Iterated assessment and feedback improves student outcomes

Lesley J. Morrell

Department of Biological and Marine Sciences, University of Hull, UK

Department of Biological and Marine Sciences

University of Hull

Kingston-upon-Hull

HU6 7RX

UK

Email: l.morrell@hull.ac.uk

Tel: +44 1482 465294

Twitter: @biosciencemum

ORCID: orcid.org/0000-0003-0256-7871

Biographical note: Lesley Morrell is a Senior Lecturer at the University of Hull, and Director of Studies for the Department of Biological and Marine Sciences, and a behavioural ecologist. She is interested in how animals respond to their environments, particularly in the context of anti-predator aggregation and environmental change. She is also interested in how students learn, and evaluating the effectiveness of learning and teaching strategies.

Word count: 6814

Iterated assessment and feedback improves student outcomes

Feedback is critically important to student learning, but reduced frequency of assignments combined with isolated or stand-alone tasks reduces the opportunity for students to engage with feedback and use it effectively to enhance their learning. Here, I evaluate student attainment during a module consisting of eight iterated tasks where the task itself is the same but the academic content differs. At the end of the module, students then self-assess their eight submissions and select two for summative assessment. I demonstrate that achievement increases over the course of the module, and that choice is valuable in allowing students to achieve higher summative marks for the course than their formative marks would suggest. Students who performed more weakly at the start of the module saw the greatest benefits from practice and choice, suggesting that these students particularly can benefit from repeated cycles of feedback and increase their marks.

Keywords: Assessment, sustainable feedback, self-assessment, iterated assessment;

Introduction

Feedback is widely acknowledged as being critically important to student learning (Hattie et al 1996, Black & Wiliam 1998, QAA 2006; Hattie and Timperley 2007), yet it has been suggested that the opportunity for students to engage with feedback has been reduced, due to a documented tendency in Higher Education towards reduced frequency of assignments and use of coursework as formative and summative assessment simultaneously (i.e. where feedback is provided only on summative assessment; Brown et al 1997, Weaver 2006, Price et al 2010, Boud & Molloy 2013). Consequently, students do not always make effective use of feedback when preparing subsequent assignments (Gibbs & Simpson 2004, Orsmond et al 2005, Glover & Brown 2006, Scott

et al 2011), and many staff believe that students take little or no notice of it (Glover & Brown 2006, Crisp 2007). Feedback is known to make a positive difference to learning (Black & Wiliam 1998, Gibbs & Simpson 2004) and forms an important part of the learning cycle (Carless et al 2011), but providing feedback does not always translate to improvement in student work (Sadler 2010), perhaps because staff see feedback as being more important than students do (Brown 2007, Carless 2006), and students see assignments as tasks rather than learning opportunities (Covic & Jonas 2008, Brennan 1995). As a consequence, feedback on summative work may not appear to have the positive impact on student development that it should.

Isolated tasks addressing different learning outcomes also reduce the possibility of effective feedback occurring (Boud & Molloy 2013), particularly if we consider feedback to be a dialogue to support learning, rather than as information provision (Askew & Lodge 2000, Carless et al 2011). When formative feedback is given on a draft piece of work, giving students the opportunity to revise the work before submission for summative assessment, they are more likely to engage with the feedback (Orsmond & Merry 2013), increasing their marks (Barker & Pinard 2014), but often fail to see the value beyond the specific piece of work (Covic & Jones 2008, Orsmond & Merry 2013), even within a subject area (Storch & Tapper 2002). In particular, a lack of awareness of different perspectives and languages (e.g. tutor versus student) on a piece of work may limit their learning opportunities and ability to engage with feedback (Carless 2006, Brown 2007, Sadler 2010, Price et al 2010). Some students, for example, may struggle to use exemplar experiences (Bloxham 2012) to improve their own work (Orsmond & Merry 2013), and high achieving and non-high achieving use tutor feedback in qualitatively different ways (Scott 2017).

Ideally, students will engage with feedback, and learn from it in ways that can be applied to other assignments, in other modules, thus closing the 'feedback loop' (Sadler 1989, Boud 2000, Orsmond & Merry 2013) and supporting the idea of feedback as a dialogue rather than directed information from tutor to student (Askew & Lodge 2000, Nicol & Macfarlane-Dick 2006; Carless et al 2011; Boud & Molloy 2013). In order for students to learn from, and apply feedback more widely, an alternative model to the common feedback-on-draft may help. By creating multiple tasks that overlap, in contrast to tasks that assess different learning outcomes, we can increase the opportunity for students to apply and learn from feedback (Boud & Molloy 2013), which can then act as 'feed-forward' (Scott et al 2011). A key constraint in applying iterated models of feedback is the availability of resources such as staff time, which can significantly limit both the amount and type of feedback that can be given (Boud & Molloy 2013). One potential solution to this is to move away from a reliance on tutor-feedback (and how it can be improved; Sadler 2013), and embrace the value of practice and self-assessment.

Assessment in itself can enhance student learning (Sadler 2013, Dann 2014). Carless et al (2011) suggest that students who engage with feedback and practice writing can develop higher-level scientific writing skills (Carless et al 2011). However, student work for the purposes of learning (rather than as summative assessment) has decreased in higher education, as the emphasis of assessment shifted from formative writing plus final exams to continuous assessment (Boud & Molloy 2013). This has generally resulted in there being fewer assignments, and, together with modularisation, reduced the opportunities for feedback (Brown et al 1997, Weaver 2006, Price et al 2010, Boud & Molloy 2013).

