

Potentials for transforming assessment for learning in the digital environment: the central role of self-judgement

David Boud
University of Technology, Sydney

Abstract

This paper addresses the issue of how the digital environment can enable new approaches to assessment for learning to be introduced in higher education courses. In particular, it focuses on the problem of how digital technology can provide opportunities for students to develop the capacity to make effective judgements about their own work. It starts from consideration of the importance of assessment for learning and what features it demands of any teaching and learning process. It identifies aspects of the digital environment that promote assessment for learning and those that inhibit it.

The paper argues that, whatever else it does, the digital environment must provide greater opportunities for students to practice making judgements about their own work than is possible in face-to-face teaching, because being able to make effective judgements is the hallmark of any graduate who is able to continue learning after graduation. The paper uses the example of a digital environment *Re:View* to show how undergraduate students' self-judgements of their work can be fostered and tracked over time.

Introduction

In recent years there has been substantial rethinking of the role of assessment in learning in schools (Black and Wiliam, 1998), and in higher education (Boud and Falchikov, 2007). This rethinking has generated a new emphasis on how assessment can be positioned to have a more productive influence on how and what students learn through a move away from a preoccupation with certification and measurement towards reconceptualising how it can operate to foster learning.

This paper examines the implications of this shift in assessment thinking and focuses on how a digital environment can aid a transformation of assessment practice. It argues that we must focus on core educational purposes involving fostering learning before giving attention to decisions for purposes of certification. This reverses the dominance that certification thinking has held over assessment practice for many decades. The paper explores this shift through focusing on developing the capacity of students to make judgements about their own learning. It examines a tool developed at the University of Technology, Sydney (UTS) that facilitate students making judgements about their work with respect to detailed standards and criteria. It uses this practice to highlight what digital environments can best contribute to assessment for learning and what should be avoided even though it may be easy to do.

My own position about assessment and digital technology is as follows. Assessment is the single greatest influence on students for good and ill. It profoundly influences what and how they learn. It communicates what educators most value and it sends messages to students about what their priorities for study should be. Historically, digital technology has been co-opted in this process, but it has sometimes operated in an unsophisticated way to entrench behavioural models of learning and emphasise low-level knowledge (e.g. through excessive

use of poorly constructed multiple-choice tests). It has done this not because of technical limitations or lack of potential to do more, but because the most common ways in which it has been used have been based on a transmissive view of knowledge acquisition. For example, online systems have been populated with poorly developed, memorisation-type items that do not focus attention on the full range of learning outcomes of courses. The major trap to be avoided is using the digital environment to lock in a primitive view of assessment. While it has been often misused to emphasise inept forms of assessment, it nevertheless offers substantial potential to transform assessment experiences. This can only occur if we are clear about how assessment influences learning and how it is experienced. The challenge is to consider both purposes and conceptions of assessment and the affordances of digital technology.

Changing positioning of assessment today

Despite much work over many years, assessment has not been transformed to really attend to concerns about learning. We cannot pretend that assessment in everyday situations in most institutions is satisfactory. There are frequent, often messy, compromises between the needs of assessment for certification and for learning. Changed thinking about assessment is needed so that assessment is not mainly about formal structured assessment activities, but is about all those things that enable us (teachers *and* learners) to make judgements about learning. This includes many activities not normally thought of as assessment (e.g. understanding standards, modelling good practice, etc.). Assessment has been preoccupied with the concepts and technologies of educational measurement that focus on attributes of norms, tests and instruments rather than how assessment practices act on students to shape their behaviour and dispositions. Notions of reliability and validity, for example, and methods such as the multiple-choice test have a place, but not the, often dominant, one that they are positioned as having.

