

Accepted version of

Boud D., Dawson P., Tai J. and Ajjawi R. (2018). Creating an agenda for developing students' evaluative judgement, In Boud D., Ajjawi R., Dawson P., Tai J. (Eds.). *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*, London: Routledge, 186-195. ISBN: 9781138089358, 9781138089341, 9781315109251

Creating an agenda for developing students' evaluative judgement

David Boud, Phillip Dawson, Joanna Tai and Rola Ajjawi

By placing evaluative judgement on the agenda for curriculum reform in higher education, this book points to an underdeveloped, but potentially highly fruitful, direction for adding value to current courses. The “capability to make decisions about the quality of work of self and others” (Ajjawi et al. [Chapter 1, this volume]) is already implicit as an outcome to be developed in university courses, but many of the preceding chapters have argued that it should be actively pursued, rather than taken for granted that it will occur.

We are not suggesting that university courses currently neglect to cultivate students' capacity to develop their evaluative judgement. A cursory glance at any course will uncover multiple examples of activities that foster aspects of judgement which can be mapped to the practices described in this book. What is not occurring, however, is a concerted effort to develop the skills of students across the curriculum and to link evaluative judgement to the substantive learning outcomes for a given discipline or profession. If university courses are promoting this, then it is not well-articulated and we do not sufficiently appreciate how they go about doing it. Furthermore, it is not clear how they might be assisted in this process. It is also not clear how we might judge students' progression along the paths to developing their evaluative judgement capability, other than observing improved work. This is rather indirect and may not help us diagnose problems in judgement rather than execution.

Many different theoretical perspectives have been introduced in this volume. These include social constructivism, cognitive psychology, practice theory, sociomateriality, ontological/epistemic fluency, and identity. These perspectives all have useful contributions to make to this area. They provide tools by which others can interrogate the various ideas which have been proposed and which we take forward here. However, it is not the role of this chapter to attempt to bring these together or suggest where any one theoretical perspective might be fruitful in the quest for further understanding of evaluative judgement and how it might be fostered.

What we seek to do in this final chapter is to point to practical directions that may be taken by those who wish to form an agenda for applying these ideas in course development. This treatment is exploratory, seeking to open up further discussion of how evaluative judgement might be considered and structured into courses.

The development of evaluative judgement involves more than introducing a set of activities to the curriculum. It involves a repositioning of the formation of students within a discipline or profession, and a reinterpretation of the notion of students' development over time. It prompts a new focus on questions such as the development of generic attributes, the structuring of a curriculum, and what constitutes progression from one level of sophistication to another. It also focuses on the role of assessment in learning and how it can contribute, not just to the judgement of students by other parties, but to the judgement by students of themselves and their peers, and where educators and other people can fit into this.

Components of evaluative judgement: a learner focus

While there is more work that needs to be done to fully conceptualise the notion of evaluative judgement and how it might be fostered, we know enough to suggest a pragmatic way forward that can assist in planning courses, especially for the area that Robert Nelson (Chapter 5) has identified as hard evaluative judgement. If we logically examine the notion of evaluative judgement and break it down into its constituent elements, we can identify things done well at present and aspects which may not be done well or at all. This analysis will vary from course to course and from unit to unit in any given course. Also, there will be variation between courses and across disciplines in terms of the objects and practices of evaluative judgement. Different kinds of task are involved, drawing on different knowledge domains and different kinds of knowledge within them. In our considerations, we need to keep in mind that, as Ajjawi and Bearman point out in their chapter (Chapter 4), standards reside in the practices of academic and professional communities and are not always explicit, or explicable. In addition, each discipline and profession has its own signature pedagogies which characterise the ways in which it is learned and these in turn will influence the development of evaluative judgement; see, for example, Dall’Alba (Chapter 2) and Bearman (Chapter 15). While it is not possible to examine these variations here, we can nevertheless begin to focus on what might be some key components of the notion. For each of the components identified, we discuss first how they might be framed and secondly consider design implications which show how the developing judgement can readily sit alongside other aspects of study and assessment.

Discerning quality

Once it is agreed that a given type of work needs to be produced, students need to appreciate what constitutes good work of that kind. As Sadler points out, it is necessary to “induct students into sufficient explicit and tacit knowledge of the kind that would enable them to recognise or judge quality when they see it and also explain their judgements” (Sadler 2010, 542). Without being able to recognise good work, they are unlikely to be able to produce it for themselves. A key part of this capability is discerning quality. While some features of high-quality work end up being codified in standards and criteria, such condensed statements might not be a good starting point for students. After all, if students could fully decrypt the language of standards and criteria they would probably already be operating at a level superior to that of the course they are studying.

