*Assessment & Evaluation in Higher Education*, 2016
http://dx.doi.org/10.1080/02602938.2016.1174187

Routledge
Taylor & Francis Group

impact on subsequent

# Are they using my feedback? The extent of students' feedback use has a large impact subsequent academic performance

Kirsten Zimbardi[a]*, Kay Colthorpe[a], Andrew Dekker[b], Craig Engstrom[c], Andrea Bugarcic[d], Peter Worthy[b], Ruban Victor[a], Prasad Chunduri[a], Lesley Lluka[a] and Phil Long[e]

*[a]School of Biomedical Science, University of Queensland, St Lucia, Australia; [b]School of IT & Electrical Engineering, University of Queensland, St Lucia, Australia; [c]School of Human Movement Studies, University of Queensland, St Lucia, Australia; [d]Endeavour College of Natural Health, Gold Coast, Australia; [e]Learning Sciences, University of Texas at Austin, Austin, TX, USA*

Feedback is known to have a large influence on student learning gains, and the emergence of online tools has greatly enhanced the opportunity for delivering timely, expressive, digital feedback and for investigating its learning impacts. However, to date there have been no large quantitative investigations of the feedback provided by large teams of markers, feedback use by large cohorts of students, nor its impact on students' academic performance across successive assessment tasks. We have developed an innovative online system to collect large-scale data on digital feedback provision and use. Our markers ($n = 38$) used both audio and typed feedback modalities extensively, providing $388 \pm 4$ and $1126 \pm 37$ words per report for first- and second-year students, respectively. Furthermore, 92% of first year and 85% of second-year students accessed their feedback, with 58% accessing their feedback for over an hour. Lastly, the amount of time students spent interacting with feedback is significantly related to the rate of improvement in subsequent assessment tasks. This study challenges assertions that many students do not collect, or use, their feedback. More importantly, we offer novel insights into the relationships between feedback provision, feedback use and successful academic outcomes.

**Keywords:** feedback; assessment; academic performance

## Introduction

Feedback on student performance is viewed as one of the most influential and effective learning paradigms (Hattie and Timperley 2007; Hepplestone et al. 2011; Jonsson 2012; West and Turner 2015). Assessment reforms in higher education recognise the need for feedback that is timely, informative and encourages positive attitudes towards future learning amongst students (Boud and Falchikov 2006; Hepplestone et al. 2011; Brearley and Cullen 2012; Knauf 2015), and there are many theoretical guides on how to improve feedback provision (e.g. Sadler 1998; Nicol and Macfarlane-Dick 2006; Hattie and Timperley 2007). It is extremely concerning, however, that a meta-analysis of the effects of feedback reported in the last 100 years found that, for 38% of the studies analysed, feedback reduced performance compared with no-feedback controls (Kluger and DeNisi 1996). Spurred by this paradox, Hattie

*Corresponding author. Email: k.zimbardi@uq.edu.au

(2008) has dug deeper into the evidence and shown that it is not the feedback that educators *provide*, but rather the feedback that educators *receive* on student progress towards learning goals, that underlies the largest improvements in student learning. Thus, leaders in higher education assessment have argued that the prevalent misuse of well-intentioned feedback would be best addressed by academics shifting their focus away from how feedback is provided, towards understanding how students use feedback to improve performance in related assessment tasks, and using this evidence to drive changes in feedback provision (Boud and Molloy 2013). Such calls for reform in higher education feedback practices raise the question of what we know about student use of feedback.

Unfortunately, how students actually use feedback, and the relationships between student use of feedback and changes in assessment performance, are poorly understood (for review see Jonsson 2012). Studies commonly report that students do not collect, or read, written formative feedback (MacDonald 1991; Sinclair and Cleland 2007). Further, it appears that, when students do attempt to use feedback, they often lack the necessary strategies to do so effectively (Furnborough and Truman 2009). A recent, small-scale but detailed study, demonstrated very clearly and poignantly that students get frustrated and ignore feedback when they do not understand what to do with it (Still and Koerber 2010). On a brighter note, the increasingly widespread adoption of online marking and feedback tools makes it easier for students to collect their feedback (Hepplestone et al. 2011). When surveyed, the majority of students report accessing their feedback and re-listening to audio comments multiple times (Brearley and Cullen 2012; Carruthers et al. 2014). With the explosion of available evidence on assessment analytics (Ellis 2013), it is now possible to undertake large-scale, systematic and objective investigations of the extent to which student use the multi-modal feedback provided on electronic assessment items, and begin to understand how feedback use impacts on students' subsequent academic performance.

In the context of multi-modal digital feedback on electronic assessment items, students report that the provision of audio feedback increases their engagement with, and the usefulness of, feedback (Ice et al. 2007; Lunt and Curran 2010; Brearley and Cullen 2012; Rhind et al. 2013). Furthermore, students report that a combination of typed and audio feedback is more useful than providing either typed or audio feedback alone (Still 2006). There are now many studies reporting that the use of audio feedback, or mixed audio and typed feedback modalities, improves the efficiency and timely delivery of detailed feedback (Merry and Orsmond 2008; Lunt and Curran 2010; Gould and Day 2013). However, most of these studies report on the use of audio feedback by a small group of enthusiastic academics (often the author(s)) and none have investigated the extent of uptake of mixed-modality feedback by large teams of casual academic markers or teaching assistants.

In the context of early stage, undergraduate biomedical science courses, cohorts of 800–1000 students are common, and learning to write in the scientific genre is essential for students to gain skills in science communication and a deep understanding of the experimental nature of scientific knowledge (Colthorpe, Rowland, and Leach 2013). Across many contexts, escalating cohort sizes continue to increase the divide between academics who coordinate such courses, the feedback being provided to students by markers, and the evidence of the impact the feedback has on student performance across assessment tasks. First, such academics need access to evidence of the extent to which large teams of casual academic markers are providing feedback, and second, to the extent to which marking teams engage with new feedback tools and modalities (such as in situ audio annotations). Third, academics need evidence of the extent of feedback use by students, and, last but most importantly, academics need evidence of the impact of feedback provision and feedback use on how students improve from one assessment task to the next.

This paper describes the use of a novel feedback system which allows markers to provide students with detailed, specifically targeted and personalised written, typed or audio feedback in a timely and efficient fashion. Most importantly, this system facilitates the tracking of both the provision of feedback by markers and student interactions with that feedback. In this study, this comprehensive feedback analytics capture system has been used to investigate: (1) the extent to which large teams of markers use audio and typed feedback modalities, when marking laboratory reports across two iterations of two courses; (2) the extent to which students interacted with that feedback; and (3) the associations between feedback use and student performance on subsequent assessment tasks.