As a result, students have more limited opportunities to practice and develop their writing skills over the course of their degree programme (Hounsell 2007).

Carless et al (2011) proposed the idea of "sustainable feedback", and suggest that this can be achieved by designing assessments that facilitate student engagement over time. Thus, by giving students an iterated sequence of tasks, rather than isolated assessments, with feedback at multiple stages, the value of feedback is clearer, and students are able to develop over time (Carless et al 2011, Boud & Molloy 2013). Boud & Molloy (2013) also suggest that more than one cycle of feedback may be needed for important or challenging learning outcomes, or for "less-responsive" students, but that the effective number of iterations remains an open question. Critically, iterations should take place on relatively short timescales (although with sufficient time for feedback to be provided and acted upon; Evans 2013), to allow for consolidation of the knowledge gained from feedback before it decays (Sadler 2010).

By considering feedback as a dialogue, it should stimulate students to monitor and evaluate their own learning, and through this, feedback becomes increasingly sustainable (Carless et al 2011). Self-assessment is often considered to be key in developing students into self-directed, life-long learners (Nicol and Macfarlane-Dick 2006, Kirby and Downs 2007, Boud and Falchikov, 2006), as it can act to promote a deep approach to learning (Kirby and Downs 2007), generating more responsible and reflective learners (Dochy et al 1999), abilities associated with high achievement (Sebesta & Bray Speth 2017). The ability of students to judge the quality of their own work also improves with practice (Lew et al, 2010, Lawson et al. 2012). Iterated assessment tasks give students the opportunity to see

how assessment criteria are applied by tutors, and move towards using those criteria in their own learning (Scott 2017).

Topics in Biodiversity and Evolution: module description

Topics in Biodiversity and Evolution is a final year (level 6 of a UK undergraduate degree) module, designed to give students an insight into the biological research in the School of Biological, Biomedical & Environmental Sciences (now positioned in the School of Environmental Sciences) at the University of Hull. Students are often unaware of the links between the research of academic staff and the teaching they receive from those same staff (Jenkins et al 1998, Brew 2006): this module provides that link. A description and initial evaluation of the module is published (Morrell 2014), and thus the module is described only briefly here.

Eight research seminars (approximately 45 minutes long) are each followed by a student-led discussion with the speaker. While content and speakers varied somewhat from year to year, many of the same speakers featured across all the 6 years considered here. Students are provided in advance with two research papers relevant to the seminar, and the seminar/discussion allows them to clarify their understanding (a flipped learning approach) and explore the topic in depth. Acquisition of factual knowledge about our research (although important) is not core to the module ethos, there is no systematic building of subject knowledge as the module progresses, and the order in which the research seminars occurred differed between years. Instead, the emphasis is on the development of key communication skills, particularly scientific writing, and on the acquisition of assessment literacy (Smith et al 2011). For each seminar, students write up one of the papers as a 500-word 'news & views' article (short scientific reports

found in top journals), as an authentic science communication task (Higher Education Academy 2012).

Students receive feedback on their first submitted news & views report (within one week of submission to ensure effective use in the next, hereafter 'first report'). I then provide feedback on only two of each student's subsequent seven submissions ('randomly selected reports' or 'random' reports 1 and 2). Crucially, all marked reports are available to all students via the VLE (anonymously). So, in a class of 30, students see around 60 examples of feedback, rather than only their own, and can use the feedback on those reports to develop their own work. As such, the module gives students access to a range of alternative approaches for achieving similar marks (Sadler 2010). Feedback therefore focused on identifying both strengths and deficiencies in the work and making suggestions for how future work could be improved, enabling students to use it as feed-forward (Scott et al 2011). Each report was accompanied by a completed marking rubric and a percentage grade. A semi-categorical mark scheme, which awards marks ending in 2, 5 and 8 for grades between 42% and 75% was used throughout, with the occasional award of 60% where the mark scheme indicated 58% and 62% were equally appropriate. General class feedback was also provided at two points in the module (after feedback on the first report was returned, and after all students had received feedback on their second randomly selected report), and students are provided with exemplars of differing quality (during a specific feedback session). Students are therefore provided with feedback on all reports from multiple sources (which can enhance performance on iterated or multiple-stage assignments, Carless et al 2011).

Students must submit 7 out of the 8 reports at passing grade to pass the module (see Morrell 2014), unless approved mitigating circumstances result in fewer

submissions. Students submitting a piece of work that would not meet the pass mark are contacted individually for additional feedback, and no student submitted more than one piece of work in this category. At the end of the module, students self-assess (Dochy et al 1999; Orsmond 2011) their submissions, and select the two that they anticipate gaining the highest marks for summative assessment ('chosen' reports). Critically, as students do not receive tutor feedback on all their submissions, effective self-assessment (Dochy et al 1999; Orsmond 2011) allows them to complete the module with a summative mark that exceeds the aggregate mark that they received on the formative assessments marked by the tutor. All marking was carried out by a single marker and independently moderated by a 2nd member of staff, in accordance with standard procedures at the institution. No changes to marks were made as a result of moderation.

As part of the module, students write a reflective piece on their choice of assignments for summative assessment. Selected quotations from these reflections are used anonymously in the discussion to add a student perspective, but no systematic qualitative analysis has been carried out. Ethical approval for the analysis of student marks and use of quotations from written reflections was obtained from the Faculty of Science and Engineering Ethical Committee (approval code: FEC_25_2016). All students whose work forms part of this report have now graduated from their undergraduate degree at the University.