Over the past ten years or so there has been a major reorientation of assessment away from an almost exclusive focus on the priority of certification and measurement of student performance towards one that places learning and the way assessment can be used positively to shape study and learning activities, and thus student attention. While Black and William's (1998) paper is often used as a marker of this change, these ideas—mainly that assessment should be seen primarily as a formative influence— have found a productive home in all sectors of education and in different countries (e.g. Knight and Yorke, 2004; Carless, Joughin and Liu, 2006; Havnes and McDowell, 2008).

This emphasis on the formative has opened the way for a rethinking of assessment practice on a wide front (Boud and Falchikov, 2007). This includes not only a general focus on the impact of assessment on learning, but also reconsideration of how assessment is conceptualized. Boud and Falchikov have argued that, whatever else it does, assessment should fundamentally be about building learners' capacity to make informed judgements about their work. The rationale for this view is that the person being acted upon in assessment is the learner, and ultimately all influences through assessment and other educational interventions are manifest through the judgements learners make about their work. Only learners can decide to learn, we must conceptualise assessment as if it were a process of shaping and informing learners' judgments. Educationally, assessment is good or bad, effective or ineffective, only to the extent to which it leads to improved judgements on the part of learners about their work. In the world of practice beyond educational institutions, learners as workers make judgements about whether their work does or does not meet appropriate standards. We should be preparing students for this situation and developing in

them the capacities and dispositions that will lead them to be effective learners and judges of their learning throughout their lives.

Such a view leads to the idea that assessment activities need to satisfy many purposes: the important long-term educational capacity-building one outlined above, but also more immediate purposes of assisting students to learn about whatever they are currently engaged on, and also their grading and certification. This reverses the *de facto* view that judgement by others for certificatory purposes comes first and that the needs of learning is subordinate to that. Of course, judgement by others is vital for certification, but only when framed within the larger educational goal of informing and developing students' own capacity to judge. Assessment in practice then always needs to do double duty—for immediate purposes and for the longer-term. Acts of assessment must be judged according to whether they work for both ends.

Key features of quality assessment

Building capacity for making informed judgements is not the only concern of assessment design. Other elements are needed if it is to guide and shape learning. Previously, drawing on a model of environments that shape learning design (Boud and Prosser, 2002), I have summarized features of assessment design as follows (Boud, 2007):

What principles should a great assessment design follow?

- A. The activity itself is not just a test but also a learning experience in its own right
- B. Assessment activities are seen as valid and worthwhile by the learner
- C. It actively promotes learning and skills beyond the immediate act itself
- D. The learner is positioned as an active agent throughout
- E. The activity has a strong positive backwash effect on learning during the course prior to assessment
- F. It enables students to celebrate and portray achievements
- G. It is part of a sequence of great assessment designs over the course as a whole
- H. It arises from a productive and stimulating learning environment
- I. It is not excessively resource-intensive
- J. It requires and prompts informed judgement on the part of the learner

In short, good assessment design always promotes high quality study activities and discourages cramming for tests or focusing students' attention on any matters other than producing meaningful learning outcomes that meet the full range of curriculum expectations. Assessment needs to be completely aligned with pedagogy and purpose. This necessarily means that conventional tests and examinations are no longer the main elements of an assessment regime. For further discussion of these issues see <http://www.assessmentfutures.com>

How does the digital environment fit with this view of assessment?

The discussion so far has been neutral about the environments in which such developments can occur. However, the digital environment offers great potential for enacting the features discussed. It provides opportunities that are hard to realize in non-digital environments. Nevertheless, before dealing with the positive features that the digital environment provides we should pause to reflect that despite the wonderful innovations that have been implemented and have been celebrated internationally, the everyday experience of students and tutors with assessment online and in other digital spaces is often far from positive. For example, at a recent online conference about the use of digital technologies, I

elicited responses from participants that indicated that it was not uncommon in e-assessment to find the following negative features:

- Excessive use of close-ended questions
- Over-assessment with relentless testing
- Students working in isolation from each other
- Lack of tutor involvement beyond initial design
- Despite appearance of personalization, lack of explanations that connect with individual student's experience

The digital environment in itself can lock in a primitive view of assessment or embrace a more creative one. Whether it does one or the other is not a function of the digital environment, but what conception of assessment is adopted. When teachers are confronted with a new digital environment, they can be easily tempted back into older views of assessment and not incorporate those that they are still trying to fully understand.

