

Accepted version of Broadbent, J., Panadero, E. and Boud, D. (published online 21 June 2017). Implementing summative assessment with a formative flavour: A case study in a large class, *Assessment and Evaluation in Higher Education*, DOI: [10.1080/02602938.2017.1343455](https://doi.org/10.1080/02602938.2017.1343455)

Implementing summative assessment with a formative flavour: A case study in a large class

Jaclyn Broadbent¹, Ernesto Panadero² & David Boud^{1,3,4}

¹Centre for Research on Assessment and Digital Learning, Deakin University, Geelong, Australia.

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

³Faculty of Arts and Social Sciences, University of Technology Sydney

⁴Institute for Work-Based Learning, Middlesex University, London

Corresponding author email: jaclyn.broadbent@deakin.edu.au

Abstract

Teaching a large class can present real challenges in design, management, and standardization of assessment practices. One of the main dilemmas for university teachers is how to implement effective formative assessment practices, with accompanied high quality feedback consistently over time with large classroom groups. This article reports on how elements of formative practices can be implemented as part of summative assessment in very large undergraduate cohorts ($n = 1500$ in one semester), studying in different modes (on- and off-campus), with multiple markers, and under common cost and time constraints. Design features implemented include the use of exemplars, rubrics and audio feedback. The article draws on the reflections of the leading teacher, and discusses that for summative assessment to benefit learners, it should contain formative assessment elements. The teaching practices utilised in the case study provide some means to resolve the tensions between formative assessment and summative assessment that may be more generally applicable.

Keywords: formative assessment, summative assessment, large classes, audio feedback; exemplars

Introduction

The type of assessment practices used in the classroom have a major impact on students' learning and academic achievement (e.g., Black & Wiliam, 1998). For example, summative assessments are used for grading purposes to enable comparisons between learners, and to ensure standards are met (Shute & Kim, 2014). On the other hand, formative assessments, also known as assessment for learning, are used for enhancing students' learning and the development of self-regulated learning practices (Nicol & MacFarlane-Dick, 2006). Generally, formative assessment is thought to be an effective strategy for learning (Shute & Kim, 2014). However, even if instructors are aware that some assessment practices are more beneficial than others, university teachers may face several constraints that affect their choice of assessment practice.

This is particularly so for large classes. Teaching a large class can present real challenges in design, management, and standardization of assessment practices. These challenges include: (1) reduced and sometimes absent face-to-face class time for online/distance learners; (2) heavy reliance upon sessional staff to maintain acceptable staff-to-student ratios; (3) issues of equity and consistency across multiple campuses and multiple marking and teaching staff; (4) challenges in finding ways to provide high quality, individual feedback; and (5) the need to provide formative assessment experiences while still meeting summative needs. These challenges often mean that established, quality pedagogy – which is often designed, tested, and evaluated in much smaller contexts – requires modification to meet the needs of large class teaching. In turn, these concessions threaten to undermine the quality of the pedagogy.

Although there is much research available regarding quality pedagogical assessment practice, there has been little research showing how these elements can be transferred to the large class context. The aim of this paper is to show how formative

assessment elements can be integrated in very large cohorts of undergraduate students, under the common cost and time constraints of the Australian higher education sector. The following case study is based on a subject with an annual enrolment of more than 2100 students. The subject's biggest semester has 1500 students, divided over four campuses (including an online campus) and split between 25 teachers with approximately 30 students per tutorial. While these assessment practices were implemented as a means of addressing and reducing some of the challenges of assessment in large class teaching, rather than for research purposes, it is worthwhile to present these and discuss their implications for consideration by other lecturers with large groups.

Formative assessment and its relationship to student learning

Since the seminal work by Black and Wiliam (1998), formative assessment has become one of the most prolific areas of research within education. Formative assessment is “contextualised and aims to build a comprehensive picture of learners’ characteristics. It is an integral part of a learning process, and it takes place several times during a course rather than only at the end” (Strijbos & Sluijsmans (2010), p3). It often contains a number of assessment features such as (a) the role — characterised as assessment for learning; (b) the frequency — intermittent and often; (c) the format — constructed responses and authentic contexts; and (d) the feedback – global and specific with suggestions on how to improve (Shute & Kim, 2014). By contrast, summative assessment — while beneficial for comparing learners and ensuring standards are met — is often thought of as a major event, infrequent, objective, with feedback that focuses on the completed assessment event, if feedback is considered at all (Sadler, 1989).

There is evidence to suggest that formative assessment improves student outcomes such as increased academic performance, self-regulated learning, self-efficacy, etc. (e.g.

Black & Wiliam, 1998; Kingston & Nash, 2011; Panadero & Jonsson, 2013), with formative feedback having been shown to be the single most important factor in learning (Hattie & Timperley, 2007). For this reason, it is crucial for teachers to consider how every assessment practice and associated activity is arranged, and the purposes behind them (Boud, 2000a; Brown, 2004).

Despite the above benefits, tensions exist in the practical implementation of formative assessment. First, assessment is often constrained by the need for university teachers to produce numerical marks and grades to be formally recorded by the university by the end of the semester. In Australia, this often means that formative assessment practices are additional to the required summative (for grades) assessment. Second, if formative assessment practices can only be added if summative assessment needs are already met, the workload of teaching staff need to be taken into consideration. This is particularly the case if the formative assessment requires frequent testing (i.e., continuous evaluation), and is to be accompanied by sufficient quality feedback information for students. Lastly, if formative assessment practices are provided and workload needs can be met, there is a dilemma in how to encourage students to engage with the tasks, when students are reluctant to undertake any tasks which are not graded. For example, a large-scale study by Jessop, El Hakim and Gibbs (2014) involving 23 different programmes, found that “that most students did not value, complete or even notice the presence of [ungraded] formative assessment tasks” (p. 77).

These are tensions that are relevant to both large and small class contexts. Giving quality feedback information to students can be a time-consuming activity and resourcing may be expected to increase in cases of personalised feedback. However, in large class contexts it is not as simple as giving more marking per marker or having more markers, as there is a need for quality checks and evaluation of consistency in marking. The larger the

number of students the more complex it is to deliver that type of feedback and, no less important, to verify that the information influences the student's future achievement (Boud & Molloy, 2013a).