## Methods

### *Educational context*

This case study was conducted at a large, research-intensive Australian university, and was approved by the institutional ethical review board. The two, semester-length (13 week) courses described in this study are the first- and second-level courses required for students who wish to major in biomedical science (specialising in physiology, pharmacology, neuroscience or molecular and cellular biology), and are offered in two semesters per year. These level 1 and 2 biomedical science courses have been described in detail previously (Zimbardi, Bugarcic, et al. 2013; Good et al. 2015). Briefly, during the practical classes, students undertake hands-on laboratory experiments, and are given increasing levels of autonomy in the experimental design as they progress through the two courses (Zimbardi, Bugarcic, et al. 2013). Students complete a report at the end of each practical/module. Each course has a designated practical coordinator (faculty member) who oversees the laboratory classes, is responsible for training of teams of teaching assistants, and moderation of the marking and feedback that teaching assistants provide. Assessment in the first level course offerings differ slightly between semesters, as the semester 1 course commences with a task that is identical to the others in format, is graded and has feedback provided, but is formative only. Assessment in the second level course offerings are identical each semester.

### *Assessment items analysed in this study*

Overall, the level 1 reports induct students into the conventions used for a scientific report, and the level 2 reports extend students to strive for the level of publishable scientific article. The task requirements and marking standards (judged using criteria rubrics) are consistent within each course, with only the topic changing for each subsequent report. Thus within each course, these assessment items have ‘overlapping features’ (Boud and Molloy 2013, 20), where the key learning outcomes in scientific writing are kept consistent so that feedback from one report is directly applicable to subsequent reports.

In the level 1 course, students are guided through the practicals by an interactive, online laboratory manual built using LabTutorTM software (ADInstruments, New

Zealand). This system is also used to scaffold students through the basic structure of the scientific article genre. At the end of each practical class, students export a document from the LabTutor system containing answers and the data their group has collected during the class. Students then work individually on the discussion section and submit an individual report online 48 h after each class. The reports are assessed by the students' teaching assistant who provides feedback on the report at least five days before the next practical class.

In the level 2 course, students complete two practical modules, each consisting of three classes in which they develop new technical skills, plan an experiment, and then conduct that experiment, collect and analyse the data. One week after the final class of the module, students electronically submit an individual report, written to the conventions of a scientific journal article. As in the level 1 course, the students' teaching assistant marks and provides feedback on the report. Students receive this feedback prior to the second class of the second module, in which they are also provided with generalised verbal feedback. In addition, students have opportunities to discuss the feedback with their teaching assistant in the final class of the module, before their second report is due, again one week after the completion of the module.

### *Student cohort*

During the period of this study, there were 733 and 972 students enrolled in the level 1 course, and 251 and 92 students enrolled in the level 2 course in 2013 Semesters 1 and 2, respectively. There were slightly more female (55%) than male (45%) students, the vast majority (90%) were 17–21 years old, domestic students (91%), with high- to mid-range university entry scores (84%). Just over half (53%) of the students taking these courses were enrolled in a three-year Bachelor of Science (B.Sc.) or four-year dual degree combining B.Sc. with another degree, or the Bachelor of Biomedical Science (four-year research-focused programme). An additional 21% were enrolled in one of four sports science degrees, and 12% were enrolled in the combined B.Sc./Medical programme (in which students undertake an accelerated two-year B.Sc. programme and then enter a four-year, graduate entry medical programme). The remaining students were enrolled in a large range of single and dual degrees (e.g. arts, engineering), specialty degrees and applied science degrees.

### *Feedback analytics capture system*

The feedback analytics capture system is a 'rich media' marking system, which aims to help markers provide students with timely, detailed and situated feedback, with learning analytics integrated into the feedback process. A full walkthrough of the system is available online (Zimbardi, Colthorpe, et al. 2013). Briefly, the system consists of three components: (1) administration interface that processes submissions for marking or moderation, (2) iPad marking application that allows markers and moderators to provide in situ feedback in the form of audio, typed and handwritten annotations, and to mark student work using course-specific criteria and standards rubrics, and (3) feedback viewer which logs the interactions of the students with the feedback document and annotations.

### *Administration interface*

The administration interface is a web-based application for course coordinators that assists in assigning submissions to markers working in large teams, and in moderating the marking and feedback provided by these large teams of markers. This interface is capable of converting a number of document formats for processing, including student work submitted digitally to online repositories and learning management systems (e.g. BlackboardTM, Washington DC, USA). Once submissions are loaded, they are automatically associated with the respective student (or groups of students), and assigned to markers. The marking workflow can be examined at any time by the coordinator, who can also moderate marked assignments at multiple stages of the marking process. Coordinators are presented with a submission workflow screen, which shows the stage at which all submissions are at within the marking process (Figure 1(A)).

### *iPad marking application*

UQMarkup is an iPad application used by markers to provide feedback on student work. This presents markers with all submissions that are assigned to them, and allows them to search and filter to find a specific submission (Figure 1(B)). Once a marker has chosen a submission to provide feedback on, they are presented with a