Aims and objectives

Here, I evaluate student attainment on *Topics in Biodiversity and Evolution* (described above) across the 8 similar, iterated assessment tasks. I ask the following research questions:

(1) *Does the iterated nature of the assessment allow students to increase their marks over the duration of the module?* The module structure requires students to engage with the same assessment task (a short written report) each week, but the topic of the report differs. A positive answer to this question would indicate that students are able to engage with the feedback, understand the requirements of the assessment, learn about tutor expectations and apply that knowledge to subsequent assignments, where the topic is different but the assignment task is the same.

(2) *What is the value of choice?* Not all student submissions are assessed at the point of submission (practise has value in itself; Hounsell 2007, Boud & Molloy 2013). At the end of the module, students self-assess (Nicol and Macfarlane-Dick 2006), using the knowledge they have gained through the module) and select their two best for summative submission. If there is value is self assessment and choice, then their chosen reports should receive higher marks, as previously shown by Morrell (2014)

(3) *Who benefits from feedback and choice?* Non-high achieving students are known to use tutor feedback in qualitatively different ways to high-achieving students, focusing on superficial deficiencies rather than more significant issues (Scott 2017). Thus, we might predict that high-achieving students benefit most from the module structure, as they are better able to adopt the changes that lead to substantial increases in marks. Alternatively, the opportunity to apply the feedback and choose stronger reports may benefit non-high achievers.

Analysis of news & views marks

I used the marks for the news & views reports from 6 academic years (spanning 2011-2016; N=887 marks from 180 students) to evaluate 1) the effectiveness of the assessment strategy in allowing students to improve their marks 2) the value of allowing students to self-assess and choose their two best reports for summative assessment and 3) whether the feedback approach particularly benefits any particular achievement group (based on marks awarded for the first report as a baseline). All analyses were carried out in R v 3.3.2 (R Core Team 2016). Initial analysis showed no effect of cohort on marks, and so marks are pooled across the 6 years.

Do marks improve over time?

To assess whether marks improve over the course of the module, I used a linear mixed effects model with assignment number (1 to 8) as a continuous variable, accounting for student identity as a random effect to control for the repeated measures nature of the data (multiple marks per student). I also assessed whether the marks awarded differed depending on the number of pieces of individual feedback each student had received (none before their first report, 1 before the submission of their second randomly selected report, 2 before the submission of their third randomly selected report and 3 before the submission of their final chosen reports), again using mixed effects models with student identity as a random effect. Pairwise comparisons were achieved by re-levelling the data. Assumptions of normality were confirmed by visual inspection of plots of residuals and QQ plots (Crawley 2007).

What is the value of choice?

To assess the effect of number of previously unmarked reports that were chosen for final summative assessment (0, 1 or 2) on the increase in marks between marked and

chosen reports, I first calculated the difference in marks between the three marked reports and the two final chosen reports ("mark difference"). I carried out an ANOVA on log-transformed "mark difference +10", followed by a Tukey HSD test to identify significant pairwise differences. The constant (10) was added to ensure all values were positive prior to transformation (a small number of students, 7/180, selected two reports that did not receive marks higher than their three marked reports). ANOVA on log-transformed mean mark for the two chosen reports was used to assess whether final marks differed depending on number of unmarked reports chosen. The assumptions of normality were confirmed by visual inspection of plots of residuals and QQ plots.

Who benefits from choice?

To assess which students, in terms of their achievement early in the module, benefited the most from being able to choose assignments, I carried out a series of correlations between marks using linear models. 95% confidence intervals were calculated to assess whether significant slopes differed significantly from 1 (the expectation if early marks accurately predict final marks). Correlations were carried out between:

- First mark and mean randomly selected report marks

- First mark and mean final chosen report marks

- First mark and mark difference (+10, log transformed)

Subsequently, the relationship between the mean mark for the chosen reports and the interaction between first mark and number of unmarked reports chosen was examined using linear models. The assumptions of normality for all models was confirmed by visual inspection of plots of residuals and QQ plots.

Results

Do marks improve over time?

Marks improved over the course of the module (figure 1: estimate = 0.873, s.e. = 0.101, t = 8.634, df = 346, p < 0.001; figure 1), rising from a mean of 58.6% (SD=7.3, median=58%) on the first report to 64.0 (SD=7.3, median=65%) on the eighth submission. Marks increased between the first report and each randomly selected report, and again between the final randomly selected report and their chosen reports (figure 1, table 1).

What is the value of choice?

As previously reported (Morrell 2014), students choosing work that had not previously been marked increased their marks by significant more than those that chose only from previously marked work (ANOVA: $F_{2,165}$ = 6.652, p = 0.002; TukeyHSD: zero vs one report: p = 0.001, zero versus two reports: p = 0.025, one versus 2 reports: p=0.9). However, there is no overall difference in final marks between students choosing none, one or two previously marked reports (F = 0.884, df = 2, p = 0.415).

Who benefits from choice?