The larger volume in marking also produces problems that do not necessarily occur on a smaller scale. First, substantial marking loads necessitate reliance on sessional marking staff to help assess all assignments (i.e. staff who are paid on a casual, hourly basis). In Australia, for example, Percy, Scoufis, Parry, Goody & Hicks, (2008) identified that 80% of all undergraduate first-year marking and teaching was completed by sessional staff, and there are no obvious signs that this has lessened since. Second, a large number of markers and increased reliance on sessional staff increases the risk of variable quality. Skill set, level of education, expertise, and experience often varies between markers, with some sessional staff having limited prior marking experience or expertise in the particular assignment set. Training these markers in providing suitable information that is likely to influence students' subsequent work is therefore essential. However, organising meetings and checks of consistency can be challenging when markers are part-time, located at different campuses, great distances away, or where students are enrolled in a variety of study modes (on-campus, off-campus and blended study modes).

Given the above constraints, it is not always practical, or even cost or time effective, to add formative assessments on top of the summative assessments required to meet university needs. However, there are a number of ways to implement assessment in large classes that maximise the release of formative information with a lower cost effort for the teacher. In this case study, there are three of special interest: rubrics, exemplars, and audio feedback.

Rubrics

Research about rubrics has grown exponentially in recent years (Dawson, 2017). Rubrics provide details of the standards by which students' assessment can be judged for quality, and the extent to which learning outcomes have been met (Panadero & Jonsson, 2013). Rubrics are documents containing evaluative criteria, quality definitions, and a scoring strategy (Popham, 1997). They can serve summative purposes while increasing the reliability and validity of multiple scorers or one scorer evaluating several pieces (Jonsson & Svingby, 2007) and formative purposes (Panadero & Jonsson, 2013). In relationship to the latter, students should be able to use the rubric to self-assess their work by these same standards, before submitting that piece of work for grading.

Nevertheless, students do not always have a strong understanding of notions of quality captured by the rubric because of the sometimes-limited descriptive information provided in the rubric, and their appreciation of the concepts and terminology used (Dawson, 2017). One way to develop students understanding is to embed rubrics in classroom activities via modelling and feedback (Panadero, 2011) or to co-create them with students (Boud, 2000b; Fraile, Panadero & Pardo, 2017). However, while these strategies are well suited to small class situations, co-creating rubrics is more difficult in a large class. How does one co-create a rubric with 1500 students in a course, with 50 tutorials and 25 teachers? Even if there were dedicated tutorial time allocated for explaining and critiquing the rubric, not all students would attend, and not all teachers would provide the same learning experience.

Exemplar exposure prior to completion of assessment

Exemplars provide illustrations of what addressing a task well would look like. They allow teachers to share knowledge of different quality work (Sadler, 2002), “exemplars are not standards themselves but are indicative of them ... they specify standards implicitly”

(Sadler 1987, p200). Exemplars can be implemented in multiple ways (essay, poster, oral presentation, etc.), and can be constructed from previous student work, current student work, or created by teaching staff (Carless, 2015). Exemplars have been found to be well received by students (To & Carless, 2015), and are a cost-effective way to provide feedback information (Scoles, Huxham, & McArthur, 2012). Exemplars provide concrete examples that help the student to understand what constitutes quality work, and to appreciate the bases upon which quality of work is judged (Sadler, 2002). Further, exemplars can provide insights into the student's own work and refine their understanding of what is required (Handley & Williams, 2011). For greater gains, exemplars with appropriate commentary (e.g., in-class discussions) can help students understand the marking criteria and standards (Hendry, Bromberger, & Armstrong, 2011). While class discussion of exemplars may be more easily implemented in small classroom settings, the challenge for large classes is ensuring that consistent and equitable discussions occur between all teacher and student groups, and that suitable commentaries are available. One solution to this problem is the use of annotated online exemplars, which have found to be effective in numerous studies (Bell, Mladenovic & Price, 2013; Handley & Williams, 2011).

Audio Feedback

Feedback has been defined as:

“a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, in order to generate improved work” (Boud & Molloy, 2013b, p. 6).

If we want to ensure that it is productive then it needs to involve more than provision of information; it requires the establishment of conditions that make it likely that useful

information may be taken up and acted upon, leading to improvement (Lipnevich, Berg & Smith (2016). The nature of the comments provided are of great significance, as some types of comments are more worthwhile and likely to lead to action than others (Hattie & Timperley, 2007). Providing feedback information is not the same as justifying the grade awarded; it has quite a different purpose for influencing students' subsequent behaviour. There is little feedback information in a grade or a number, especially on how to improve and correct aspects that went wrong. Thus, to be part of a formative process the feedback needs to have clear goals (where am I going?) qualitative information about current performance (how am I doing?), as well as information about how to improve subsequent performance (where to next?; Hattie & Timperley 2007). However, it is also evident that high quality feedback takes a large effort to produce, is not always valued by students, and in a large class context it can be difficult to disseminate in a timely and consistent fashion.

One way to overcome the challenge of disseminating feedback to large student cohorts is through audio information instead of text-based or written information (Lunt & Curran, 2010). Providing students with audio information can overcome the time and location constraints that arise from engaging in face-to-face, individualised discussions (Jonsson, 2012). Studies in methods of feedback to students have shown that compared to written feedback, audio feedback can provide significantly more detail and depth, be more personal, allowing for greater expression, tone, and nuance, and is often preferred by students over written feedback (Carruthers et al. 2015; Lunt & Curran, 2010; Merry & Orsmond, 2008; Nemec & Dintzer, 2016). For the marker, giving audio feedback has been shown to be quicker to provide than written comments of the same quality (McCarthy, 2015), and it provides a sense of teacher presence through students hearing the teacher's voice (Oyarzun, Conklin & Barreto, 2016). Audio feedback is not without its practical challenges of course, as students have reported difficulty downloading large audio / video files (McCarthy, 2015),

difficulty playing the files (Henderson & Phillips, 2015), difficulties replaying specific parts of a long audio file, and it is also unsuitable for students with hearing impairments (Lunt & Curran, 2010). Teaching staff have reported difficulty finding a quiet environment to record their comments (Henderson & Phillips, 2015) and lack of familiarity in using technology to provide audio feedback (Cann, 2014).