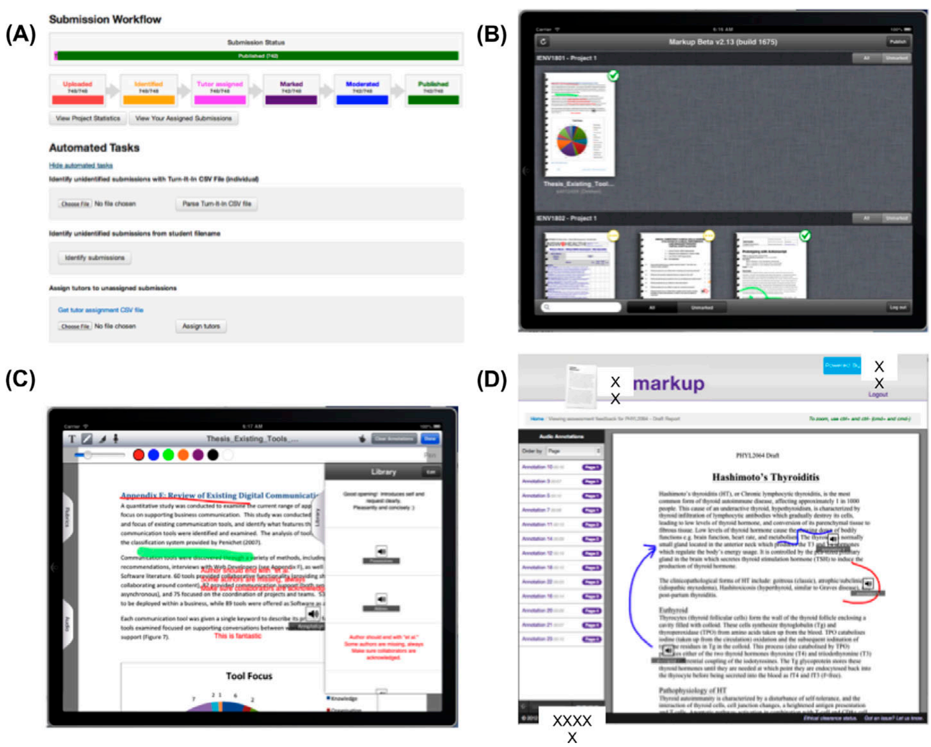


Figure 1. (A) Administration interface; (B) iPad library of submissions available for marking or uploading; (C) iPad feedback view, with examples of each type of annotation embedded; (D) Feedback viewer showing annotated document.

full screen representation of the document (Figure 1(C)). A number of tools are available at the top of the interface, to allow the marker to annotate either by handwritten annotation (freehand or highlighter), typed annotation (via an onscreen keyboard) or situated audio annotation. The application provides three tabs that can be shown or hidden by the user, including a marking criteria rubric, a list of all audio annotations contained within the document and an annotation library.

#### *Feedback viewer*

Once a submission has been marked, moderated by the coordinator if needed, and published by the coordinator, students receive an email notification that feedback is available. The email contains a unique link which allows students to log into the secure system and access a web-based presentation of their submission with embedded annotations (Figure 1(D)). The feedback viewer is designed with two goals in mind: (1) to allow for in situ feedback without requiring additional software for the student, and (2) to capture information about how the student engages with the feedback. All information is time stamped so that information can be gathered on when and how long students viewed and interacted with each page and annotation.

### *Data collection and analysis*

A total of 5960 reports submitted and marked through the feedback analytics system in semester 1 and 2, 2013 were used for this study. The modality, length (duration or number of words) and position of each feedback annotation provided on all reports was recorded. To compare the extent to which markers used the audio and text annotation modalities, we converted the audio annotation duration to an estimated word count. Specifically, a randomly selected sample of 160 audio annotations were transcribed and their word count determined. Talking speed was relatively stable across audio annotations, as they contained on average $164 \pm 6$ words per min, which is also consistent with previous reports of 450–500 words per three-minute audio feedback comment (Brearley and Cullen 2012). Then the length (in minutes) of each audio annotation was multiplied by 164 to allow a direct comparison of the length between typed and audio annotations. Clickstream logs were collected as students accessed their marked documents, recording the clock time of each click as student interacted with the marked document. Students' academic performance on each report was also recorded. Throughout this study, quantitative analyses were performed using R 3.1.1 (R Development Core Team, Auckland, NZ). The results are expressed as mean and standard error of the mean (SEM) or as frequencies as appropriate. Statistical comparisons were made using multiple analysis of variance (MANOVA) with Tukey post-testing and considered significant if the adjusted $p$ value was less than 0.05.

## Results and discussion

The increasing availability of online technologies, which allow the provision of multimodal feedback annotations that can be generalised or situated, has certainly increased the variety and flexibility of feedback delivery options (Hepplestone et al. 2011). However, it has also increased the variability of feedback provision and changed the way in which students interact with feedback, leading to potentially greater variability in student outcomes (Hepplestone et al. 2011; Jonsson 2012).

The primary aims of this study were to understand how the different modalities of feedback available through online technologies affected feedback provision, both across successive student assessment tasks and within large cohorts with multiple markers, and to examine the way students interacted with the feedback provided. Analyses of feedback analytics data from laboratory reports submitted for assessment (Table 1) in two biomedical science courses in level 1 ($n$ = 1705 students) and level 2 ($n$ = 343), in Semesters 1 and 2, 2013, have shown that there are significant differences in the ways in which markers use the different modalities of feedback (Figures 2–4). In addition, the data demonstrates that there are substantial differences in the way students interact with their marked reports and feedback within them across the semesters (Figures 5–8).

### *Feedback provision*

#### *Annotation position*

Markers used the *in situ* feedback feature of the UQMarkUp tool well, with most of the typed and audio annotations placed near the relevant section of student writing, and very few reports receiving 'summary' annotations either on the first (9%) or last (0.01%) page of the report text. Previous studies investigating the impact of audio annotations have highlighted that students view the separation of audio comments from the relevant section of their work as confusing, for example when a single overall audio comment is provided, or when audio and text files are provided separately (Ribchester, France, and Wakefield 2008; Rodway-Dyer, Knight, and Dunne 2010). In this regard, the advent of completely online assessment submission and marking systems, which allow in situ embedding of feedback regardless of modality, represent a notable improvement in feedback provision.

#### *Annotation modality*

Although markers used both the audio and typed annotation modalities, there was a clear difference in the extent to which markers used the two modalities between the year levels. In the first-year reports, markers provided slightly more typed annotations per report than audio annotations ($p < 0.001$; Figures 2(A) and 2(B)). This was reversed for the second years' reports, where markers provided nearly twice as many audio annotations than typed annotations per report ($p < 0.001$). However, on average, audio annotations contained nearly eight times as many words as the typed

Table 1. Number of assignments processed through the feedback analytics capture system used for the analysis in this study.

| Course | Semester | Report 0* | Report 1 | Report 2 | Report 3 |
|---|---|---|---|---|---|
| Level 1 | 1 | 682 | 674 | 671 | 667 |
| | 2 | | 897 | 870 | 854 |
| Level 2 | 1 | | 239 | 241 | |
| | 2 | | 85 | 80 | |
| Total | | 682 | 1895 | 1862 | 1521 |

*Report 0 is a formative task that was only offered in the Level 1, Semester 1 course.

