Evaluative judgement and peer assessment: promoting a beneficial reciprocal relationship

Joanna Tai
Deakin University,
Australia

There are many pedagogical benefits of peer assessment: it can develop content knowledge, students' feedback skills, and afford additional sources of feedback for students. Furthermore, peer assessment can contribute to the development of students' evaluative judgement, a core capability for independent practice and lifelong learning. However, peer assessment is frequently seen as problematic, due to logistical issues, and concern from both staff and students around the ability of peers to contribute meaningfully to learning. Somewhat paradoxically, students' evaluative judgement is likely to contribute to successful peer assessment. Technological solutions for peer assessment can have a significant role in improving uptake of peer assessment practices. If such implementations also focus on the core requirements/principles of evaluative judgement development, this may be one way to improve the success of peer assessment. This paper provides a rationale for the inclusion of peer assessment within curricula. It introduces the concept of evaluative judgement; highlights the benefits and challenges currently faced within peer assessment, and identifies desired functionalities for peer assessment and evaluative judgement that could be implemented through technological means.

Keywords: evaluative judgement; peer assessment, assessment for learning

Introduction

Peer assessment encompasses a wide range of activities that students can do with each other. This might be viewing and providing marks on an oral presentation; marking up and providing feedback on a written piece of work; students marking each other on clinical skills; a team of students assessing each other on their teamwork skills; or even students correcting their peers' short answer questions according to an answer key. Essentially, these are all activities where "students judge and make decisions about the work of their peers against particular criteria" (Adachi, Tai, & Dawson, 2018b, p. 454). However, how do students come to be able to judge and make decisions of work against criteria? The notion of developing students' evaluative judgement can help to explain how this can be facilitated within educational settings. However, peer assessment may also contribute to the development of evaluative judgement (Tai, Ajjawi, Boud, Dawson, & Panadero, 2017). This conceptual paper aims to explore how developing students' peer assessment abilities, and developing students' evaluative judgement, interact and contribute to each other. Peer assessment very frequently also involves the use of technology: given the reciprocal nature of peer assessment and evaluative judgement, this paper will then consider how technology might contribute to the development of evaluative judgement through implementations of peer assessment.

Evaluative judgement: what is it, and why is it important?

Evaluative judgement is an emerging concept within higher education. It has been defined as "the capability to make decisions about the quality of work of self and others" (Tai, Ajjawi, et al., 2017, p. 5). At first glance, it may seem just like a more complicated way of referring to self and peer assessment, however it entails more than just participating in assessments. Evaluative judgement also requires an understanding of quality, and the ability to apply understandings of quality. Evaluative judgement is therefore crucial for independent practice, which may feed into self-regulated and lifelong learning practices. Quality, here, is a holistic concept that cannot be devolved, broken up, or otherwise itemised to exist as a checklist or other basic instrument which anyone could use.

The need for students to have a workable understanding of quality has been raised in a range of interconnected fields. When Sadler (2010) discussed the student's role in feedback, he identified that students must be able to process information and see the gap between their actual performance, and the expected performance. Boud (2000) approached students' understandings of quality from the concept of sustainable assessment: that students needed to partake in in assessments which prepared them for the future, rather than undermining their

independence. Here, Boud suggested that assessments should function so that students might be able to judge the quality of their performance in future instances where no formal assessment existed. Boud & Falchikov (2006) also argued that learners must *become* the assessor to foster long-term learning. Concepts such as phronesis, or "know-how" (Hager, 2000), connoisseurship (Eisner, 1985) and tacit knowledge (Eraut, 2000) also involve expertise of the individual. All these ideas touch on aspects of developing expertise, understanding quality, and in some cases, being able to judge quality. Evaluative judgement draws these ideas together, and provides a more articulate justification for a range of commonly used pedagogies, heretofore used to 'improve learning' but without an underlying conceptual rationale (Tai, Ajjawi, et al., 2017).

Evaluative judgement itself is discipline and context dependent (Bearman, 2018). Evaluative judgement therefore is not a generic skill that exists within all students in exactly the same way: we aim therefore to develop their evaluative judgement in a particular area. For instance, a civil engineer might be very good at identifying quality in concrete composition and pouring, but have little understanding of what constitutes quality for poems. A surgeon observing the use of laparoscopic equipment by a fellow surgeon will be able to comment on the skill of their peer in a qualitative manner, but may be unable to judge how well a consultant psychiatrist treats a patient with a mental illness. Returning to learner in higher education, an English literature major may be able to describe and execute a well-written persuasive essay, but be less able to write a policy advisory document. For a particular discipline, specific content knowledge, skills, attitudes, dispositions, and/or other qualities are required for evaluative judgement to be properly exercised. Indeed, students' ability to self-assess accurately is also somewhat context and task dependent (Boud, Lawson, & Thompson, 2015). Thus, there is an argument for implementing activities in all courses, and all units of study that develop students' evaluative judgement.