Pay-offs of digital technology for assessment for learning

As we are approaching a time at which we can assume that all students will have access to a range of straightforward digital technologies and be familiar with their basic operations without additional instruction, a range of possibilities is now available. While the uptake of these may differ in different societies with different resources, there is a qualitative change occurring worldwide.

What then are the particular advantages of digital technology for assessment for learning? The most fundamental of these is the shift from 'teacher performing teaching' to 'learner performing learning'. That is, there is a fundamental change of perspective from prioritizing what teachers do to what students do. Without this shift, little of the new technologies can be used. While digital media can of course be (mis)used as solely transmissive mechanisms, digital technologies permit and encourage participation. Certainly in their recent manifestations in social networking, they discourage passivity and foster interaction. This means that they can be used to increase the time students spend on meaningful tasks and engage them actively in most processes. Students can spend more time-on-task, probably the most ubiquitous factor identified in what contributes most to learning (Hattie, 2009). Learning involves students being pro-active and *doing* rather than *listening/reading* and digital environments can be set up to do this. In fact, they are increasingly having a pre-disposition to do this, in that the ubiquitous technologies of Web 2.0 are all participative and engaging. This is unlike the traditional classroom that is structured and predisposed towards transmissive and non-participative interaction.

The second advantage follows from the first. With interactions, and the continuous record of digital activity, come greater opportunities for feedback and responsiveness. Teachers have access to an extended repertoire of ways of responding to a variety of students and a variety of student needs. They do not need to spend their time on knowledge transmission and explanation so they are released to focus on enabling students to engage with knowledge, understand it, learn how to produce useful work and help them monitor the development of their skills. There is a qualitative shift from the teacher telling to the teacher helping students engage.

The third advantage comes from the radical change in relationships that occur between and among students and with students and others in the digital environment. Communication with each other is not discouraged as it may be in the traditional class, but positively encouraged. Collaborative learning can flourish

because students can communicate with each other all the time independent of time and location. Working with others doesn't involve complicated planning and a particular physical environment. It is commonplace.

These advantages are mere potentials for action until they are realized. Any opportunities that technology provides can be inhibited. We must therefore look at the potential negatives. By and large these are not negatives of the environment, but how it is used. They are therefore in the control of the teacher and others designing the curriculum.

The first is the possibility of distraction. Digital environments are seductive and offer so many opportunities that it is easy to be distracted into activities that might be interesting, but do not keep the goal of learning and the particular learning outcomes desired, firmly in mind. Literature on e-learning is packed with interesting developments whose connection with basic educational purposes is not evident.

The second is the possibility of regression on the part of the teacher and their desire to move back to comfortable old ways of acting. These include being in control, of fostering student-teacher interactions rather than student-student interactions, of using technology as more efficient transmission or testing mechanisms without the rethinking that is required to deal with the challenge of thinking seriously about learning and how it can be fostered.

There are particular traps to be avoided that can inhibit learning. These include the possibility of yet more ways to constrain students and produce compliance, of not involving students in meaningful decisions about their own learning and not realizing the new opportunities for feedback that don't involve unilateral judgements by the teacher. The particular trap, but also affordance of the technology itself is that it is always there, never tired. But because the basic elements of a course normally need to be pre-coded it can lock teachers in to early decisions and not be sufficiently responsive. Balancing this, databases are available to support assessment tasks and good records can be kept for both students and teachers. Lots of things are easy to do, but they don't transform assessment and don't necessarily support learning. In short, we need to resist the temptation of seeing the digital environment as a behavioural technology rather than a participative and constructivist one.