After reaching a basic familiarity with the kind of work that is to be produced, and having developed some of the discourse and knowledge needed to discuss it, a useful starting point can be to expose students to samples of work of various qualities so that they can begin to appreciate what distinguishes one from another – perhaps through the use of exemplar pedagogies like those discussed by Carless et al. (Chapter 11). This is a process of discernment through induction. Learners are given practice in noticing what differentiates one piece of work from another. “Work” is used here to describe any product or performance that learners aspire to be able to enact for themselves.

Discerning quality is a central feature of variation theory, developed by Ference Marton. He argues that learning is a process of making increasingly sophisticated discernments. The expert can distinguish many more features of a phenomenon or problem than a novice. In this view, learning is the building of a repertoire of variations in the phenomenon of interest. This body of work provides many studies and illustrations of how quality can be discerned (Bowden and Marton 1998).

Developing the capacity of students to discern quality is not new. It has been advocated by authors on study skills for at least forty years. A simple example of this approach was established by Graham Gibbs in the 1970s (1981). First-year undergraduate students are given examples of essays written in response to an assignment similar in kind to the one for which they are preparing. Through discussion in small groups, students distinguish which essay was better than the others. They are then asked to identify exactly how one is better or worse than the others. These differences are noted and a list of distinguishing features assembled by the group. Thus, a set of distinguishing features is generated and students are introduced to the notion of criteria. Students take these features and use them to prepare their own essays. The power of the process is that students do not need to be given direct instruction on quality features, and so long as the exercise is undertaken in diverse groups with sufficient discussion, students realise that they have the capacity to begin to discern quality for themselves. Later variations on such approaches are taken up in the chapters by Henderson et al. (Chapter 12), Thompson and Lawson (Chapter 14), Tai and Sevenhuysen (Chapter 16) and Carless et al. (Chapter 11), who enter into some specific discussions on how we can get this to occur in various contexts and configurations.

Discernment of quality is not something that begins or ends at first year, as Nelson recognises (Chapter 5). It is a persistent feature of learning and practicing anything, as illustrated in the chapters by Carless et al. (Chapter 11), Henderson et al. (Chapter 12) and Rees et al. (Chapter 18). Thus, opportunities to prompt discernment of quality can be included at transition points in programmes, when students move from producing one kind of work to another that is qualitatively different. For example, students moving from small-scale reports or essays might need to revisit their discernment of quality when faced with producing a research report or Honours project. Indeed, there is some suggestion that moving from one form of assessment to another that requires a quite different approach might require additional efforts in discernment, as students tend to regress a little when producing work in a new form (Boud, Lawson, and Thompson 2015).

While making judgements of the work of others and benefiting from feedback processes is likely to develop connoisseurship, it does not necessarily lead to learners being able to generate good work themselves. So then, what do we need to assist learners in the transition from appreciating the work of others to generating good work of their own? A learner could develop high levels of evaluative judgement, and be assessed as such, but not be able to create good work. The transition from discernment to manifestation is an important one.

Judgement processes

How is a judgement made with respect to any learning outcome or type of work? This is one of the most underdeveloped areas in our understandings of evaluative judgement. When educators pass judgement on students' work they traditionally use implicit yardsticks to judge quality, and they are not very good at tapping their own communities of practice to reaching a common view (Price 2005). More simply, they compare students against each other to gain a spread of marks. While the latter process has the virtue of being much easier to undertake, it is now no longer acceptable. In an era of explicit learning outcomes, students must be judged against a defined standard using appropriate criteria (Boud 2017). Officially, comparisons between students are residue of an earlier policy era; however, these comparisons are difficult to escape from as they have shaped much current assessment practice, and there is evidence they may contribute to the calibration of our judgement.

Judgement against a standard, rather than with respect to other students, is now required. It is the norm in most work settings. In technical areas, the notion of a standard is quite straightforward. Indeed, industry quality standards for products and processes are ubiquitous

and are, for example, published internationally by the International Organization for Standardization (ISO). Companies are held accountable for consistently meeting ISO norms. In most aspects of university work it is not possible, or even desirable, to be quite so formal or rigid. However, some notion of a standard is needed, both for certification of students and to assist them in meeting these requirements. While this may come from a written specification, standards are more typically embedded in examples of work: in texts, in course materials and in captured performances. These provide a yardstick for judgement.