Ways of developing students' evaluative judgement

Developing evaluative judgement, on the other hand, may be a transferrable skill: the types of learning activities and processes that develop evaluative judgement are likely to be similar across professions and disciplines, even if they are not exactly the same. Several ways of developing evaluative judgement have been proposed. This includes the use of exemplars, the use and co-creation of rubrics, self-assessment, peer assessment, and feedback (Ajjawi, Boud, Dawson, & Tai, 2018). All of these activities are proposed to be helpful in helping students gain a more holistic understanding of quality. While rubrics might be seen as a way to break up a holistic understanding into various criteria, there are still elements of quality that are unlikely to be able to be captured within the standard range of rubrics, some of which can even be single descriptor rubrics (Dawson, 2015). Many of these activities are assessment related. Other common assessment activities that might develop complex and holistic understandings of quality include critical reflection, and the use of an annotated portfolio of work.

Evidence for these ways of developing students' evaluative judgement is scarce to date. This is likely due to the relative novelty of the term, rather than a lack of investigation into how students develop understandings of quality that they are able to employ in their own judgements of work. The work that does exist is in the field of exemplar use, and peer assessment and feedback. To & Carless (2015) found that in-class use of written essay exemplars, including peer and teacher discussion, helped students to be able to identify and judge quality. Similarly, Nicol (2014) found that peer assessment and feedback helped students to identify quality within essay work. Tai, Canny, Haines & Molloy (2016) also identified that peer feedback and discussion regarding performance of clinical skills contributed to students' evaluative judgement. Peer assessment may be particularly powerful in developing evaluative judgement, as when acting as an assessor of peers, students must exercise their evaluative judgement, thereby providing opportunities to practice and refine their judgement skills.

Peer assessment functions, benefits, and limitations

Peer assessments has multiple beneficial functions, including the development of transferable skills, providing an authentic form of assessment (when viewed in terms of future work and life-long learning skills), promote students' learning, including the development of evaluative judgement, the provision of feedback to students in a timely fashion, and developing students' feedback skills. It may also alter the nature of educators' time and input required for a particular learning activity (Adachi et al., 2018b; Rust, Price, Donovan, & Brookes, 2010).

Which benefits arise depend on how the peer assessment is constructed: there are several choices to be made concerning the use of peer assessment. Adachi, Tai and Dawson (2018a) identified a total of 19 design elements which contribute to the overall make-up of a peer assessment, including one cluster of decisions concerning the use of peer assessment. This included the subject area, intended learning outcomes, whether it is a process (e.g.

team work, communication skills) or a product (e.g. essay, presentation) that is being assessed, and whether that assessment counts towards grades (and how much it contributes towards that grade). For example, where students are asked to formatively assess each other on their participation in a group project to provide feedback to team members, students will gain feedback on their performance, and it may be seen as an authentic form of assessment, as a form of 360º performance appraisal. Asking students to mark peers' lab reports against a rubric may foster the marker's learning of criteria and standards of work, and also mimics the peer review process which academics undertake for publishing, lending to its authenticity.

The limitations of peer learning, and particularly peer assessment, have been explored extensively. Where peer assessment is used for summative purposes, there may be significant backwash effects on students, and students may act differently as a result (Boud, Cohen, & Sampson, 1999). This may include preferential grading of friends or the deliberate down-grading of competitors: as Falchikov (2007) points out, "we cannot escape the tension between co-operation and competition that permeates education" (p. 139). These issues may be reduced through the formative use of peer assessment, reducing the incentive for deliberately inaccurate peer assessments. General concerns around the accuracy of peer assessment also persist, which feed into the more general idea that peers are unable to contribute to feedback and assessment, i.e. that students are the "blind leading the blind" (Carless, 2013; Tai, Haines, Canny, & Molloy, 2014). However, peer assessment has been found to become more accurate (when compared to educator assessments) when an understanding of criteria and standards is established (Falchikov & Goldfinch, 2000; Kulkarni et al., 2015). It is likely, therefore, that evaluative judgement is important for peer assessment to be done well, in the understanding of quality work, and the application of standards to judgements.

