59–78.
<https://doi.org/10.1080/1470329032000172720>
- Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20(4), 265–269.
<https://doi.org/10.1016/j.learninstruc.2009.08.002>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2017). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*. doi:10.1007/s10734-017-0220-3
- Topping, K. (1998). Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249. <https://doi.org/10.2307/1170598>

Topping, K. J. (2009). Peer Assessment. *Theory Into Practice*, 48(1), 20-27.

doi:10.1080/00405840802577569

van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: evaluating seven course designs. *Assessment & Evaluation in Higher Education*, 31(1), 19–36. <https://doi.org/10.1080/02602930500262346>

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41–54.

<https://doi.org/10.1016/j.edurev.2008.11.002>

van Zundert, M. J., Sluijsmans, D. M. A., Könings, K. D., & van Merriënboer, J. J. G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. *Educational Psychology*, 32(1), 127–145.

<https://doi.org/10.1080/01443410.2011.626122>

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>

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing*, 19, 80–96.

<https://doi.org/10.1016/j.asw.2013.11.008>