Identifying and utilising standards is a central feature of learning. While standards may be provided to students in many courses, and in vocational courses they are enshrined in professional standards statements, they invariably need to be contextualised and explicated for any given task, and they need to be performed, as Ajjawi and Bearman point out (Chapter 4). Working with and appreciating standards is a pre-requisite for developing judgement based on them. Modelling of this use is needed from teachers, who can represent the kinds of processes that they are helping to prompt in students.

Modelling and embodiment, or even representations of standards, are insufficient in themselves. Rubrics present an example where standards are contextualised by teachers but may remain opaque to students. While rubrics (a marking grid representing the various learning outcomes for an assignment and criteria and levels of performance for each) are manifestly a representation of the standards required for a given assignment, they are often not in a form that can be readily appreciated or, when feedback is provided in terms of the categories of the rubric, understood. An example of how this might be addressed is given in the chapter by Ajjawi and Bearman (Chapter 4), which describes how students can go through a process of co-constructing a rubric for an assignment and thus appreciate, through direct experience, how standards and criteria are created and instantiated in the form of a rubric.

Judgements may be made through comparison with the work of others, as given in the example above, or with oneself, as when a learner seeks to improve their own work from previous assignments or through working on drafts. Students need to discern the key features of their own work, and that of others, and then judge what else is needed to improve their work.

Fostering judgement processes requires opportunities within courses for students to engage in practicing judging and to receive feedback information on the judgements they make. This is needed whether or not rubrics are involved, and can be undertaken on a regular basis with each new class of students. Importantly, students must not just appreciate the standards that they will apply; they must also get opportunities to compare their judgements with those who have greater expertise. An example from our own teaching experience is to ask students to appraise their work against a set of features of good work that they have generated in class. These are used to compile an assignment attachment sheet, on which each student writes qualitative statements indicating how their accompanying assignment addresses the criteria that have been agreed. When their marked work is returned, they receive comments solely on how well they have judged their own performance. This focuses on discrepancies between their own judgement and that of the assessor. Thus, feedback here is primarily focused on improving students' judgement of work of this kind, and secondarily on improvement of the work itself. This process can be enhanced when, prior to submission, students share their draft with a peer, receive comments and modify their own work accordingly. Of course, this can only be done effectively in a standards-based context in which students are not penalised when the work of their peers improves. Further examples of how feedback may be optimised to develop students' evaluative judgements can be found in the chapters by Henderson et al. (Chapter 12) and Johnson and Molloy (Chapter 17).

Managing biases

All judgements are subject to beliefs, assumptions and biases of different kinds, as Joughin and Dawson point out (Chapters 6 and 10 respectively). It is the nature of human judgement that a fully objective assessment cannot be made. The best we can do is to recognise our own biases and seek to manage them. This is just as much a problem when teachers or assessors judge students' work as it is when students do it for themselves or each other. A whole repertoire of approaches has been developed to manage biases in marking, which can be adapted to the present context (e.g., the Better Judgement website).

The most important step for this aspect of developing evaluative judgement is to accept that the judgements that anyone makes will be biased, and to recognise where the sources of bias may reside. As well as the biases and their sources discussed by Joughin (Chapter 6), the most common bias in self-judgement is the desire to do well and interpret one's own work in the best possible light. For example, when confronted with a marking scheme with model answers, many students are tempted to give themselves the benefit of the doubt. For this reason, it is vital for students to access multiple sources of information and be open to a multiplicity of judgements by others that they are encouraged to reflect on. These sources include different teachers, different peers and anyone to whom their work is visible or who could be invited to view it. We recognise our biases through seeing our work through the eyes of others. Temptations for bias are even stronger when there are incentives for distortion, for instance through the allocation of grades or rewards for the quality of the work, rather than the quality of the judgement. Strategies for promoting metacognitive checking of evaluative judgement processes are presented by Lodge and colleagues (Chapter 7).

Assessing quality trustworthiness of sources and others

Any given learner usually has access to considerable material and human resources that could potentially assist them in making effective judgements of their work. However, some of these sources could lead to worse judgements, or potentially even harm the development of their evaluative judgement capability. How then can they know what to rely on and what to be sceptical of? As far as material resources are concerned, in the era of the Internet this is part of the basic requirement of distinguishing what counts as a trustworthy information source (or not) for any given purpose. For example, when online source evaluation is not explicitly taught, students tend towards heuristics that can be somewhat unhelpful, such as looking to confirm pre-existing beliefs or going on the reputation of the source rather than the quality of the information (Metzger and Flanagin 2013). The development of skills for judging the quality of sources of information need to be considered alongside other aspects of learning skills development around the use of sources and evidence for any aspect of academic work. Interventions intending to develop students' information literacy or domain-specific critical thinking skills could fruitfully consider these capabilities through an evaluative judgement lens.