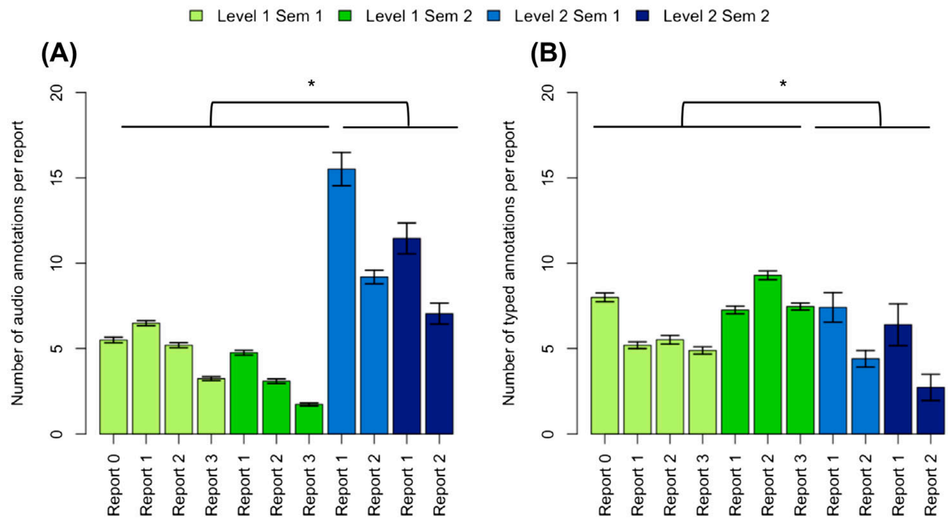


Figure 2. Number of audio and typed annotations for level 1 course ~~(Report 0: $n = 681$, Report 1: $n = 1559$, Report 2: $n = 1535$, Report 3: $n = 1516$~~ and ~~for the~~ level 2 course ~~(Report 1: $n = 299$, Report 2: $n = 295$)~~. Shown is the (A) Mean ± SEM number of audio annotations per report across two semesters of level 1 and 2 subjects; (B) Mean ± SEM number of typed annotations per report across two semesters of level 1 and 2 subjects.
Note: *Indicates a significant difference in the number of annotations between first- and second-year reports ($p < 0.05$).

annotations ($p < 0.001$; Figures 3(A) and 3(B)). Therefore, the total amount of feedback provided in audio annotations, in terms of word length, was significantly greater than typed annotations for all reports ($p < 0.05$; Figures 4(A) and 4(B)).

### *Differences in feedback provision across year levels*

When comparing between year levels, we found that markers provided significantly more audio annotations on second-year reports than on the first-year reports ($p < 0.001$; Figure 2(A)). The audio annotations were also approximately 30% longer on second-year reports than those provided to first-year students ($p < 0.05$; Figure 3(A)). Thus, although markers provided more typed annotations ($p < 0.05$; Figure 2(B)) that were longer ($p < 0.05$; Figure 3(B)) on the first-year reports compared with second-year reports, the second-year students still received nearly three times as much feedback per report as the first-year students overall ($p < 0.001$; Figure 4).

These differences in the extent of feedback provision between the year levels were reproduced across the two semesters, suggesting that course context plays a role in shaping how markers provide feedback. This may be related to the differences in task length between first and second year, with the second-year reports being of greater average length (1909 ± 150 words) than the first-year reports (1269 ± 99 words). Indeed, reports from second-year students also took longer to mark (28.4 ± 1.0 min) than first-year reports (13.9 ± 0.2 min). In addition, marker feedback behaviours are likely to be influenced by task design. In our context, the more structured practicals used in first-year results in a much higher degree of similarity between first-year reports than second-year reports (Zimbardi, Bugarcic, et al. 2013).

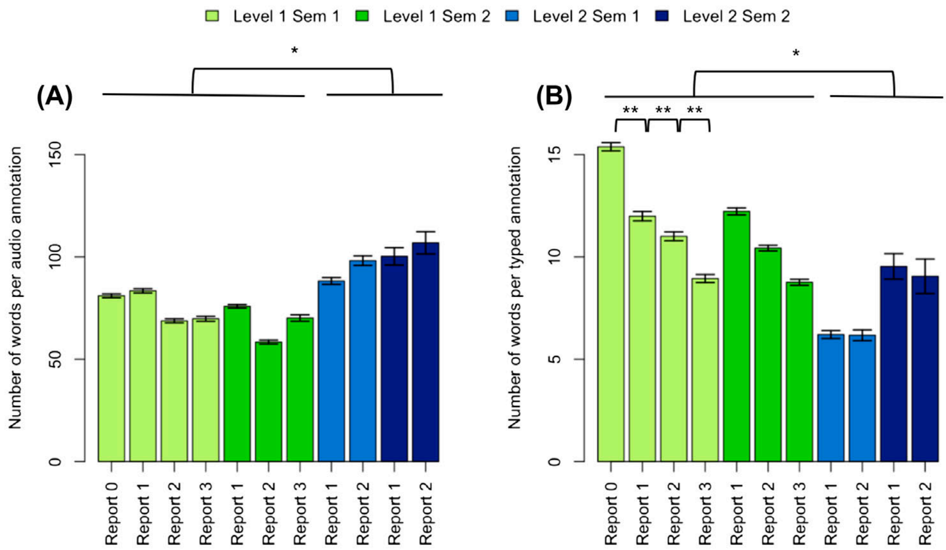


Figure 3. Number of words provided in audio and typed feedback for level 1 course ~~(Report 0: *n* = 681, Report 1: *n* = 1559, Report 2: *n* = 1535, Report 3: *n* = 1516~~ and ~~for the~~ level 2 course ~~(Report 1: *n* = 299, Report 2: *n* = 295)~~. Shown is the (A) Mean ± SEM number of words in audio annotations in reports across two semesters and levels; (B) Mean ± SEM number of words in typed annotations in reports across two semesters and two levels. Notes: *Indicates a significant difference in the length of annotations between first- and second-year reports ($p < 0.05$); **indicates a significant difference in the length of annotations between reports ($p < 0.001$). NB: The *y*-axis scale for A is 10× the *y*-axis scale for B.