Aim of this study and problems to be addressed

As mentioned earlier, the need to produce numerical marks and grades for summative purposes and improvement-oriented information for formative purposes adds to workload for both staff and students. We believe to meet both needs, summative assessment should have formative elements. Although there is much research available regarding good pedagogical assessment practice, there has been little research showing how elements can be transferred to a large class context. In particular, how does one translate formative assessment practices successfully into large classes, while still meeting summative assessment needs? For example, how do you add formative elements such: (a) focusing less on assessment outcomes and support learning during the learning process, (b) providing high quality feedback to students so they can improve their future performance, (c) having iterative low stakes assessment, and (d) ensuring students understand standards?

This paper outlines how formative assessment practices can be integrated with summative assessment in a context involving 1500 students (across multiple campuses, taught by 25 teachers in 50 tutorials) by (a) using exemplars with detailed explanations of marking rubrics to focus on learning and understanding of standards, (b) using small-stakes iterative linked assessment that scaffold tasks to support learning, and (c) giving high quality audio feedback information consistent across markers, that guides the student for the next assessment aimed at enhancing their understanding of quality for that task.

A case study focuses on the design features of the course unit, the progressive reflections of the teacher, student satisfaction regarding the unit, and on the issues that have been confronted in multiple iterations. Since challenge and level of approach require explication not often afforded in standard empirical pieces, our intention is to both illuminate the implementation of ideas often discussed in the abstract and to be of practical value to those confronting large scale teaching challenges. The following description is based on teaching practices developed and implemented in a large classroom environment by the first author over the years 2010-2016, inclusive. Support for these practices is demonstrated through university student satisfaction surveys, in-class student surveys, online resource access data, and student results.

Method

The educational context

The course unit

This case study is based on an Australian University first-year psychology subject, which is part of an accredited psychology course. There are requirements from both the university and the course accrediting body that psychology subjects are awarded summative grades. Within this context, University policy stipulates that no more than 20% of the mark allocation can be from online unsupervised tests, and that the maximum weighting of any assessment task (including examinations) is 60% of the mark for the unit. Further, feedback mechanisms that extend beyond providing a score are strongly encouraged. This means that multiple assessment tasks are used in each subject, often with personalised feedback, and this has workload implications. If formative assessments are to be implemented, they are additional to the summative assessment tasks in the unit, and if personalised feedback is to be provided for these formative tasks, there are added workload and cost implications.

The subject in the case study focuses on the psychological aspects of health behaviour change and has an annual intake of 2100+ students across 3 semesters, co-ordinated by the first author in the role of unit chair. In the largest cohort (1500+) unit content is taught and assessments marked by around 25 sessional staff who each teach and/or mark 1-3 groups of 30 students. The unit is delivered in blended mode across four campuses (hundreds of km apart), as well as in an online-only mode (approximately 300 of 1500 students enrolled). Students enrol from over 40 different courses, which means students' academic preparedness, discipline experience, backgrounds, capabilities, and learning needs differ widely; therefore, engaging all students is a significant challenge. The unit is staffed and funded at the same level of resourcing per student as any other in the faculty.

Assessment activities

Students complete several assessment pieces to determine their final course grade. This includes 10 quizzes (worth 10% combined) each consisting of 10 multiple choice questions and available to students online via the learning management system (LMS), an end of semester 100 question multiple choice exam (worth 45%), and three linked journal assessments spread out over the 11-week semester, with an in-class test regarding the content of the journal assessments (for a total of 45%). The three-linked journal assessment task is the particular focus here.

The journal assessment is an authentic assessment task. Authentic assessment is a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills (Mueller, 2005). This particular assessment is designed to engage students in their own health behaviour change (e.g., increasing frequency and/or duration of exercise), and to apply their learning of

behaviour change to their own context. Throughout the semester, students are taught a range of concepts and techniques (e.g., self-efficacy) designed to progress someone towards a health behaviour change goal. Importantly, the goal is self-established by each student to increase agency, and by initiating self-goals we aim for activation of self-regulated learning strategies.

Within the three linked assessments, each assessment requires the student to video him or herself reflecting upon, and giving insight into, progress on his/her health behaviour change, as well as completing a structured written component that incorporates the concepts of behaviour change they have been learning about. The skill and understanding required by the student to complete each journal is built upon from the previous journal, so that learning that has occurred early can be reused later to improve performance. The assessment is scaffolded in level of challenge and withdrawal of support, with each subsequent journal assessment getting progressively more difficult with less support e.g., the third journal assessment component requires more independent critical thinking skills and evaluation, with less dedicated class time to support the activities than the first or second journal assessments. See Figure 1 for the assessment outline.