The mark for the first report was a significant predictor of the mark for both the randomly selected ($F_{1,166}$ = 70.64, r^2 = 0.294, p < 0.001, figure 2a) and chosen ($F_{1,166}$ = 56.75, r^2 = 0.250, p < 0.001, figure 3b) reports. However, the 95% confidence intervals for the slopes of the fit lines did not overlap with 1 (randomly selected: 95% CI = (0.393, 0.635), chosen: 95% CI = (0.376, 0.644)) suggesting that those students performing less well on the first reports were able to improve more over the course of the module than those performing well at the start. Indeed, the mark difference (increase

in marks between the marked 3 and chosen 2 reports) is negatively correlated with the mark received for the first report ($F_{1,166} = 15.45$, $r^2=0.065$, $p < 0.001$; figure 2c), although it should be acknowledged that students with lower marks for the first report have more scope for growth.

Further illustrating the importance of choice, there is a significant interaction between the number of unmarked pieces of work chosen and the mark for the first report on the mark for the chosen reports (table 2, figure 3). For students choosing no unmarked pieces of work, there is a positive correlation with the first mark received (table 2), the slope of which does not differ significantly from 1 (95% CI: 0.503-1.147). However, the slope of the relationship between the first mark and mean chosen mark depends on the number of unmarked pieces of work chosen. For students choosing 2 unmarked pieces of work, the slope is significantly shallower and the 95% CI overlaps with zero (-0.117 – 0.598), indicating that the mark for the first report is no longer a significant predictor of the mark for the chosen report (figure 3). The slope for students choosing one marked and one unmarked piece of work is intermediate between the two, with a CI overlapping with neither zero nor 1 (0.333-0.657).

Discussion

The marks attained by the students increased over the course of the module. A single cycle of feedback was sufficient to improve achievement (marks increased from report 1 to the first randomly selected report for each student). Further cycles of feedback resulted in greater increases in achievement. As all students had access to both their own feedback and that of their peers, it is not possible (without detailed information on student engagement, which was not available) to distinguish whether the increase in marks over time is due to interaction with only their own feedback or also with that of

their peers, or which aspects of the feedback the students used. However, the students' written reflections revealed that they engaged with the module and used the feedback on others' work to improve their own:

> 'The positive effect of feedback that was given over the course of the module cannot be understated as It clearly helped me to produce a better standard of work in comparison to my first submitted report' - 2015/16 student
>
> 'The positive aspect of this, is that I have been able to view other people's work and how other students have interpreted both the subject and format of the assignment. I have also been able to take away and use some of their feedback and hopefully improve my work by implementing their feedback into my assignments.' – 2015/16 student

As previously reported (Morrell 2014), marks for the chosen reports were significantly higher than either the first or randomly selected reports, and those students choosing unmarked work gained a greater increase in marks, but this analysis of a larger data set shows that students increased their marks between their third marked report (random 2) and their final chosen reports, suggesting that giving students the opportunity to self-assess work and select their two best resulted in a further increase in marks.

The greatest uplift in marks was seen in those students performing more weakly at the start of the module, suggesting that the combination of practice and choice is particularly beneficial for these students. Of course, students performing strongly at the start of the module (although only 5/180 students achieved a first class mark (>70%) on their first assignment) have potentially less scope for improvement that those gaining lower marks, but crucially, this finding demonstrates that weaker students can and do engage with the feedback strategy and increase their marks. However, in order to fully benefit, students also need to engage with the self-assessment and choice aspect of the

module, as it is primarily those students choosing one or two previously unmarked reports that drive the uplift.

These results support the suggestion that iterated assessment tasks, together with timely feedback, can enhance student learning (Boud & Molloy 2013; Carless et al 2011). Although the students were provided here with identical tasks (news & views reports), with the same learning outcomes, the scientific content of each task was different, moving away from a model of using feedback to improve a draft version (Barker & Pinard 2014) towards encouraging students to learn from and apply the feedback to new yet similar tasks (Boud & Molloy 2013; Carless et al 2011). The increase is marks in unlikely to be due to an increase in subject knowledge, as the module does not build scientific content systematically from week to week. Indeed, the order in which different seminars were presented varied from year to year, and topics varied in their familiarity to students, partially dependent on their previous module selections.

Instead, students consolidate their knowledge of "what makes a good report", thus better understanding the requirements of the assessment, and how to access the higher marks. They then actively use that knowledge in preparing new reports within the module, developing their skills in scientific writing, and the critical expression of ideas. Ideally, they should also be able to take aspects of that knowledge and apply it to tasks in other modules. Some students highlighted this through comments on end-of-module evaluations:

> 'Throughout this module I have learned to write more concisely while including all the relevant information in order to complete the news and views style articles. This is a skill that has been hugely helpful while completing reports for other modules' – 2014/15 student

'Overall this module has allowed me to gain skills in scientific writing and understanding, over a wide range of topics, which will be transferable to other modules, and later on in life' – 2014/15 student

Although it was not possible to quantitatively assess the impact beyond a single module, there can be real benefits to students if they are able to apply feedback across modules (Boud & Molloy 2013). This ability (and opportunity) is sometimes thought to be lacking, particularly in modularised degree programmes and where feedback comes at the end of a module (Brown et al 1997, Weaver 2006, Boud & Molloy 2013). The findings here highlight the need to give students the opportunity to embed feedback knowledge through direct implementation, using it in a new context, rather than a unidirectional approach where feedback is provided but potentially not engaged with (Orsmond et al 2005, Glover & Brown 2006, Scott et al 2011, Orsmond & Merry 2013).