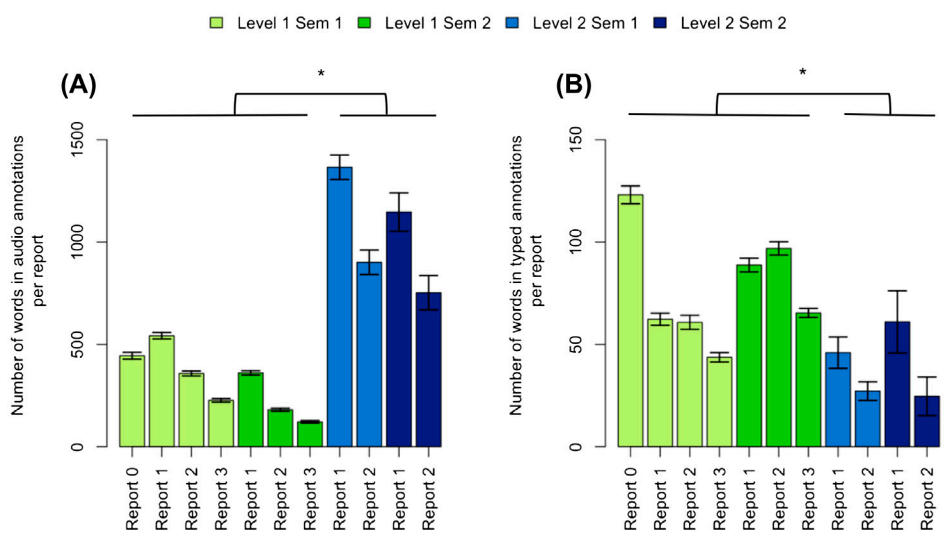


Figure 4. Total amount of feedback provided in audio and typed feedback for level 1 course ~~(Report 0: *n* = 681, Report 1: *n* = 1559, Report 2: *n* = 1535, Report 3: *n* = 1516~~ and ~~for the~~ level 2 course ~~(Report 1: *n* = 299, Report 2: *n* = 295)~~. Shown is the (A) Mean ± SEM number of words in audio annotations in reports across two semesters and levels; (B) Mean ± SEM number of words in typed annotations in reports across two semesters and two levels. Note: *Indicates a significant difference in the total amount of feedback between first- and second-year reports ($p < 0.05$). NB The *y*-axis scale for A is 10× the *y*-axis scale for B.

In addition, there are higher expectations in terms of scientific reasoning in the second year compared with first year.

Is it possible to reattach this line to its parent paragraph on page 8? perhaps by making the deletions suggested in the above figure captions?

*Changes in feedback provision across the progressive assessment tasks*

With the exception of the formative task (Report 0), markers tended to provide more audio annotations on the earliest reports in each year and semester, with the number of audio annotations then declining on subsequent reports across each semester (Figure 2(A)). Interestingly, despite this decline, the average length of each in situ audio annotation was relatively consistent, with averages across semesters and year levels at $28.8 \pm 0.2$ s per annotation. With an average talking speed of 164 words/min, this would suggest that markers tend to give feedback in approximately 80 word 'sound bites', effectively a short paragraph of information. This contrasts with previous studies in which markers provided a single, considerably longer, audio annotation (Ribchester, France, and Wakefield 2008; Gould and Day 2013), but may equate to a similar amount of audio feedback being provided overall on equivalent tasks (Ice et al. 2007; Lunt and Curran 2010).

While audio annotations were relatively consistent in length, the typed annotations declined between successive reports, at least for the first-year course. In semester 1, first-year students received the longest typed annotations on their formative task ($15.4 \pm 0.2$ words), but this dropped by the final report ($8.9 \pm 0.2$ words; $p < 0.001$; Figure 3(B)). Although there was no formative task in semester 2, the length of the typed annotations provided to those first-year students declined at a similar rate between successive reports ($p < 0.001$), finishing at the same minimum by the final report ($8.8 \pm 0.1$ words; Figure 3(B)) seen in semester 1. In contrast, the length of the typed annotations provided on second-year reports remained unchanged across successive reports in each semester ($p > 0.05$; Figure 3(B)), and were similar or even shorter than the typed annotations the first-year students received on their final reports ($p < 0.001$; Figure 3(B)).

Notably, the *number* of typed annotations provided on second-year reports nearly halved from the first to the second report ($p < 0.001$; Figure 2), resulting in an overall decline in the amount of typed feedback provided to second-year students across the successive reports ($p < 0.001$; Figure 4). Taking the total amount of audio and typed feedback provided on each report together (Figure 4), it is clear that markers provide a declining amount of feedback across the successive, linked assessments embedded in these curricula. While this might seem concerning when viewed alone, it is quite likely that the decline in feedback provision is linked to a similar decline in student use of the feedback, and, inversely, to improvements in report quality (indicated by increasing academic performance) across successive reports.

### Feedback use

*Proportion of students viewing report feedback*

The vast majority of the first-year students opened their marked reports (92%), although the proportion that did so tended to decline slightly across the semester, with the final report being opened by fewer students (83%; Figure 5). Surprisingly, the reverse was true in second-year students. In comparison to the first-year students, fewer second-year students opened any of their reports (85%), with only 63%

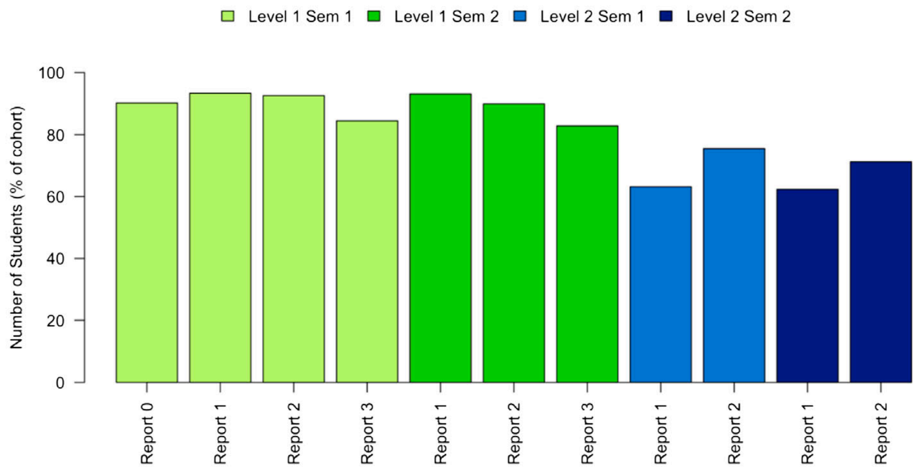


Figure 5. The proportion of students who looked at their feedback (shown as the proportion of students who received feedback in each cohort). Number of students in each cohort: Level one semester 1: $n = 682$; Level one semester 2: $n = 897$; Level two semester 1: $n = 239$; Level two semester 2: $n = 85$.

opening their first report and (a non-overlapping) 74% opening their second report. Notably, this pattern was consistent in each semester (Figure 5).