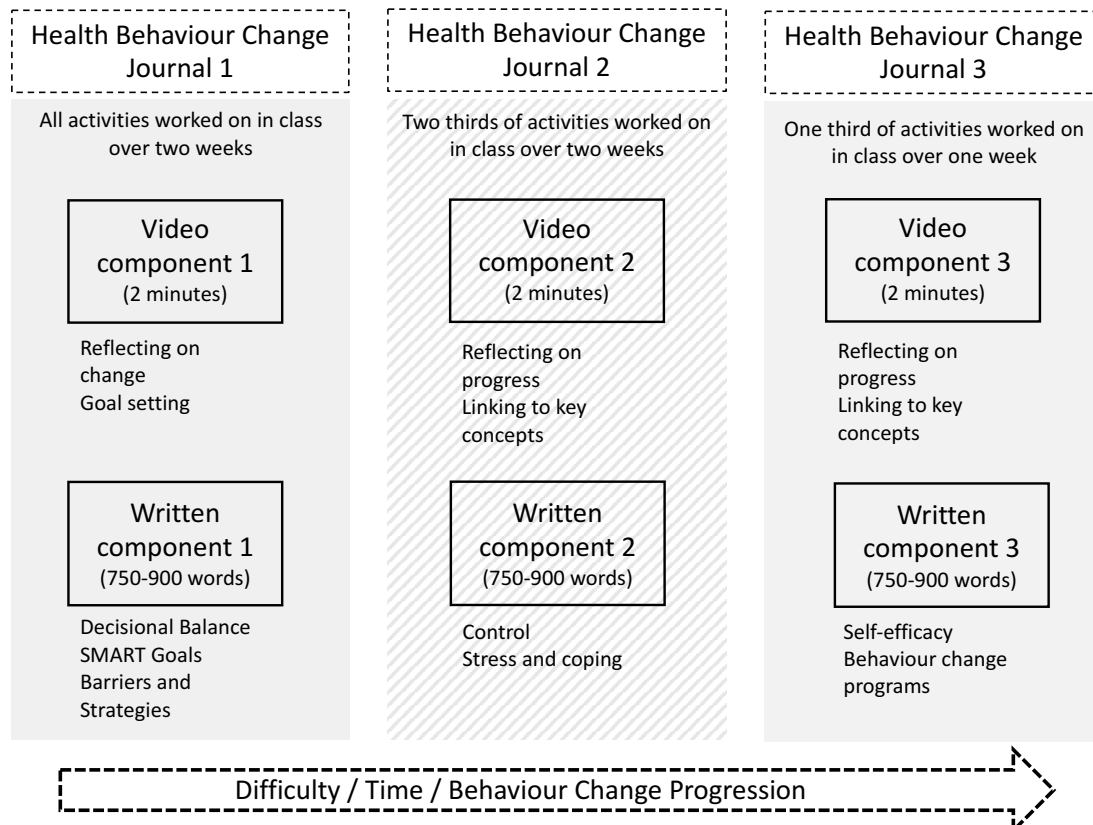


Figure 1. Three linked health behaviour journal assessments

The journal assessment includes deliberate design features such as: (a) being authentic to encourage student agency and self-regulation (Schmitz, Klug & Schmidt, 2011), (b) building in challenge over time, while (c) targeting similar skills throughout each of the assessments. The journal assessment tries to balance the tension between formative and summative assessment. First, the iterative nature of the journals provides students with opportunities to take on-board feedback and improve over time – a key feature of assessment for learning (Boud & Molloy, 2013a). A key consideration when determining the number of assessment tasks was to ensure that the tasks were not so small that they became meaningless, and that there was enough time between the returning of comments and the next assessment so that students could improve their learning. Harland, McLean, Wass, Miller, and Sim (2015) highlight the disadvantage that too many very small

assessment pieces can have on student's learning (e.g. controlling student's behaviour through assessment and a decrease in the quality of learning). Second, students are often reluctant to undertake any tasks which are not graded (Jessop et al., 2014), thus the low-stake grades awarded for each journal (worth 11.25%) are used to motivate students to perform the tasks. The grading of the tasks also has a secondary benefit of providing an assessment of learning to fulfil the university requirements associated with the unit. Lastly, while some of the characteristics of this assessment do increase workload for teaching staff, it is less than if formative assessments were to be added as an addition to the summative assessment.

Learning and assessment practices that provide a formative emphasis

Using an exemplar to develop a student's notion of quality

As discussed earlier, students do not always have a strong understanding of the notions of quality captured by the rubric (Dawson, 2017). To ensure students understand what quality looks like, receive the same messages about the standard by which the rubric is applied, an authentic online exemplar was created, which includes a detailed application of the rubric. The exemplar used unfolds over the same time period as the students' own assignment. A real health behaviour change (by the unit chair) is documented, rather than a fictitious case, because it is intended that students relate to the experience of behaviour change in real time and vicariously experience both the assessment and the process of behaviour change. Further, this kind of behavioural modelling should be more powerful than ways in which exemplars are typically used. The online exemplar follows the same structure and questions as the journal assessment (although as by journal three, some questions are only partially answered), is of a high distinction standard (above 80%), is assessed using the

same rubric used to grade the students' journal assessment, and is made available online prior to the students' assessment submission date. In addition to this, the lecturer provides an in-depth recorded video explanation of the rubric's application, discussing how the marker makes a judgment about the application of the rubric, and explaining why the exemplar met a particular criterion, how the exemplar could have been improved, and potential pitfalls that were avoided and what poor performance may look like.

While students have been shown to benefit from seeing a range of performances (McConlogue, 2015), it was decided that only one exemplar of high quality would be used, rather than multiple exemplars of differing quality. It was reasoned that one exemplar better fitted the aim of having an authentic exemplar (discussed above). Furthermore, as students would already see three exemplars (one for each journal iteration), there was concern that providing too many exemplars may overwhelm them, and as they were first-year students, they may find it difficult to differentiate between levels of quality given their inexperience with university assessment. As only one exemplar was shown, it was decided that the exemplar should be of high quality, as there was concern that students may be unable to isolate and then analyse the qualities of a poor-quality exemplar and work out how to improve it (Handley & Williams, 2011). Lastly, by creating and sharing an online exemplar and rubric discussion, we ensured the same message reaches every student in the same way, without the message being changed or diluted by 25 different sessional staff.

Providing feedback to students so they can improve their next performance

A key to formative feedback is to improve learning (Shute, 2008). In the case study journal task, we provide information that not only focuses on current performance, but focuses on supporting learning during the learning process so they can improve their future performance. Markers give personalised information on each of the student journal

assessments, provided in the form of a five-minute (on average) audio recording, alongside a graded rubric. Note: Recordings are chunked to a maximum of three minutes per recording, and separate recordings are made for separate sections for ease of navigation. Markers are trained to highlight what the student has done well and how s/he could improve performance. The information given to students addresses both specific issues in the current journal assessment, as well as detailing how they can improve in subsequent journal posts. Some examples of the types of comments given by markers during the audio recording can be seen in Table 1.

Table 1. Examples of the features of information given by markers

Salutation and interpersonal communications	Addressing student by name, may include commenting on personal circumstances (either known from class, what they have written in their journal, or other communication e.g. email), or on effort, or prior performance.
Avoid ambiguous statements	Ambiguous statements are avoided, and concrete examples are used to increase the student's understanding. For example, while lovely to hear, statements like "that was fantastic" and "you did a great job", do not provide the student with any information about what exactly they did well.
Explicit statements about good performance	Markers explicitly state what the student did well, so that the student is able to replicate the skill e.g. 'I liked how you argued X by giving examples. I particularly like when you wrote X and Y....'.
Explicit statements about improving poor performance	The same practice as above is applied to any criticisms of the student's work, particularly as students may not have done 'X' because they don't know how.
Constructive guidance on how to do better in the next journal	At specific touch points, assessors use strategies to help orient students to what quality looks like in successive assessment (e.g. feedforward). An example of information used to influence subsequent behaviour is: <i>"However, what was just missing here was really explaining how those strategies would actually work to increase your sense of control. This is going to be really important to do in your next assessment in question one where you need to come up with strategies of how to increase your self-efficacy for each of Bandura's four ways of building self-efficacy..."</i>
Send off and good luck	Mostly relationship building, encouraging students to incorporate comments into the next journal, giving best wishes, and invitation to discuss further.