Peer assessment design for developing evaluative judgement

Given this reciprocity between peer assessment and evaluative judgement, what features of peer assessment therefore contribute to the development of evaluative judgement? Falchikov (2007) argues strongly for the use of scaffolding in peer learning, and then, once learners have developed their skills, for the gradual fading of support to allow students to become independent practitioners. Peer assessments should therefore commence with strong scaffolding: the articulation and discussion of standards, and quality of work, should be a starting point. Establishing a shared understanding is important for academics in the marking and moderation of work (Bloxham, 2009; Sadler, 2013); the same applies here. By participating in peer assessments, students may also come to understand the complexity and subjectivity within assessment processes (Bloxham, Den-Outer, Hudson, & Price, 2016). To develop students' skills in peer assessment, there should then be many opportunities to practice and undertake peer assessments, to help in understanding quality in particular areas – either in terms of the topic or content knowledge, or in terms of the format of assessment. To combat concerns around the legitimacy of peer assessments, the subject of the peer assessment should be something that students could reasonably expected to already have a level of knowledge about. Furthermore, students should be given explicit guidance around what that subject is, to indicate the boundaries and acknowledge the limits of their expertise (Tai, Canny, Haines, & Molloy, 2017). All of this contributes to a learning environment and culture where peer assessment is frequent, normalised, appropriate, and therefore expected. Using peer assessment in this manner, however, also implies that it occurs largely in addition to educator assessment and feedback, especially in the earlier years. This may represent a significant workload for educators, and so the role of technology in peer assessment will be important in promoting its uptake.

How can technology facilitate peer assessment and the development of evaluative judgement?

Technology is omnipresent in higher education. In one way or another, technology has a role in facilitating learning, and this is especially the case for distance learners. Practically, technology can help in the running of peer assessment activities. Allocation of pairs or groups can usually be done through the Learning Management System and with this comes the ability to exchange or share work to be assessed, and discuss tasks with each other. Beyond this, specific peer assessment software can also afford functionalities around providing grades, marks, or feedback to peers, and provide in-built mark-up tools for commenting on work. However, the implementation of these functionalities can be time consuming, requiring the support of educational developers and institutional information technology services (Adachi, Tai, & Dawson, 2016). This is likely to have an impact on who implements peer assessment, and how it is implemented.

Beyond the logistics of peer assessment, we should also consider what might be done to foster evaluative judgement, both in the immediate peer assessment, and across time. The key components of evaluative judgement should be supported: students should have easy access to standards, criteria, and representations of these things. This might be as simple as providing annotated exemplars and rubrics. Given the promising

research on video feedback, audio or video explanation of standards may provide a different media which helps with student processing of information (Mahoney, Macfarlane, & Ajjawi, 2018). Real-time discussion opportunities with peers and educators may also assist in developing understandings of quality prior to the act of peer assessment, and these might be facilitated online where a record of conversation is easily made for others to consult. Students are also likely to benefit from multiple opportunities to conduct peer assessments and practice their evaluative judgement capabilities: this might be automated through an adaptive system that provides peer work of varying quality, to provide a wider range of experiences.

Evaluative judgement is a continually developing capability. We should therefore also develop systems that allow for the longitudinal evidencing of evaluative judgement development. Though the judgement itself is complex and qualitative, students might gain feedback on the relative accuracy of their judgements. This could include visualisation of progress over time, and opportunities for students to reflect on their developing understandings of quality, and how they have applied them to their own work, and work of others. Such a repository would remain accessible across units and even courses of study.

Implications for practice and research

Practice and research which involves the term 'evaluative judgement' is relatively novel, though similar ideas have been discussed for many years. Therefore, the first goal for practice and research in this area could be seen as the explicit use and employment of evaluative judgement as a concept. Within the bounds of peer assessment, this may be especially helpful in providing a rationale for peer assessment, and a new way of communicating the desired outcomes of peer assessment.

Research which explores the development of evaluative judgement so far has occurred in face-to-face settings. The use of technology in these settings has been limited, as many activities are based around in-class discussion. Therefore, exploring how evaluative judgement is and can be developed while learning at a distance, and/or learning online is a key next step. The types of peer assessments undertaken in such learning environments are also likely to be designed differently, and so describing these accurately, and exploring their connection to evaluative judgement, will also be crucial.

Overall, evaluative judgement represents an exciting conceptual advance that draws together assessment and pedagogy in higher education. Peer assessment is likely to contribute to the development of evaluative judgement, and evaluative judgement to successful peer assessment. Within the context of the present higher education environment, the creative use of technology will be key in promoting both practice and research in this field.