Assessing student work often involves making a qualitative judgement, which cannot always be easily reduced to a formal marking scheme or rubric that can be applied by others (but see Scott 2017), and which develops through experience of the marking process (Sadler 2010). Some students felt that the module indeed helped them to better understand mark schemes, potentially developing some of this knowledge:
'It has improved my understanding of mark schemes and how to apply it more successfully to my work' -2014/15 student
'This module has taught me the importance of reading the mark scheme' – 2015-16 student

The student role in feedback is often neglected (Hounsell 2007, Carless et al 2011, Boud & Molloy 2013), and Carless et al (2011) recommend that Askew & Lodge's (2000) definition of feedback as a dialogue to support learning is adopted, rather than as a directed monologue (Scott 2017). Through iterated feedback on similar work, with specific reference to a marking scheme, students can monitor the quality of their own work through their increased understanding of the marking process and tutor

expectations (Sadler 1989, Sadler 2010). However, when feedback is directed towards individual students, the opportunity to view different aspects of, or approaches to, quality is missing. Students generally only have access to their own approach to a particular piece of assessment, and so cannot access multiple examples across the range of marks, and multiple examples of work of the same level, but taking different approaches. Here, where feedback is directed at the student group more widely, there is the potential for this type of learning to occur.

For tutors, the ability to make a quality judgement develops through experience of the range of overall quality in the set of submissions, and the comparability of submissions of similar quality that differ in their execution (Sadler 2010). In this module, the provision of freely-available, marked work by peers, together with the feedback on those pieces of work, should allow students to access a range of different approaches and begin, perhaps, to recognise both the role of judgement in awarding marks (Sadler 2010, Scott 2017), and that there are different ways to produce good quality work (a "rich experiential assessment space"; Sadler 1989; Sadler 2010). This could be particularly beneficial to those who struggle to improve their own work following individual feedback (Sadler 2010, Pryor & Crossouard 2008). The approach is designed to allow students to recognise that there are different aspects to the quality of a piece of work, that is, that there are different ways to do good work, as highlighted by one student:

> 'It was also extremely beneficial to be able to access fellow students reports and
> feedbacks, as I never received a first in any of my reports it was crucial for me to
> be able to see what I needed to produce in order to hit those higher grades.' -
> 2014/15 student

Sadler (2010) points out that training students to recognise these different levels of quality is necessary to allow them to monitor the quality of their own work during the

writing process (Sadler 1989), but it is also critical in the self-assessment of their work on completion, and the development of life-long learning skills (Nicol and Macfarlane-Dick 2006, Kirby and Downs 2007, Mok et al 2006). Further research is needed into the extent to which students are engaging with feedback that is not their own, and how they are using it.

Here, I show that while overall students perform better following iterated feedback and self-assessment, students performing more weakly at the start of the module make good use of the opportunity to improve their marks, often by one or two categories of 10 percentage points (equating to 1-2 degree classifications in the UK). This is a significant margin given the necessity of applying the feedback to a new task on a different topic. Some students, however, do not improve their marks to such an extent, and this may reflect a variety of factors. Some students lack confidence in self-assessment and select from amongst previously marked reports (Morrell 2014), resulting in a small increase, and reflecting a wider lack of confidence in self-assessment abilities among students (Scott 2017). Other students may struggle to apply the feedback from both their own and other students' work to a new, albeit silimar, task (Morrell 2014), as this is something that they lack experience of, given the extensive use of summative work and modularisation in higher education (Brwon et al 1997, Weaver 2006, Price et al 2010, Boud & Molloy 2013). Others may not engage fully with the module structure and not make the expected use of the feedback they have access to.

In summary, iterated assessment tasks appear to be valuable approaches to improving student attainment, providing students engage with the task and make use of the wide range of feedback provided on different approaches. Both practice and choice result in an increase in marks: the improvement could be greater if all students received

personal feedback on all pieces of work, or if the choice of summative tasks was made by the tutor. This would carry an approximately three-fold time cost to the tutor in this example, but overcome some of the disquiet previously expressed by students on the module regarding not having all their own work marked (Morrell 2014). Complete tutor marking would reduce the development of conscious self-assessment, which would become a much less critical part of the module. Embedding self-assessment within modules, either through a feedback dialogue (Askew & Lodge 2000, Nicol & Macfarlane-Dick 2006; Carless et al 2011; Boud & Molloy 2013) or as an assessment in itself (as here) should encourage students to engage with the process and could act to promote self-assessment abilities. The approach is scalable beyond the small class sizes here: multiple tutors could contribute to the marking and feedback, if combined with clear expectations (marking criteria) and between-tutor moderation of marks and feedback during the module. The approach is not restricted to particular subjects, and is applicable across any context where particular writing skills are required, or where a particular approach takes practice to acquire, so long as the selected format for the assessment was appropriate to the norms of the discipline.

References

Askew, S., & Lodge C. 2000. "Gifts, ping-pong and loops – Linking feedback and learning". In *Feedback for Learning* edited by S. Askew*, 1-17, London: Routledge.

Barker, M. and Pinard, M. 2014. "Closing the feedback loop? Iterative feedback between tutor and student in coursework assessments". *Assessment and Evaluation in Higher Education* 39: 899-915

Black, P. and Wiliam, D. 1998. "Assessment and classroom learning". *Assessment in Education: Principles, Policy & Practice* 5: 7–74.

Bloxham, S. 2012. "'You can see the quality in front of your eyes': grounding academic standards between rationality and interpretation". *Quality in Higher Education* 18 (2): 185-204.