There are also specific aspects of each journal designed to scaffold or replicate learning from the previous journal assessment. At these touch points, assessors use strategies to help orient students to what quality looks like in successive assessment tasks. This information goes beyond justifying the grade awarded to look ahead to the next assignment, offering constructive guidance on how to do better in future work. Each marker is required to give at least two statements of feedback information that specifically links to the next assessment piece. This information helps clarify what high performance encompasses in relation to future assessment, possibly facilitates self-reflection, and aims to elicit a student's best possible performance in subsequent assignments – all aspects of formative feedback (Hattie & Timperley, 2007). The feedback information for each assessment is returned to students with sufficient time for them to take it into account in their next journal assessment.

Feedback information is only useful if it is of high quality and consistent.

As discussed earlier, providing high quality feedback in large classes relies on sessional marking staff with the risk of variable quality from markers of varying degrees of capability. So, while moderation practices are not directly related to formative assessment practices, it is important to ensure consistency among markers - both in terms of information provided by them to students and their interpretation of the rubric. As such, a rigorous grading and feedback moderation process is implemented at the start and mid-way through the marking process. Like many moderation processes, detailed resources are made available for markers and a subsample of student assignments are blind double-marked (see below Figure 2). The unit chair as well as one or two other senior staff (the moderation team) complete all the moderation. During blind double marking, personalised audio feedback information is sent to each marker from the moderation team offering examples of how to enhance their feedback. This provides an opportunity for them to discuss their understandings

and re-mark the student's work. This process is repeated until each marker demonstrates the level of quality and consistency required by all markers across the cohort. Markers are also encouraged to observe others' performance, and this feedback loop enables the markers to judge how their marking compares to others can also aid moderation (for more information see Broadbent, *in press*).

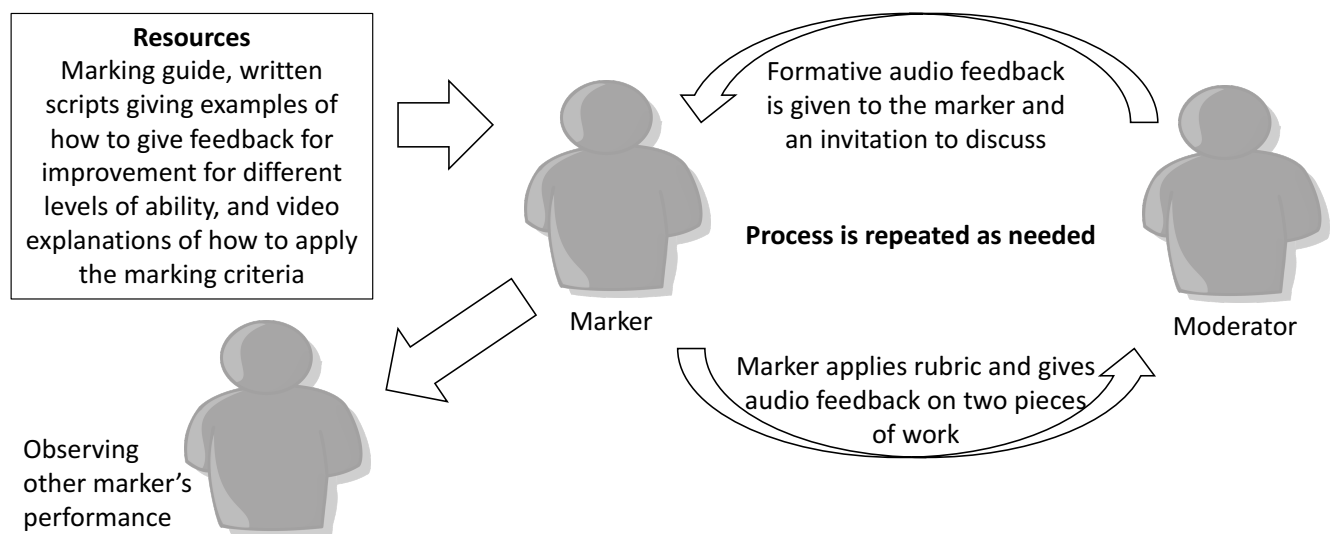


Figure 2. The marking and audio moderation process.

Figure 3 provides an overview of the overall feedback process; that is, the implementation of: (1) an exemplar to allow the student to observe another performance and to view a discussion of how the rubric is applied; (2) iterative linked assessment, that is scaffolded in difficulty, with audio feedback information containing information that can be used in the next assessment; and (3) a rigorous moderation process to reduce disparate understanding of assessment standards and to ensure consistency in marking and feedback from tutors.