Of course, there are many other positive affordances such as greater user-control, the availability of multi-media and multi-modal activity and so on. But the basic ones suffice for consideration of assessment.

What are the most strategic uses of digital technology to enhance learning through assessment?

There are three key areas for development of assessment arising from the discussion above: feedback, collaboration and the development of self-judgement. (Other implications for assessment are well discussed in JISC, 2010).

Feedback

Feedback in non-digital environments suffers from the very widespread misconception that the process is complete when useful information is provided to students. We know though that for feedback to have a positive effect, the impact of information on students' work must be identified and comments adjusted accordingly (Narciss, 2008). Reproducing conventional assignment completion and marking schemes does not lead to feedback. It is only through knowledge of the effects of information provided to students that a feedback

effect can be observed and the kinds of information most useful for improving learners' work determined. Digital applications need to enable this to happen. They can do so by, for example, speeding turn-around times on assignments, enabling richer comments with less demand on teachers than are possible in written comments (each through personal audio or video comments) and by enabling changes in student work to be readily tracked. They can even help us move away from the transmissive model of feedback in which the teacher is the possessor of useful information that is communicated unilaterally to students to one in which students identify the kind of feedback they need and can respond to teachers to enable it to be given in ways that are helpful to them. Taking this a step further, feedback from other students and other parties can be readily organized in a digital environment.

Collaboration

Collaboration is a strong feature of digital environments and has been discussed extensively. However, what does it make possible for assessment? It clearly facilitates the use of group tasks. Students do not need to be in the same place or work at the same time in order to collaborate on joint activities. The ease of conduct of joint tasks means that more can be deployed and that group activities can be used for formative purposes with relatively little teacher input.

Collaboration can be one to one or as part of a larger team. Most important of all for assessment is that the collaborative feature can address one of the biggest gaps of all in conventional assessment thinking: the almost exclusive dominance on assessment of individual students by teachers. One to one collaboration can lead to students commenting on each other's work and the active use of peer assessment as a positive pedagogic strategy can be readily deployed. The collaborative facilities of digital technology are providing one of the main ways in which the persistent criticism of educational institutions by employers—that they do not develop the vital skills of collaboration—can be addressed.

Self-judgement

While self-testing with immediate guidance on all answers is easy to do with digital technology, and is especially worthwhile in technical areas, the development of students capacity for self-judgement involves far more than this. The feature that needs to be added is that of identifying and applying standards and criteria to ones' work. Competent graduates need much more than advanced knowledge, they need to be able to use this knowledge to undertake complex tasks and tell for themselves that they are making a good job of this. They do not do this in isolation from others so they also need collaborative skills as part of self-judgement and they need to know how to use information from other people in productive ways to change what they do.

One example in this area is the e-portfolio. E-portfolios have been developed to provide a place for students to portray their work and reflect on it, but they have been under conceptualised and it is often difficult to know what any given type of portfolio is contributing to learning outcomes. Nonetheless, development in this area is of great significance and the potential for innovation is great.

Another example is the range of special purpose tools that have been designed to aid cooperation and collaboration in assessment and that involve students making and sharing judgements. From my own context at UTS, SPARK^{PLUS} and *Re:View* are good examples. SPARK^{PLUS} is a web-based self and peer assessment kit (<http://spark.uts.edu.au>). It not only enables students to confidentially rate their own and their peers' contributions to a team project, but also allows students to self and peer assess individual work and improve their judgment through benchmarking exercises. *Re:View* is a web application developed to aid marking, feedback and graduate attribute development (<http://reviewsecure.com>).