When it comes to direct information from other people, further considerations apply. A feature which may be overlooked in making judgements is an affective one. When undertaking a feedback process with another person, how does one judge what reliance to place on their comments? Can they be trusted in what they provide? This is already an active but overlooked consideration when students receive information from staff (Telio, Regehr, and Ajjawi 2016). Some comments are believed and others are not. The basic question to be asked is: does this person have my best interests at heart and do I think that what they say has credence? Only when the answer is in the positive can the information provided be regarded as secure. With no knowledge of the other person, they can only be judged in terms of their position (a teacher, a peer) and on the content and manner of the information they offer.

Information which appears to be disrespectful can be ignored, and the rest treated cautiously. The question to consider is: is what the other person communicates significant enough and trustworthy enough to take the effort to act on it?

The converse of judging the trustworthiness of others is being trusted oneself; that is, ensuring that information provided can be relied upon and be seen to be trustworthy by the other person (Carless 2013). Trust is a reciprocal process: trust begets trust, and when offering comments to others it should be expected that the trustworthiness of the provider is taken into consideration by the recipient. Maximising one's own trustworthiness to others involves being interested in their learning, eliciting what they want, appreciating the appropriate criteria to use and communicating that to them, pitching the level of comments suitably for what you know of them, and contextualising information suitably; in short, knowing what you are doing and communicating it authentically.

A challenge comes when there is different information coming from different sources. Clearly, the value of what others say relative to the work considered is of first importance; balancing different inputs then involves making judgements about the source and the perspectives they hold, and their relationship to the recipient and the work.

Seeking opportunities for practice

Evaluative judgement is a complex process which needs as much practice as does generating the work that is being judged. It is an embodied process that needs to be enacted, not just studied or thought about, as Dall'Alba emphasises (Chapter 2). She draws attention to the importance of developing evaluative judgement over time, as students become the professionals they aspire to be. Development needs to be undertaken across the different domains and subject matters of a programme. It is not a generic process which, once learned, can be applied universally, but rather, as Bearman highlights, is instead embedded within disciplinary and professional practices (Chapter 15). This implies that multiple opportunities are required for the practice of judgement across a course or programme, many of which require accompanying feedback processes using multiple sources of feedback information, as discussed by Thompson and Lawson (Chapter 14) and Ellis (Chapter 13). The obvious occasions for practice are when tasks are undertaken as part of the normal curriculum. However, when new ideas are being introduced or new processes of development are used, there may be a need for explicit instruction, practice and feedback in how to judge appropriately. As Goodyear and Markauskaite discuss (Chapter 3), it is also important for students to gain exposure to situations in which intrinsic feedback from the environment can "bite them back", allowing students to thus learn from the world in which they operate rather than from any deliberate pedagogic intervention.

A well-designed curriculum would map these occasions and opportunities for practice over units and years, and provide mechanisms to allow students to track their development. This requires digital enablement to store artefacts, record information about judgements and feedback, allow easy navigation across occasions, and provide clear summaries of progress in developing judgement as well as meeting programme learning outcomes. There are various devices to aid different aspects of this (for instance, see Thompson and Lawson, Chapter 14), such as e-portfolios, specialist platforms (e.g., Re:View) and dash boards; however, they need to be well integrated into the usual learning management systems to be effective. Forums for reflection are also needed, to assist learners in utilising the information available through such enabling devices to take responsibility for managing the development of their own evaluative judgement across a programme. Such approaches require a programmatic view of curriculum design and assessment, rather than the fragmented one common in a unit-centred approach to courses. As tempting as it may be, developing one course unit focused on

evaluative judgement is unlikely to have much effect; evaluative judgement must be integrated into day-to-day activities across units.

Enabling conditions of courses

While we have emphasised the need for activities to be included in the curriculum to promote the development of evaluative judgement, such activities cannot be enacted independently of the many other aspects of teaching and learning that pursue other desirable goals. From the learning-centred perspective we have been discussing in this chapter, the most important condition is emphasised by Boud and Falchikov (2007) in their notion of promoting informed judgement: the self-identification by students as active learners. This refers to the need for all students to navigate the transition between seeing themselves as *consumers* of courses – the responsibility for which lies with others – to *learners*, as they realise that they are the principal agents for their own learning and that only at their own initiative can they benefit from the opportunities available to them. Different students make this transition at different points, some long before they enter university. It is only through an effective transition pedagogy, which starts in first year and ensures that all students have shifted in this way (Kift 2015), that effective development of evaluative judgement can proceed. Transition is nonlinear, and continues throughout and beyond the university; evaluative judgement should be revisited when major transitions occur, such as when learning key threshold concepts, or when entering placements in the workplace.