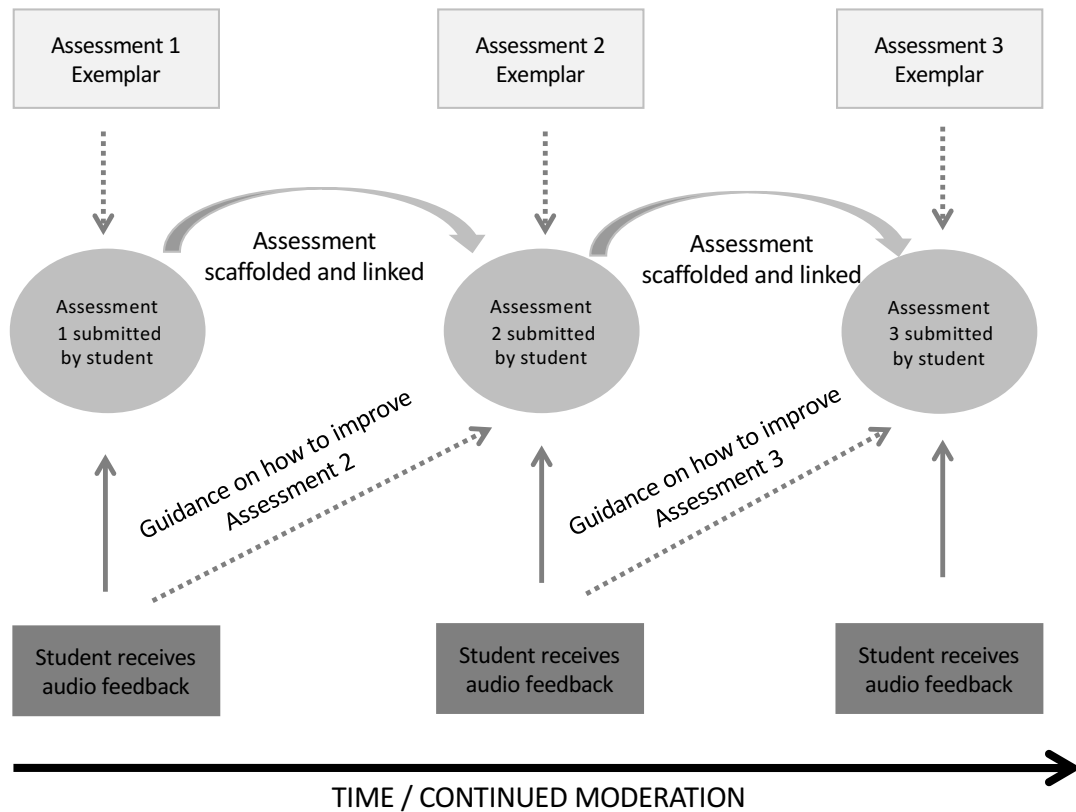


Figure 3. Diagram of how the assessment, exemplar and feedback in the case study fit together.

Supporting data

Support for these practices is provided through university-wide student satisfaction surveys, in-class student surveys, online resource access data, and student results.

Results

Student satisfaction surveys

Within the university, students are surveyed regarding their satisfaction with each of subjects in which they are enrolled. Of interest here is the data for the question “Feedback on my work in this unit helps me to achieve the learning outcomes”. While this question related to feedback over the entire semester, and is not solely related to the health behaviour change journal assessment task, we believe the data lend support to the feedback methods used in this subject. As illustrated in Table 2, students were highly satisfied with the feedback they received. Table 2 also shows that with the addition of each component (exemplar, audio feedback information, and design to influence students’ subsequent work) over time, student satisfaction appears to correspondingly increase. While these data do not demonstrate causation, it is generally consistent with the approaches outlined here. For example, in 2010, students could view the exemplar (although it had no audio discussion of the rubric), and received written instead of audio feedback information, with no guidance on future assessment from their marker. During this time, student satisfaction for feedback received surpassed the University averages for the same question (72%; versus case study subject 79%). We can surmise that even implementing the exemplar alone (without audio feedback information / or guidance on future assessment) had a positive impact on student satisfaction. The addition of audio feedback information in 2011 (instead of written comments) again corresponded with an increase in student satisfaction from 79% to between 87% - 90%, averaging 88.3% over six semesters. A further corresponding student satisfaction increase occurred when marks had to give at least two feedback statements that linked to the next assessment piece (2014). Satisfaction scores increased from 88% to 94%. As training and moderation of markers was improved, satisfaction continued to increase, and resulted in a rise to very high satisfaction levels of 94% - 99% agree or strongly agreed in 2014-2016. The University average, as a comparison, 81% agree or strongly agreed in semester 2, 2016. Most importantly, 75% and 66% of students in the case study across semesters 1 and 2 in 2016

strongly agreed that the feedback was helpful, compared to only 39% of students in the University average (see Table 2 for details).

Table. 2
Student's satisfaction with the feedback they received

	% Agree
Pre-audio feedback but with exemplar	
BENCHMARK: Deakin, Sem 2, 2010 (n=31598)	72%
HBS110, Sem 2, 2010 (n responses=360)	79%
Audio feedback given	
Sem 1, 2011 (n responses =70)	87%
Sem 2, 2011 (n responses =426)	88%
Sem 1, 2012 (n responses =81)	90%
Sem 2, 2012 (n responses =485)	87%
Sem 1, 2013 (n responses =73)	89%
Sem 2, 2013 (n responses =540)	88%
Feed-forward feedback given	
Sem 1, 2014 (n responses =89)	94%
Sem 2, 2014 (n responses =534)	94%
Sem 1, 2015 (n responses =130)	99%
Sem 2, 2015 (n responses =533)	96%
Sem 1, 2016 (n responses =87)	95%
Sem 2, 2016 (n responses =318)	94%
BENCHMARK: Deakin, Sem 2, 2016 (n responses =29218)	81%

On reflection, we found giving audio feedback to be a time-efficient and cost-effective way to give feedback. Some inexperienced markers initially took longer to orientate themselves to both the assessment and using audio feedback, however we suggest this is often the case for inexperienced markers regardless of mode of feedback. In the current context, markers were paid for a similar amount of time to give audio feedback, as they previously were to give written feedback. While this suggests equity in workload and cost implications across the two modes, the greatest benefit was that the quality of the feedback

improved as a result of implementing the audio feedback (as discussed above). Therefore, we conclude that for the level of quality achieved, audio feedback is a more time-efficient and cost-effective way to give feedback.

Student grades

A further observation in support of these teaching approaches is the relative stability of grades across the three assessments. One might expect that as the journals became more challenging, and as in-class teacher support and assistance from the exemplar was reduced, that students would perform less well in latter journal assessments using the common rubric. However, this is not the case, and students perform similarly across all three-journal assessments, with no reduction in academic achievement, despite the increase in challenge. As mentioned earlier, after submitting three journal assessments, students sit an in-class/online test applying the knowledge learnt in the journals to a novel situation. Despite the in-class test remaining comparable, as the process of giving feedback and training of markers was refined, students' mean scores on this test continue to increase: 66% (semester 1, 2014); 68% (semester 2, 2014); 73% (semester 1, 2015); and, more recently 76% (semester 2, 2015). While learning does not occur in a vacuum and there may be other possible explanations for this improvement, at least at first blush it appears that students were able to meaningfully use the feedback they received in subsequent assessment pieces.

Resource use by students.

Unsurprisingly, the exemplar journal is the most frequently accessed online resource by students in this subject. For example, in semester 2, 2015 ($n = 1553$ enrolled students), the

three online exemplars had 8,157 combined views by students; an average of 5.25 views per student.