Boud, D. 2000. "Sustainable assessment: Rethinking assessment for the learning society". *Studies in Continuing Education* 22(2): 151–167

Boud, D. & Falchikov, N. 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. & Molloy, E. 2013. "Re-thinking models of feedback for learning: the challenge of design". *Assessment and Evaluation in Higher Education* 38(6): 698-712.

Brennan, M.J. 1995. "Essay writing in nursing: alerting students and teachers to the educational benefits". *Nurse Education Today* 1: 351–56.

Brew, A. 2006. *Research and Teaching: Beyond the divide*. Basingstoke: Palgrave Macmillan.

Brown, G., Bull, J. and Pendelbury, M. 1997. *Assessing Student Learning in Higher Education*. London: Routledge.

Brown, J. 2007. "Feedback: the student perspective". *Research in Post-Compulsory Education*. 12(1), 33-51

Carless, D. 2006. "Differing perceptions in the feedback process". *Studies in Higher Education* 31(2): 219-33.

Carless D., Salter, D., Yang, M. & Lam J. 2011. "Developing sustainable feedback practices". *Studies in Higher Education* 36(4): 395-407.

Covic, T. & Jones, M. K. 2008. "Is the essay resubmission option a formative or a summative assessment and does it matter as long as the grades improve?"*Assessment & Evaluation in Higher Education* 33: 75-85.

Crawley, M.J. 2007. *The R Book*. Chichester: John Wiley & Sons.

Crisp, B.R. 2007. "Is it worth the effort? How feedback influences students' subsequent submission of assessable work". *Assessment & Evaluation in Higher Education* 32: 571–581.

Dann, R. 2014. "Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice". *Assessment in Education: Principles, Policy & Practice*, 21(2): 149-166

Dochy F., Segers, M., & Sluijsmans, D. 1999. "The use of self-, peer and co-assessment in higher education: A review". *Studies in Higher Education* 24(3): 331-350.

Evans, C. 2013. "Making sense of assessment feedback in higher education". *Review of Educational Research* 83 (1): 70-120.

Gibbs, G. and Simpson, C. 2004. "Conditions under which assessment supports students' learning". *Learning and Teaching in Higher Education* 1: 3–31.

Glover, C. and Brown, E. 2006. "Written feedback for students: Too much, too detailed or too incomprehensible to be effective?" *Bioscience Education* 7 doi:10.3108/ beej.2006.07000004.

Hattie, J., Biggs, J. and Purdie, N. 1996. "Effects of learning skills interventions on student learning: A meta-analysis". *Review of Educational Research* 66: 99–136.

Hattie, J., & Timperley, H. 2007. "The power of feedback". *Review of Educational Research* 77(1): 81-112

Higher Education Academy 2012. *A Marked Improvement: Transforming assessment in higher education*. York: Higher Education Academy.

Hounsell, D. 2007. "Towards more sustainable feedback to students" in *Rethinking Assessment in Higher Education. Learning for the Longer Term* edited by Boud, D. and Falchikov, N, 101-113 London: Routledge

Jenkins, A., Blackman, T., Lindsay, R. and Paton-Saltzberg, R. 1998. "Teaching and research: Student perspectives and policy implications". *Studies in Higher Education* 23: 127–141.

Kirby, N.F. & Downs, C.T. 2007. "Self-assessment and the disadvantaged student: potential for encouraging self-regulated learning?" *Assessment & Evaluation in Higher Education*, 32 (4): 475-494

Lawson, R., Taylor, T., Thompson, D. G., Simpson, L., Freeman, M., Treleaven, L., & Rohde, F. 2012. "Engaging with graduate attributes through encouraging accurate student self-assessment". *Asian Social Science*, 8(4): 3–12.

Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). "Accuracy of students' self-assessment and their beliefs about its utility". *Assessment & Evaluation in Higher Education*, 35(2): 135–156.

Mok, M. M. C., Lung, C. L., Cheng, D. P. W., Cheung, R. H. P. and Ng, M. L. 2006. "Self- assessment in higher education: experience in using a metacognitive approach in five case studies", *Assessment & Evaluation in Higher Education*, Vol. 31(4): 415-433

Morrell, L.J. (2014). "Use of feed-forward mechanisms in a novel research-led module". *Bioscience Education* 22(1)*:* 70-81 doi:10.11120/beej.2013.00020.

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): 199-281.

Orsmond, P. & Merry, S. 2013. "The importance of self-assessment in students' use of tutors' feedback: a qualitative study of high and non-high achieving biology undergraduates". *Assessment and Evaluation in Higher Education* 38(6): 737-753.

Orsmond, P. 2011. *Self- and Peer-assessment: Guidance on Practice in the Biosciences*. Leeds: UK Centre for Bioscience, Higher Education Academy.

Orsmond, P., Merry, S. and Reiling, K. 2005. "Biology students' utilization of tutors' formative feedback: A qualitative interview study". *Assessment & Evaluation in Higher Education* 30: 369–386.

Price, M., Handley, K., Millar, J. & O'Donovan, B. 2010. "Feedback: all that effort, but what is the effect?" *Assessment and Evaluation in Higher Education*, 35(3): 277-289.

Pryor, J. & Crossouard, B. 2008. A socio-cultural theorisation of formative assessment, *Oxford Review of Education*, 34(1): 1-20

Quality Assurance Agency (QAA) 2006. *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education. Section 6: Assessment of Students*. Gloucester: QAA.