It is difficult to identify the cause of this difference between first- and second-year students without further investigation. At this stage, it could be hypothesised that this reflects a lower engagement with assessment by second-year students (Loughlin et al. 2013), or that there were differences in the in-class opportunities for students to gain verbal feedback on their first report. It is also possible that, because the first-year course included more subsequent reports and the feedback for the first report was released earlier in semester, first-year students had more prompts and opportunities to access the feedback on their first report than the second-year students. A preliminary analysis of the dates on which students access their feedback (Zimbardi et al. 2014) suggests that the majority of both first and second-year students access the feedback on the first report before the second report is due, but a more thorough analysis, along with additional data on student use of other feedback sources, is needed to convincingly test these hypotheses.

## *Feedback viewing duration and pause times*

In terms of the duration for which student had reports open, in each semester and year level the duration declined markedly from averages ranging between 3–7 h for the non-final reports, down to 30–90 min for the final report (Figure 6). When the patterns of open duration across the cohorts were examined in more detail (Figure 7), it was apparent that subsets of students interacted with their marked reports for distinct periods of time. For the non-final reports in each semester (Figures 7(A) and 7(C)), the majority of students were divided approximately equally between those who opened their report for between 1 min and an 1 h (41%), and those who opened them for greater than an hour (43%), with a smaller subset who opened their reports for less than 1 min (6%) or did not open their report at all (10%).

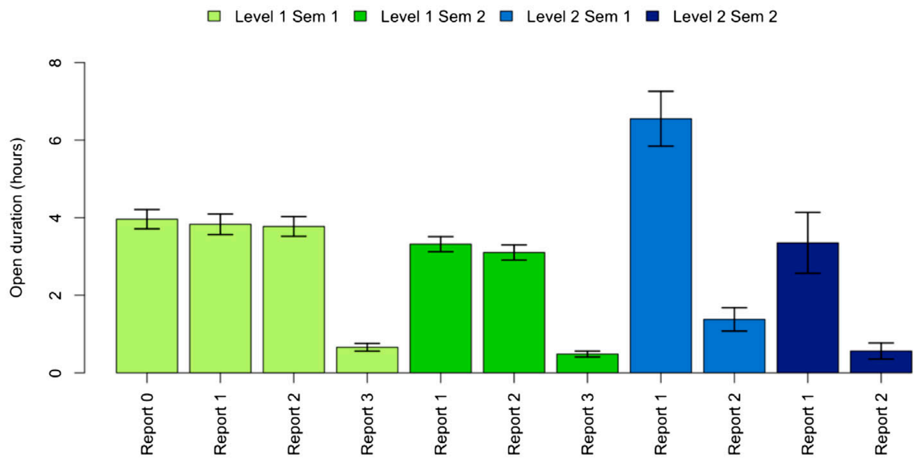


Figure 6. The duration students had their report feedback open for the level 1 course ~~(Report 0: $n = 682$, Report 1: $n = 1571$, Report 2: $n = 1541$, Report 3: $n = 1521$)~~ and for the level 2 course ~~(Report 1: $n = 324$, Report 2: $n = 321$)~~.
~~Notes:~~ Data are expressed as the mean ± SEM time for which each report was open, in hours.

For Figures 2-5 the “Data are expressed…” line is in the main caption instead of as a seperate note. Better to keep that format consistently throughout the paper?

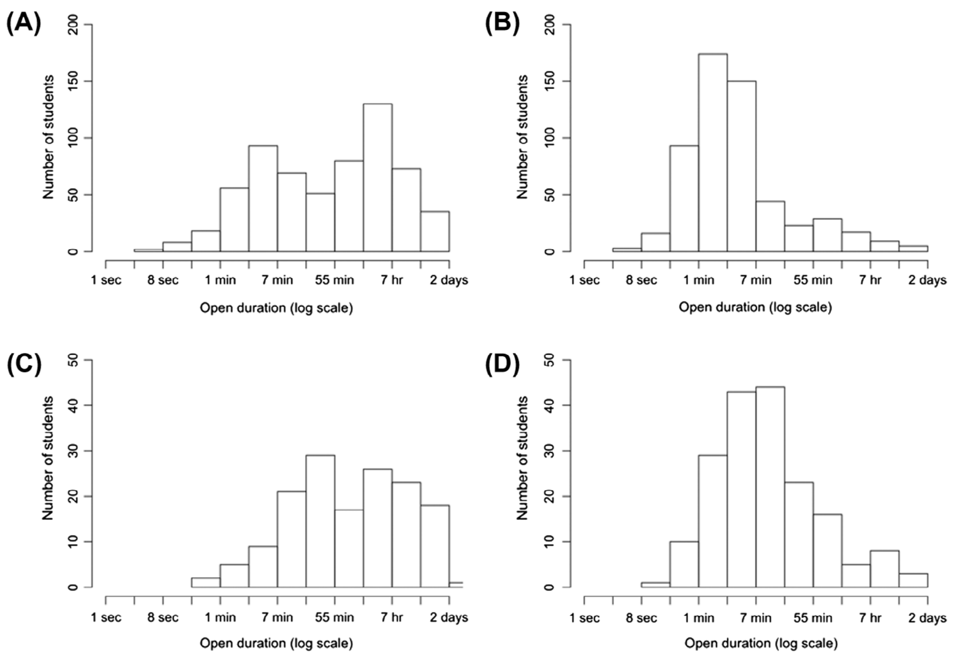


Figure 7. The number of students who had their feedback open for various durations following the level one semester one course (A) formative report (Report 0: $n = 682$), and (B) final report (Report 3: $n = 667$), and the level two semester one course (C) first report (Report 1: $n = 239$), and (D) final report (Report 2: $n = 241$).
~~Notes:~~ Data are presented as a frequency histogram showing the number of student across the log transformed open duration.

students

It is possible that these durations represent students engaging in different categories of behaviour (Warnakulasooriya, Palazzo, and Pritchard 2007): students who work through the feedback thoroughly (intermediate users); students who work through the feedback thoroughly and also use it directly to inform their subsequent report writing (long users); and a small tail of students who glance through their reports very quickly (short users), perhaps primarily to check their grade. This categorisation is supported by the timing and duration of openings across the assessment period. In the first year cohorts it was observed that there was a cluster of openings of marked reports shortly after their release, and another cluster in the period 48 h prior to the due date for the next report, with the latter group being of longer open durations. It is also supported by the pattern of open durations for the final reports from each semester, which have on average, much shorter open durations (Figure 6) and far greater proportions of students falling into the categories of shorter open durations, of <1 min (17%), or between 1 min and an 1 h (57%; Figures 7(B) and 7(D)). The duration and timing of students' interaction with feedback occurring on the non-final tasks suggests that the students perceive these tasks to be sufficiently similar to one another for the feedback to be useful for the subsequent tasks (Boud and Molloy 2013), but also suggests that one of the key drivers for student interaction with feedback is the immediacy of its use on similar assessment tasks.