In-class surveys

In 2013, students were asked about their perceptions of the online exemplar ($n=309$), 83% of students agreed or strongly agreed it motivated them to learn; 94% agreed or strongly agreed it helped them understand what was required in the assessment; 84% agreed or strongly agreed it helped improve their own work, and 85% agreed or strongly agreed it helped them understand health behaviour change techniques.

Discussion

The aim of this paper was to show how elements of formative assessment can be integrated with summative assessment in very large cohorts of undergraduate students, who are studying in different enrolment modes, with multiple markers, under common cost and time constraints.

Reflections about the use of the key elements: exemplars with rubrics and audio feedback

In this case study, the use of an exemplar accompanied by an annotated rubric provided a way to: (a) disseminate information about how a task will be assessed and the rubric will be applied; (b) grant equity of instruction for students that is at a time (and place) of convenience for them, and importantly also for the teacher; (c) give instruction and information efficiently in a cost-effective manner. Previous literature (e.g., Sadler, 2002) confirms that exemplars are a valuable tool to enhance students' understanding of the marking standards/criteria as well as notions of quality required for the task. As argued by

Nicol and Macfarlane (2006), knowing what good performance looks like and encouraging self-assessment capabilities fulfils the objectives of good formative assessment.

As discussed earlier, co-creating rubrics with students (Fraile et al., 2017), modelling self-assessment practice (Panadero, Jonsson & Strijbos, 2016) and using in-class discussions about exemplars (Hendry et al., 2011) may be more beneficial for students, than the transmission from expert to novice through exemplars with audio explanations, as we have done here. However, in many circumstances these may not be possible to implement. In our case study, we have taken elements of formative pedagogical practice and modified them to fit the exigencies of a large class context. First, by making the exemplar available online, it can be accessed anywhere at any time by the students. Second, by giving a detailed explanation of how the rubric is applied, the online exemplar provides a means for all students to access the same (consistent) information regardless of enrolment mode, physical location, who one's classroom teacher is, or how regularly the student attends class. Third, the authentic, real-person, real-time character of the exemplar anecdotally appeared also to contribute to student engagement, although we have no formal data on this.

The use of multiple small, low stakes, iterative summative assessments in this case study allowed us to adopt some formative practices such as: (a) building students' skills over time, and (b) providing students the opportunity to use their feedback to improve subsequent performance (Hattie & Timperley, 2007; Shute, 2008). It was not feasible for us to have multiple ungraded assessment tasks that required feedback on top the required summative assessment; however, we were able to design an assessment piece that contained smaller low stakes tasks, which built upon the same - or similar set of - skills for a summative grade. Designed correctly, feedback must not only address current performance but should also focus on improving future performance as well (Boud & Molloy, 2013a).

One way to create personalised, detailed, and time-efficient feedback is through audio feedback. As discussed earlier, much more can be said, in a shorter period of time, than can be written. Lunt and Curran (2010) estimate that 6-minutes of writing is equal to one minute of audio feedback. In our case, we found that by replacing written feedback with audio feedback, and restricting markers to the same amount of time that they previously had when giving written feedback, we were still able to improve the quality of the feedback, at no extra cost of time or money. Cann (2014) also found that audio feedback, if used as a replacement, rather than a supplement to written feedback was more time efficient. Providing higher quality feedback information without increasing workloads clearly has its benefits when marking large number of assignments within a semester, especially when the marking teachers do not have to compromise on quality. Further, feedback embedded into the assessment design provides an opportunity to make the provided information useful and usable for students in their future work, and ensure that the feedback loop is closed. Still, providing audio feedback is only beneficial to students when all markers engage with the task. In the current context, most markers embraced the audio feedback, although a few did have what Cann (2014) calls “technical inertia”, where teaching staff unfamiliar with media tools were apprehensive to try audio feedback. We found that if we provided support and a training video on how to use the audio tool, this was easily overcome.

However, feedback information can only enhance students’ notions of quality if the marker is able to make expert judgements about what quality looks like and can communicate to the student how this can be exemplified in their own work. To ensure that markers have a shared understanding of the standards that should be exhibited, that they provide high quality feedback information, and are consistent in delivering this message across the student cohort, a rigorous grade and feedback moderation and training process is required. We used audio feedback to achieve this aim as well. It is more personalised than written feedback, and

overcomes challenges of having to get a large number of people in the same room (or the very least available at the same time) for a meeting, as well as the challenge of having markers at different physical locations. Anecdotally, providing markers with detailed formative feedback helped to develop the markers' skills early, and cultivated their self-sufficiency, accuracy, and expertise in the grading process. This is particularly important in a higher education context when many teaching academics are casually employed, with less opportunity for professional development (May, Peetz & Strachan, 2013).

Conclusions about balancing assessment purposes in a large class

In educational contexts, such as the one presented here, the tension between different purposes of assessment becomes particularly salient. On the one hand, with a large group of students in a subject taught by 25 sessional tutors across different campuses, the summative purpose of having clear and shared standards and scoring systems is a basic requirement. Here, it is of crucial importance to have shared assessment practices and high inter-rater reliability in marking to ensure fairness, to avoid tensions with students, such as wanting to change to the "easiest" teacher's classes, or receive complaints about unbalanced workloads derived from different assessment methods.

On the other hand, focusing purely on the summative purposes is insufficient to enhance students' learning. Aspects such as providing enough opportunities to receive useful information and, even more importantly, to actually use it, engaging with exemplars, rubrics, and modelling, and building the capacity to make judgements of their own learning are also basic requirements for the design of the subject. All these assessment practices turn into powerful pedagogical elements when used with formative and sustainable purposes. Take for example rubrics which have both strong summative effects (Jonsson & Svingby, 2007) and formative effects (Panadero & Jonsson, 2013). When both are combined, a balanced use

should result in a more powerful learning environment, one in which summative and formative practices are aligned and students can have a sense that what is actually being promoted is their learning rather than simply recording their performance (e.g., grade).