R Core Team 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Sadler, D.R. 1989. "Formative assessment and the design of instructional systems. Instructional Science" 18: 119–144.

Sadler, D.R. 2010. Beyond feedback: developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education* 35: 535-550

Sadler, D. R. 2013. "Opening up feedback: Teaching learners to see". In *Reconceptualising Feedback in Higher Education: developing dialogue with*

students edited by Merry, S., Price, M., Carless, D., & Taras, M., 54-63. London: Routledge

Scott, G.W. 2017. "Active engagement with assessment and feedback can improve group-work outcomes and boost student confidence". *Higher Education Pedagogies* 2 (1): 1-13

Scott, J., Shields, C., Gardner, J. and Hancock, A. 2011. "Student engagement with feedback". *Bioscience Education* 18 doi:10.3108/beej.18.5SE

Sebester, A.J. & Bray Speth, E. 2017. "How should I study for the exam? Self-regulated learning strategies and achievement in introductory biology". *CBE Life Science Education* 16(2): ar30, doi: 10.1187/cbe.16-09-0269.

Smith, C.D., Worsfold, K., Davies, L., Fisher, R. & McPhail, R. 2011. "Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy'". *Assessment and Evaluation in Higher Education* 38(1): 44-60.

Storch, N. and J. Tapper. 2002. "A useful kind of interaction? Evaluations by university students of feed-back on written assignments". *Australian Review of Applied Linguistics* 25: 147–67

Weaver, M.R. 2006. "Do students value feedback? Student perceptions of tutors' written responses". *Assessment & Evaluation in Higher Education* 31: 379–394.

Table 1: Analysis of marks for the first, two randomly selected and final chosen reports. Pairwise comparisons were achieved by re-levelling the data in the model. Significant p-values are highlighted in bold.

		Value	SE	DF	t-value	P
(Intercept)		58.590	0.577			
First	Random 1	3.493	0.510	704	6.846	**<0.001**
First	Random 2	5.445	0.520	704	10.466	**<0.001**
First	Chosen	8.243	0.442	704	18.633	**<0.001**
Random 1	Random 2	1.952	0.518	704	3.764	**<0.001**
Random 1	Chosen	4.750	0.440	704	10.786	**<0.001**
Random 2	Chosen	2.798	0.452	704	6.193	**<0.001**

Table 2: Results summary table of the effect of the mark for the first report on the mean mark for the two chosen reports, accounting for the number of unmarked reports chosen. The intercept represents the case where both chosen reports had previously received mark. The comparison of 1 v 2 unmarked reports was achieved by re-levelling the data in the model. Significant p-values are highlighted in bold.

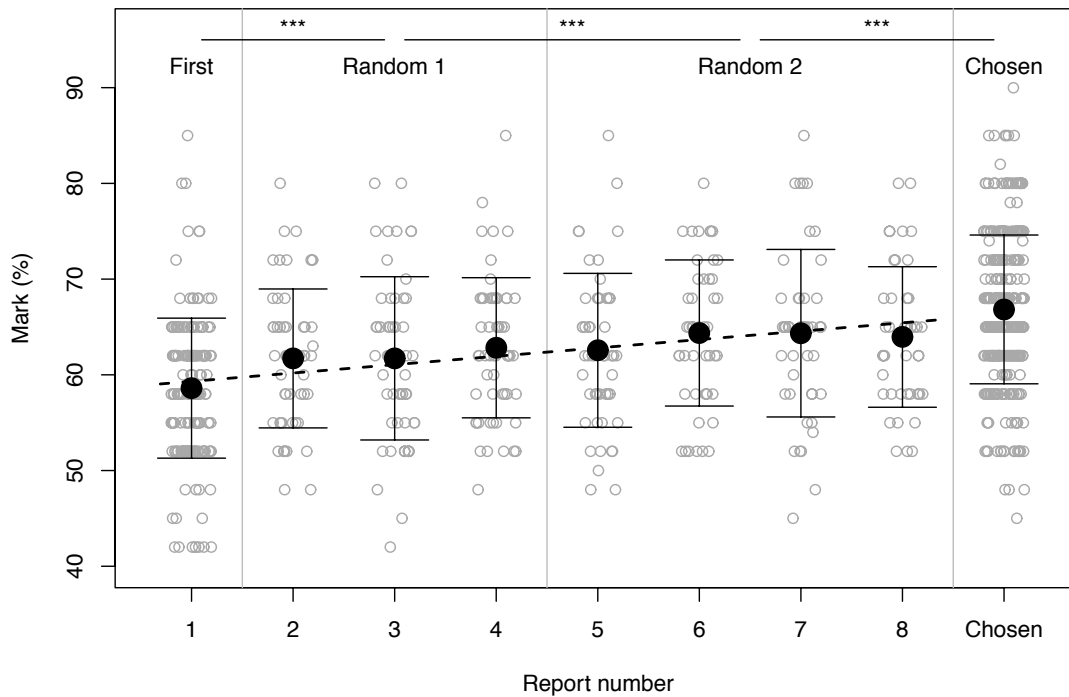|	Estimate	SE	t	p
(Intercept)	17.492	9.990		
First report	0.825	0.163	5.055	**<0.001**
Unmarked: 0 v 1	20.936	11.071	1.891	0.060
Unmarked: 0 v 2	33.818	14.530	2.328	**0.021**
Unmarked: 1 v 2	12.882	11.577	1.113	0.268
1st * Unmarked: 0 v 1	-0.330	0.183	-1.806	0.073
1st * Unmarked: 0 v 2	-0.584	0.244	-2.395	**0.018**
1st * Unmarked:1 v 2	-0.254	0.199	-1.278	0.203