The analytics used in this study are able to go beyond simple duration of openings, with the 'clickstream' data providing the first insights into the temporal patterns of student interactions with feedback. These clicks represent student interactions with the report document, such as selecting a position within it, opening an audio annotation or scrolling. Based on the total number of times students clicked an element in each of their reports, students interacted twice to three times as much with their non-final reports compared with their final reports in both courses (Table 2). In addition, the pauses between clicks may provide a useful lens for understanding how students are interacting with their feedback. On average, the amount of time students paused between clicks halved from approximately 2–3 min for non-final reports, down to around 1 min for final reports (Table 2). The vast majority of these pauses (83%) fell between 0.4 s and 3 min, with very few pauses exceeding one hour (4%; Figure 8(A) and 8(B)).

This finding suggests that, despite the often considerable duration for which students had their reports open, they were spending much of this time actively interacting with the report, rather than simply leaving it open for extended periods. This pattern of behaviour was also very consistent across all non-final and final reports at both year levels (Figures 8(A) and 8(B)), such that it is likely that this represents the 'normal' pattern by which students interact with in situ feedback. Such detailed data

provided on their

Table 2. Differences in the extent to which students interact with the feedback provided their final report compared with earlier (non-final) reports.

| | Course | Non-final report | Final report | *p* value |
|---|---|---|---|---|
| Interaction (clicks per report) | Level 1 | 105 ± 2 | 38 ± 1 | 0.001 |
| | Level 2 | 202 ± 10 | 67 ± 4 | 0.001 |
| Pause duration (min) | Level 1 | 2.7 ± 0.1 | 1.1 ± 0.1 | 0.001 |
| | Level 2 | 2.0 ± 0.2 | 1.3 ± 0.2 | 0.01 |

Data is reported as mean ± SEM for number of clicks per report and pause duration between clicks.

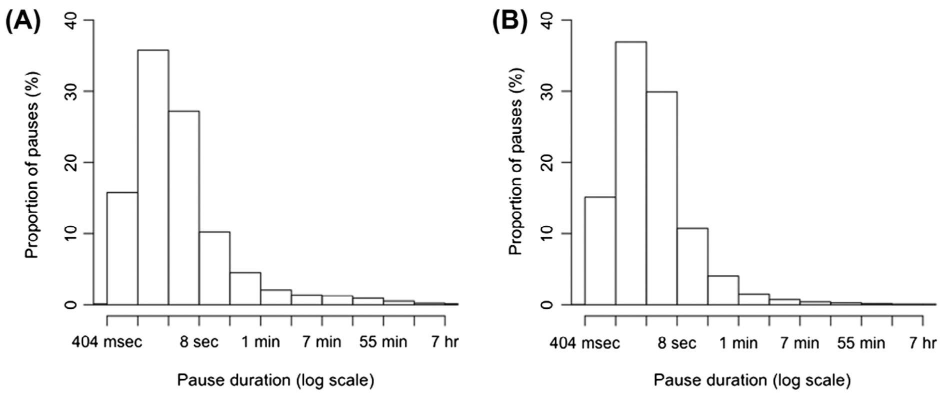


Figure 8. Frequency histogram of the duration that students pause between clicks when interacting with feedback for the (A) non-final reports ($n$ = 3686 reports) and (B) final reports ($n$ = 1509 reports).
~~Notes:~~ Data are expressed as the proportion of pauses students took between clicks, ranging from milliseconds to hours ($x$ scale is log transformed to improve clarity).

representing how students interact with their feedback, and educational materials more generally, is now becoming increasingly prevalent in the age of 'big data' and learning analytics (Long and Siemens 2011), and we are currently at the tip of the iceberg in unpacking the behavioural correlates of such patterns in clickstream data.

At this stage, we hypothesise that minuscule pauses between scrolling clicks might indicate students accessing specific parts of their marked report, while short pauses between clicks are likely to represent students reading the feedback or their submission. Moderate duration pauses might indicate students working outside their feedback (for example on their next report), and very long pauses may indicate students leaving their feedback open in the background. We are currently undertaking studies to match this type of clickstream data to student behaviours in think-aloud videos conducted in interview situations and in natural settings (at home as students use their feedback to work on their next report), and the findings are providing very interesting insights with the potential for large-scale quantitative investigations with widespread generalisability.

### *Academic performance*

Both first- and second-year students showed a steady improvement in their performance on each report within each semester (Figure 9), with the average achievement on each subsequent task being significantly higher than the preceding task, although the performance on the first task in the second-year course was lower than that of the final report in the first-year course. Potentially this reflects the increase in assessment complexity and expectations across the year levels (Colthorpe, Bugarcic, and 2015 Smith forthcoming). Interestingly, the average performance for the first-year students in each semester was comparable between the summative tasks ($p$ = 0.15), but their performance on the formative task in semester 1 was lower than any summative tasks, regardless of semester ($p < 0.001$; Figure 9). This suggests that students in semester 2 are not disadvantaged by the lack of the formative task, potentially the experience they have gained in completing a semester prior to commencing this