The use of Re:View to explore student judgements

Re:View is an example of a digital technology tool developed by Darrall Thompson in the School of Design at UTS. It is now a UTS product available for use elsewhere. *Re:View* can be set up for anything from a single module of study to the entire program for an undergraduate degree. It provides a space in which the task-specific criteria for every item of assessment can be recorded and made visible to students and for comments to be made by tutors about these in the light of students' completed work. There is also an opportunity for students to be rated and to rate themselves according to these criteria for each task. *Re:View* enables students to self-assess online against these criteria before the lecturer or tutor adds their marks and comments. It can be set to any grade categories but as it is designed to give criteria-based comments to students, it does not display percentage marks. It is not the official repository for marks and does not replace any official marks and grades systems administered by the university.

Student self-assessment is an option that academics can select for each task. Students can enter *Re:View* and rate themselves before having the ratings of tutors revealed to them. While *Re:View* has been mainly used as a vehicle for feedback and for tracking the development of graduate attributes across a program of study, we have been using it to research the phenomena of students making self-judgements about their work.

There have been considerable studies over a long period of time comparing students' marks with those of teachers (e.g. Falchikov & Boud, 1989). These show that students are reasonable judges of their own grades, but that the accuracy of prediction varies according to the expertise of the student and the level of course: stronger students are more likely to underestimate grades, weaker students overestimate; students in advanced courses are more likely to underestimate, students in introductory courses overestimate. It follows from this that it would not be surprising to find a tendency that when students encounter new subjects, their ability to make accurate judgements of their work declines. The judgement by students in terms of their rating against defined criteria and standards provides as close as we can get to an empirical study in 'natural conditions' without making an additional intervention in a course.

As part of the embedding of graduate attributes as an explicit part of all assessment tasks across an entire undergraduate course in design, data were available from students for all assessment tasks on their own rating (prior to knowledge of the rating or the grade they were awarded), and the rating awarded by a lecturer or tutor on the same scale. Our study involved tracking the assessments students made of their own work against that of the marker across all assignments in an undergraduate program in a Bachelor of Design program in an Australian university.

The total number of students in the undergraduate degree programs was approximately 1300 taking a total of 148 subjects (course modules) in any one year. Of those data was available from a total of 49 subjects for the three-year period from 2006 to 2009. Each subject had up to four tasks examined by varying numbers of tutors according to subject size.

Use of *Re:View* was voluntary for students. We selected to use data from students who had used *Re:View* to self assess over a minimum of two semesters. This meant that even though we had self assessment scores matched with tutors marks for over 13,000 criteria we only used 2196 self assessments from 182 students. As the study was focusing on long term effects of self-judgement over

time we also conducted analyses on students who had used *Re:View* over three semesters (63 students, completing 765 self assessments), and four semesters (19 students, completing 288 self assessments).

The data used in the study was the students' individual ratings for each assessment task semester by semester. Each task had a range of criteria and percentage marks were collected for each criterion as well as the total percentage mark for each task. The tutors' actual ratings of students' performance was also gathered for each criterion in each task in each subject for comparison with students' ratings. Qualitative ratings on a sliding visual scale were converted into percentages for analysis.

Two of the questions we addressed are (details can be found in Boud, Lawson, Thompson and Brew, 2010):

1. Do student's marks converge with tutors within a subject?
2. Does the gap between student judgements and tutor judgements reduce over multiple tasks?

Our data show signs of the following:

The frequently reported phenomenon (Falchikov and Boud, 1989) of students of higher ability being more accurate or even under-rating themselves, and students of lower ability over-rating themselves was repeated in our data.

Although students initially struggle to accurately self assess, with time and benchmark scores from their tutor they appear to get more accurate. This is also consistent with previous studies on this topic.

Students rate themselves significantly higher than tutors in the first task of their first three semesters of self-assessing. This gap reduces over time. By the fourth semester there was no significant difference between students and tutors in the first task in each subject. This suggests that students need time and practice to adapt across subjects to make accurate judgements.

Although students' ability to accurately assess increases within a subject, when they begin a new subject, a difference between their rating and the tutor rating is again evident. This may be due to having to understand a new set of criteria and standards for each subject and so would suggest that the increase in accuracy of judgement is not immediately transferrable.