Several contributors have identified ways in which consideration of self-regulated learning can provide useful ways of thinking about practicing evaluative judgement and the enabling conditions of courses, along with specific suggestions about how it can be introduced. Introducing evaluative judgement practice is one important way in which self-regulation can be promoted and developed, as discussed by Bennett (Chapter 9), Panadero and Broadbent (Chapter 8) and Lodge (Chapter 7).

The effective use of the elements of evaluative judgement development outlined here are contingent on having good quality feedback processes throughout a programme (Boud and Molloy 2013). These involve the careful selection of tasks, sequencing of activities so that capacity for evaluative judgement can build systematically, and feedback comments arranged from multiple sources (teacher, peers, etc.) and timed such that they can be utilised for the next task. Feedback for the development of judgement needs to be considered programmatically, not just in relation to each unit; capacities need to build over time and across different subject matter.

Conclusion

In conclusion, there is a clear agenda to be pursued to enable the general goal of developing learners' evaluative judgement within higher education. Individual teachers will readily find ways of doing so, drawing on examples given in this book. The more substantial challenge is to develop learners' evaluative judgement systematically across the curriculum across many diverse disciplinary areas. At first sight, it may appear that this is an additional burden for an already overloaded programme. However, closer consideration suggests that this agenda is simply one of ensuring that graduating students are able to do what the curriculum claims. A focus on evaluative judgement is a focus on the core feature of education itself: there is no point in graduating students who do not know what they can and cannot do. They need to be able to judge good work when they see it – whether it is their own work or that of others. Anything less is a failure on the part of the institution to graduate capable students who can continue their learning.

References

- Bowden, J., and F. Marton. 1998. *The University of Learning: Beyond Quality and Competence in Higher Education*. London: Kogan Page.
- Boud, D. 2017. "Standards-based assessment for an era of increasing transparency." *Scaling Up Assessment for Learning in Higher Education*, edited by D. Carless, S. Bridges, C. Chan, and R. Glofcheski, 19-31. Dordrecht: Springer.
- Boud, D., and N. Falchikov. 2007. "Developing assessment for informing judgement." In *Rethinking Assessment for Higher Education: Learning for the Longer Term*, edited by D. Boud and N. Falchikov, 181-197. London: Routledge.
- Boud, D., R. Lawson, and D. Thompson. 2015. "The calibration of student judgement through self-assessment: disruptive effects of assessment patterns." *Higher Education Research and Development* 34 (1): 45-59. doi:10.1080/07294360.2014.934328.
- Boud, D., and E. Molloy. 2013. "Decision-making for feedback." In *Feedback in Higher and Professional Education*, edited by D. Boud and E. Molloy, 202-218. London: Routledge.
- Carless, D. 2013. "Trust and its role in facilitating dialogic feedback." In *Feedback in Higher and Professional Education*, edited by D. Boud and E. Molloy, 90-103. London: Routledge.
- Flinders University. "Better Judgement." Better Judgement Website. <http://www.flinders.edu.au/medicine/sites/better-judgement/>.
- Kift, S. 2015. "A decade of Transition Pedagogy: A quantum leap in conceptualising the first year experience." *HERDSA Review of Higher Education* 2: 51-86.
- Gibbs, G. 1981. *Teaching Students to Learn: A Student-Centred Approach*. Milton Keynes: The Open University Press
- Metzger, M. J., and A. J. Flanagin. 2013. "Credibility and trust of information in online environments: The use of cognitive heuristics." *Journal of Pragmatics* 59: 210-220. doi:10.1016/j.pragma.2013.07.012.
- Price, M. 2005. "Assessment standards: the role of communities of practice and the scholarship of assessment." *Assessment & Evaluation in Higher Education* 30 (3): 215-230.
- Sadler, D.R. 2010. "Beyond feedback: developing student capability in complex appraisal." *Assessment & Evaluation in Higher Education* 35 (5): 535-550.
- Telio, S., G. Regehr, and R. Ajjawi. 2016. "Feedback and the educational alliance: examining credibility judgements and their consequences." *Medical Education* 50 (9): 933-942. doi:10.1111/medu.13063.4649