A key feature of the overall assessment and feedback design is the deliberate and iterative use of a variety of interventions, which progressively build student capacity for good work and include indicators of success within them. The most important feature of this set of practices is not the use of any particular strategy, but the ways they have been put together. What we have demonstrated here is a proof-of-concept over multiple iterations of high quality feedback, use of formative elements in summative assessment with defensible summative grading within the context of a course unit with large student numbers offered in mixed modes across different campuses as well as online. For teaching practitioners thinking of implementing summative assessment with a formative flavour, based on our experiences in this case study, it is worth considering:

- a) having an exemplar, or set of exemplars, available online;
- b) giving a detailed explanation of how the rubric/ marking criteria are applied to the exemplar, communicated through video, as done here, or by annotation;
- c) If adding formative assessment with personalised feedback is impractical (as it is here due to University / accrediting body requirements for graded assessment, and the increased workload that would accompany the addition of formative assessment on top of the required summative assessment), consider breaking down larger assessment into linked summative assessments that builds upon the same, or similar, skills. This way, students can take their learning from a current assessment piece and apply it to a subsequent assessment piece;

- d) design feedback information so that students' current performance is linked to improving their future performance;
- e) using audio feedback as a replacement to written feedback as way to increase the quality of the feedback (while adhering to the original time and workload restrictions); and
- f) provide formative audio feedback to markers to increase consistency in marking and feedback, in a time-efficient manner.

While this all requires thoughtful planning, it can be done within typical cost constraints so long as the learning implications are fully considered and monitored. We have argued that summative assessment is more beneficial to learners if it takes elements from formative assessment. We hope to have shown that formative elements can be used efficiently, and to the benefit of the students, in large class contexts.

References

- Broadbent, J. (in press). Large class teaching: how does one go about the task of moderating large volumes of assessment? *Active Learning in Higher Education*.
- Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, 38(7), 769-788.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- Boud, D. (2000a). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), 151-167.
- Boud, D. (2000b). Development through self assessment. In Taylor, K., Marienau, C. and Fiddler, M. *Developing Adult Learners: Strategies for Teachers and Trainers*. San Francisco: Jossey-Bass, 63-66.
- Boud, D. & Molloy, E. (2013a). Rethinking models of feedback for learning: the challenge of design. *Assessment and Evaluation in Higher Education*, 38, 6, 698-712.
- Boud, D. & Molloy, E. (2013b). What is the problem with feedback? In Boud, D. and Molloy, E. (Eds) *Feedback in Higher and Professional Education*. London: Routledge, 1-10.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Cann A. (2014). Engaging students with audio feedback. *Bioscience Education*, 22(1), 31-41.
- Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.

Carruthers, C., McCarron, B., Bolan, P. Devine, A., McMahon-Beattie, U., & Burns, A.

(2015). 'I like the sound of that'—an evaluation of providing audio feedback via the virtual learning environment for summative assessment. *Assessment & Evaluation in Higher Education*, 40(3), 352-370.

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360.

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69-76.

Handley, K., & L. Williams. (2011). From Copying to Learning: Using Exemplars to Engage Students with Assessment Criteria and Feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95–108.

Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: high-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528-541.

Hattie, J. & Timperley, H. (2007) The power of feedback. *Review of Educational Research*, 77, 81-112.

Henderson, M., & Phillips, M. (2015). Video-based feedback on student assessment: scarily personal. *Australasian Journal of Educational Technology*, 31(1).

Hendry, G., Bromberger, N. & S. Armstrong. (2011). Constructive guidance and feedback for learning: The usefulness of exemplars, marking sheets and different types of feedback in a first-year law subject. *Assessment & Evaluation in Higher Education*, 36(1), 1–11.

- Jessop, T., El Hakim, Y.E. & Gibbs, G. (2014). The whole is greater than the sum of its parts: a large-scale study of students' learning in response to different programme assessment patterns. *Assessment & Evaluation in Higher Education*, 39(1), 73–88.
- Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63-76.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.
- Lipnevich, A. A., Berg, D. A., & Smith, J. (2016, June). Toward a model of student response to feedback. In *Handbook of Human and Social Conditions in Assessment* (pp. 169-185). Routledge.
- Lunt, T. & Curran, J. (2010). 'Are you listening please? The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education*, 35(7), 759-769.
- May, R., Peetz, D., & Strachan, G. (2013). The casual academic workforce and labour market segmentation in Australia. *Labour & Industry: a journal of the social and economic relations of work*, 23(3), 258-275.
- McCarthy, J. (2015). Evaluating written, audio and video feedback in higher education summative assessment tasks. *Issues in Educational Research*, 25(2), 153-169.
- McConlogue, T. (2015). Making judgements: investigating the process of composing and receiving peer feedback. *Studies in Higher education*, 40(9), 1495-1506.
- Merry, S., & Orsmond, P. (2008). Students' attitudes to and usage of academic feedback provided via audio files. *Bioscience Education*, 11(1), 1-11.

- Mueller, J. (2005). The authentic assessment toolbox: enhancing student learning through online faculty development. *Journal of Online Learning and Teaching*, 1(1), 1-7.
- Nemec, E.C. & Dintzner, M. (2016). Comparison of audio versus written feedback on writing assignments. *Currents in Pharmacy Teaching and Learning*.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.
- Oyarzun, B. A., Conklin, S. A., & Barreto, D. (2016). Instructor Presence. *Handbook of Research on Innovative Pedagogies and Technologies for Online Learning in Higher Education*, 106.
- Panadero, E. (2011). Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics. (*Unpublished doctoral dissertation*). Universidad Autónoma de Madrid, Madrid, España.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.
- Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.) *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 311-326). Springer International Publishing.
- Percy, A.M., Scoufis, S., Parry, A., Goody, A., & Hicks, M. (2008) The RED Resource, Recognition – Enhancement – Development: The contribution of sessional teachers to higher education. Sydney: Australian Learning and Teaching Council. Retrieved from: <http://bit.ly/1VaLoJ6>
- Popham, W. J. (1997). What's wrong – and what's right – with rubrics. *Educational Leadership*, 55(2), 72–75.

- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2), 119-144.
- Sadler, D. R. (2002). "Ah! ... So That's Quality." In *Assessment Case Studies, Experience and Practice from Higher Education*, edited by P. Schwartz and G. Webb, 130–136. London: Kogan Page.
- Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with university students. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 251-266). New York: Routledge.
- Scoles, J., Huxham, M., & McArthur, J. (2012). No longer exempt from good practice: using exemplars to close the feedback gap for exams. *Assessment and Evaluation in Higher Education*, 38(6), 631-645.
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In *Handbook of research on educational communications and technology* (pp. 311-321). Springer New York.
- Strijbos, J. W., & Sluijsmans, D. M. A. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20, 265-269.
- 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*, 40(6), 746-764.