This finding is shown again when we examine the accuracy of students' judgements from their first attempt at self-judgement to their last. Again those students who self assessed for three semesters or less did not show that they were able to accurately self assess when compared to the tutors marks, but those students who self assessed for more than three semesters did show more accuracy in their judgements.

This is an initial study that we do not claim is anything other than suggestive. It gives tentative support to the idea that students can become more effective at judging their performance with practice. But it does so only with students who volunteer to undertake the process. It suggests that there is a regression effect when students are confronted with a new subject: there is a tendency for them to initially to be less effective in judging their performance than in a subject with which they have had prior practice. There is also an indication that students become more effective in rating themselves over semesters.

The study suggests directions for further work rather demonstrating anything on which we can securely build. Nevertheless, it shows the potential for digital

technology to do quite different kinds of assessment work using conceptions of assessment different from those typically used in studies of student performance. We intend to explore these issues further with fuller data sets and with data from other institutions that have used this process.

Implications

This paper shows some possibilities for transforming assessment to make it more directly focus on key learning outcomes. Digital technology has a key role to play in this as it can facilitate a shift from a teacher and course-centric view to one that acknowledges the agency of students and the need for them to build capacities to make judgements of their own work as a central goal of education. *Re:View* is one of many emergent tools that can contribute to this process.

The transformation we need in assessment is not about technology as such, nor about techniques and methods. It is about the centrality of learning and ensuring that whatever it takes to pursue the sophisticated formation of learners occurs. It is also about seeing assessment as a central element in promoting learning, an element that should be in the hands of students as much as of teachers.

References

- Black, P. & William, D. (1998). Assessment and classroom learning, *Assessment in Education*, 5 (1), 7–74.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Education*, 22 (2), 151–167.
- Boud, D. (2007). Great designs: what should assessment do? Keynote address, Re-Engineering Assessment Practices International Online Conference on Assessment Design for Learner Responsibility, 29-31 May 2007. <http://ewds.strath.ac.uk/REAPO7>
- Boud, D. and Falchikov, N. (2007). Developing assessment for informing judgement. In Boud, D. & Falchikov, N. (Eds.) (2007). *Rethinking Assessment for Higher Education: Learning for the Longer Term*. London: Routledge, 181-197.
- Boud, D., Lawson, R., Thompson, D. and Brew, A. (2010). Does students' practice in grade prediction calibrate their judgement? Paper presented at the Conference of the International Society for the Scholarship of Teaching and Learning, Liverpool, October.
- Boud, D and Prosser, M. (2002). Appraising new technologies for learning: a framework for development, *Educational Media International*, 39, 3&4, 237-245.
- Carless, D., Joughin, G. and Liu, N-F. (2006). *How Assessment Supports Learning: Learning-oriented assessment in action*. Hong Kong: Hong Kong University Press.
- Falchikov, N. & Boud, D. (1989). Student self assessment in higher education: a meta-analysis, *Review of Educational Research*, 59 (4), 395–430.
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Havnes, A. & McDowell, L. (Eds.). (2008). *Balancing Dilemmas in Assessment and Learning in Contemporary Education*. New York: Routledge
- JISC (2010). *Effective assessment in a digital age: a guide to technology-enhanced assessment and feedback*. <http://www.jisc.ac.uk/digiassess>
- Joughin, G. (Ed.) (2009). *Assessment, Learning and Judgement in Higher Education*. Dordrecht: Springer.
- Knight, P. T. and Yorke, M. (2004). *Learning, Curriculum and Employability in Higher Education*, London: Routledge.
- Narciss, S. (2008). Feedback Strategies for Interactive Learning Tasks. In

Spector, M J, Merrill, M D; van Merriënboer, J and Driscoll, M P (Eds)
Handbook of Research on Educational Communications and Technology.
Third Edition. New York: Lawrence Erlbaum, 125-143.