Figure 1: The mean ± SD (large filled circles and error bars) percentage mark gained for each report over the course of the module, with model prediction for the 8 submissions (dashed line). Marks for the chosen reports are also included, and the plot is subset into the 4 categories of report (first, randomly selected and chosen reports). Asterisks indicate significant differences between these subsets (*** = $p<0.001$).
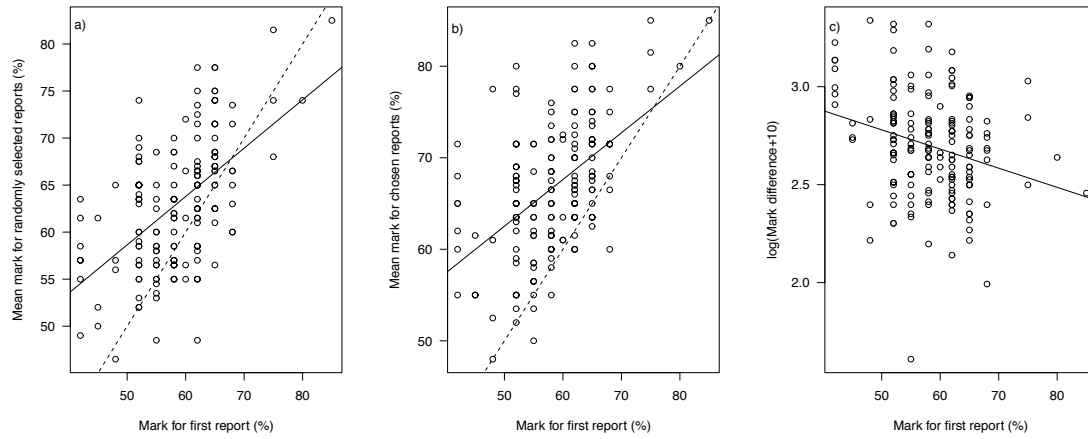
Figure 2. Relationship between **a)** the mark awarded for the first and randomly selected reports; **b)** between the first and chosen reports; **c)** between first report and the difference in marks between the 3 marked reports and the two chosen reports (mark difference). The solid line represents the model prediction (best fit line through the data) while the dashed line indicates a 1:1 relationship between the marks.
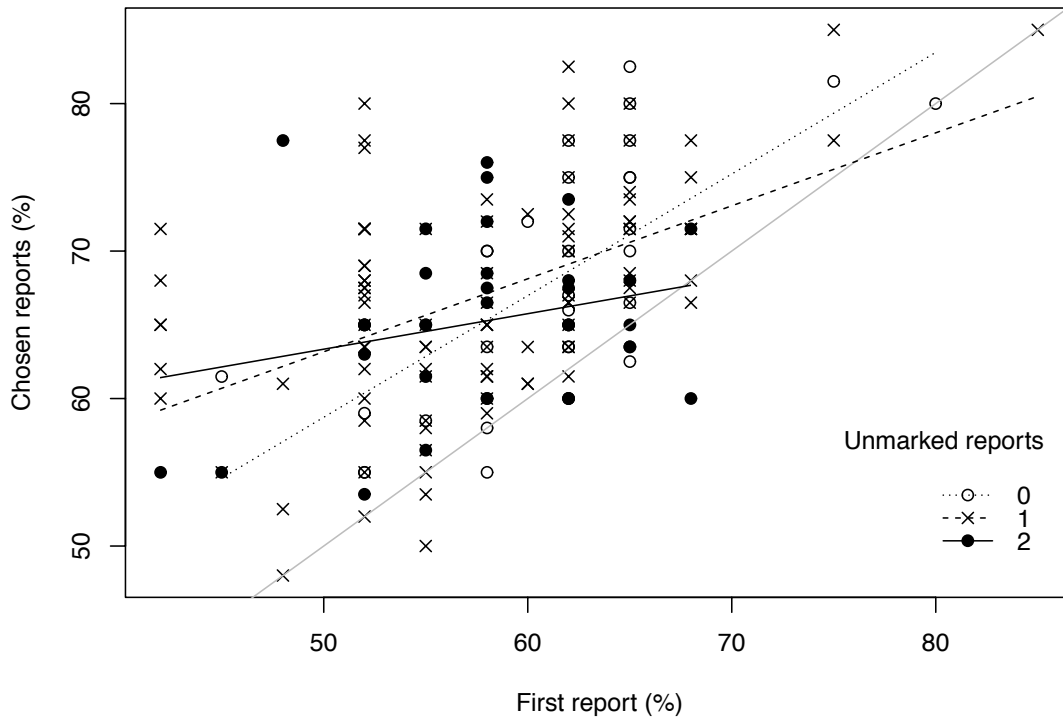
Figure 3: The relationship between the mark for the first report and the mean mark for the two chosen reports, for students choosing zero (open circles, dotted line), one (crosses, dashed line) or two (filled circles, solid line) previously unmarked reports. The solid grey line in a) represents a 1:1 relationship, and in b) represents no increase (i.e a mark difference of zero).