References

Adachi, C., Tai, J., & Dawson, P. (2016). Enabler or inhibitor? Educational technology in self and peer assessment. In *Proceedings of ASCILITE 2016, Show Me The Learning, Adelaide, November 27-30* (pp. 11–16). Adelaide, SA. Retrieved from http://2016conference.ascilite.org/wp-content/uploads/ascilite2016_adachi_concise_mon.pdf

Adachi, C., Tai, J., & Dawson, P. (2018a). 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

Adachi, C., Tai, J., & Dawson, P. (2018b). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 294–306. https://doi.org/10.1080/02602938.2017.1339775

Ajjawi, R., Boud, D., Dawson, P., & Tai, J. (2018). Conceptualising evaluative judgement for sustainable assessment in higher education. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing Evaluative Judgement in Higher Education: Assessment for knowing and producing quality work*. Milton Park: Routledge. https://doi.org/10.1080/02602938.2018.1434128

Bearman, M. (2018). Prefiguration, identities and agency: The disciplinary nature of evaluative judgement. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work* (pp. 186–195). Milton Park: Routledge. https://doi.org/10.4324/9781315109251

Bloxham, S. (2009). Marking and moderation in the UK: False assumptions and wasted resources. *Assessment and Evaluation in Higher Education*, *34*(2), 209–220. https://doi.org/10.1080/02602930801955978

Bloxham, S., Den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking:

exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, *41*(3), 1–16. https://doi.org/10.1080/02602938.2015.1024607

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–167. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/713695728

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, *24*(4), 413–426. https://doi.org/10.1080/0260293990240405

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399–413. https://doi.org/10.1080/02602930600679050

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher Education Research & Development*, *34*(1), 45–59. https://doi.org/10.1080/07294360.2014.934328

Carless, D. (2013). Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising Feedback in Higher Education* (pp. 113–122). Hoboken, NJ, US: Taylor and Francis.

Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2015.1111294

Eisner, E. W. (1985). *The Art of Educational Evaluation*. Philadelphia, Pennsylvania: The Falmer Press.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, *70*(1), 113–136. https://doi.org/10.1348/000709900158001

Falchikov, N. (2007). The place of peers in learning and assessment. *Rethinking Assessment in Higher Education: Learning for the Longer Term*, (2007), 128–143. https://doi.org/10.4324/9780203964309

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

Hager, P. (2000). Know-How and Workplace Practical Judgement. *Journal of the Philosophy of Education*, *34*(2), 281–296. https://doi.org/10.1111/1467-9752.00173

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., … Klemmer, S. R. (2015). Peer and Self Assessment in Massive Online Classes. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design Thinking Research* (Vol. 20, pp. 131–168). Springer. https://doi.org/10.1007/978-3-319-06823-7_9

Mahoney, P., Macfarlane, S., & Ajjawi, R. (2018). A qualitative synthesis of video feedback in higher education. *Teaching in Higher Education*, 1–23. https://doi.org/10.1080/13562517.2018.1471457

Nicol, D. (2014). Guiding Principles for Peer Review: Unlocking Learners' Evaluative Skills. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and Innovations in University Assessment and Feedback* (pp. 197–224). Edinburgh University Press. Retrieved from http://www.jstor.org/stable/10.3366/j.ctt14brxp5.16

Rust, C., Price, M., Donovan, B. O., & Brookes, O. (2010). Improving Students ' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education*, *28*(2), 147–164. https://doi.org/10.1080/0260293032000045509

Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, *35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 5–19. https://doi.org/10.1080/0969594X.2012.714742

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*. https://doi.org/10.1007/s10734-017-0220-3

Tai, J., Canny, B. J., Haines, T. P., & Molloy, E. K. (2016). The role of peer-assisted learning in building evaluative judgement: opportunities in clinical medical education. *Advances in Health Sciences Education*, *21*(3), 659–676. https://doi.org/10.1007/s10459-015-9659-0

Tai, J., Canny, B. J., Haines, T. P., & Molloy, E. K. (2017). Implementing Peer Learning in Clinical Education: A Framework to Address Challenges In the "Real World." *Teaching and Learning in Medicine*, *29*(2), 162–172. https://doi.org/10.1080/10401334.2016.1247000

Tai, J., Haines, T. P., Canny, B. J., & Molloy, E. K. (2014). A study of medical students ' peer learning on clinical placements : What they have taught themselves to do. *Journal of Peer Learning*, *7*, 57–80.

To, J., & Carless, D. (2015). Making productive use of exemplars: Peer discussion and teacher guidance for positive transfer of strategies. *Journal of Further and Higher Education*, 1–19. https://doi.org/10.1080/0309877X.2015.1014317