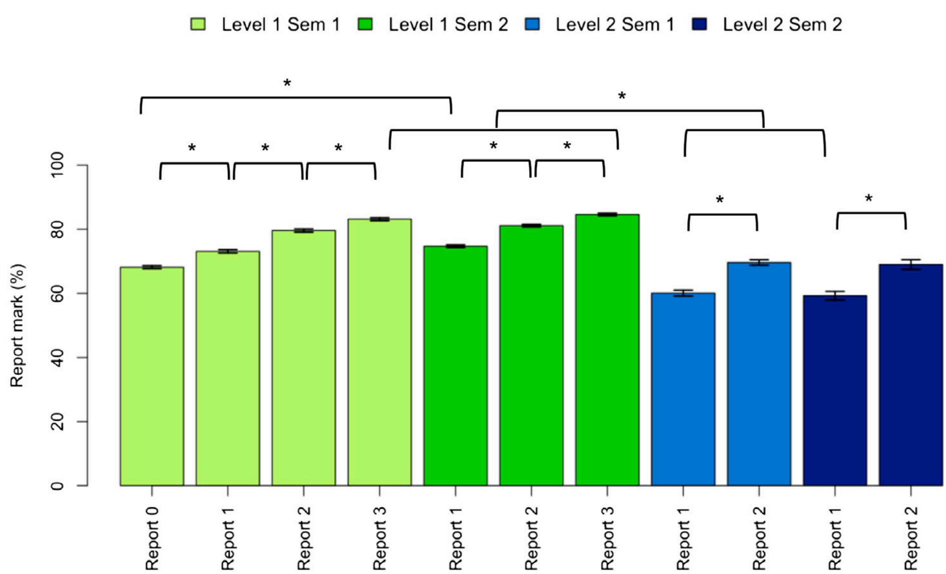


Figure 9. Overall final mark for reports in the level 1 course ~~(Report 0: $n = 682$, Report 1: $n = 1571$, Report 2: $n = 1541$, Report 3: $n = 1521$)~~ and ~~for the~~ level 2 course ~~(Report 1: $n = 324$, Report 2: $n = 321$)~~.
~~Notes:~~ Data are expressed as the mean ± SEM mark that students achieved for each report, as a percentage. *Indicates a statistically significant difference in report mark ($p < 0.05$).

course has been an effective substitute for any learning gains from the formative task (Kolb 1984).

When student performance was compared to the extent to which they viewed their feedback, a number of differences were apparent both in terms of the average mark received on any given report and on changes between reports. For example, first year, first semester students were categorised based on the duration for which they opened their Report 0 feedback, into unopened ($n = 57$), short (<1 min; $n = 26$), intermediate (>1 min < 1 hour; $n = 264$) and long (>1 h; $n = 320$) open durations, and their performance on subsequent reports compared. There were no significant differences in average marks for Report 0 between any of the groups ($p > 0.56$; Figure 10), indicating that students began with equivalent academic performance on this type of assessment task.

Students who did not open, or who opened their Report 0 for short or intermediate durations, did not show any significant improvement in their marks until Report 2 (unopened and intermediate) or 3 (short), whereas those students who opened their Report 0 for long durations improved significantly by Report 1, and continued to improve significantly on each subsequent report ($p < 0.001$; Figure 10). In addition, students who opened Report 0 for longer than one hour had significantly higher marks for Reports 1–3 than students in the intermediate and unopened groups ($p < 0.05$; Figure 10). The pattern of improvements in marks across Reports 1–3 for first-year students in semester 2 is similar; students in the long duration groups consistently outperforming other groups. Second-year students also showed a similar pattern, but, by second year, students who open their Report 1 feedback for long durations already outperform those who do not open it. This may suggest that

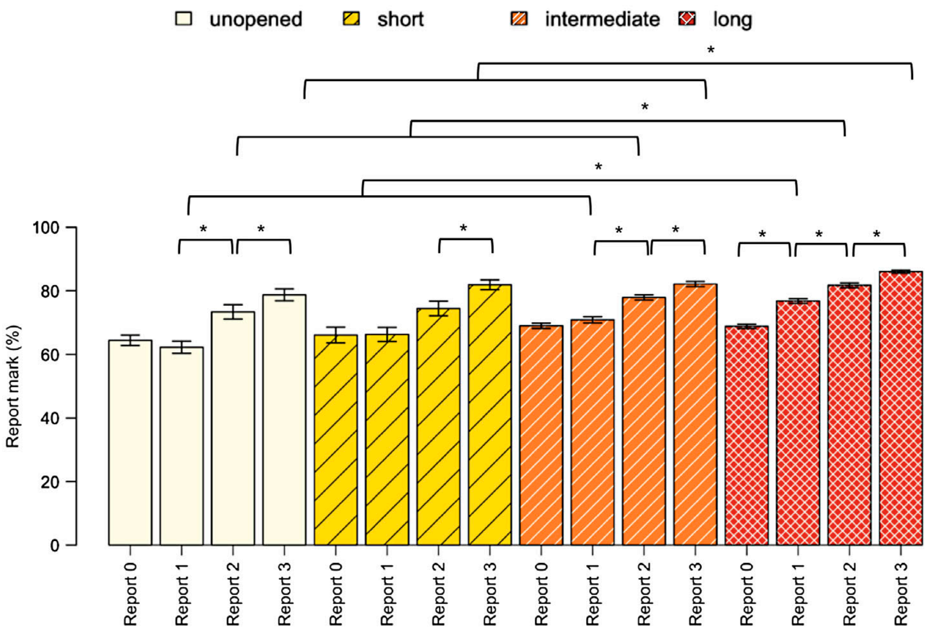


Figure 10. Final marks for reports in the level 1, semester 1 course, with students categorised based on the duration for which they opened their Report 0 feedback into unopened ($n$ = 57), and short (<1 minute; $n$ = 26), medium (>1 minute <1 hour; $n$ = 264) and long (>1 hour; $n$ = 320) open durations.
~~Notes:~~ Data are expressed as the mean ± SEM mark that students achieved for each report, as a percentage. *Indicates a statistically significant difference in report mark ($p < 0.05$).

students in second year who behave differently in relation to their feedback are carrying forward their patterns of behaviour and associated learning gains from first year.

Our use of this feedback analytics system over consecutive years will help us to investigate this longitudinal question in the future. At this stage, our findings are consistent with the views (Boud and Molloy 2013) that students who spend longer with their feedback open are more likely to have greater and more rapid improvements in achievement on subsequent tasks than those who never open or who only open their feedback briefly.

## Conclusions and implications

It is clear from our findings that the majority of students, when presented with feedback electronically, open and interact with that feedback for considerable periods of time. However, the marked decline in use of feedback provided on final reports in any given semester demonstrated that immediacy of usefulness is one of the primary drivers of student interactions with feedback. The findings of this study provide reproducible evidence, from several cohorts of students, which offers empirical support for several theoretical frameworks for feedback practice (Nicol and Macfarlane-Dick 2006; Boud and Soler 2015), and has important implications for practice.

Firstly, it suggests that it is more efficient to focus feedback provision on the earlier tasks in any sequence, as students are less likely to use and demonstrate learning gains from feedback on final tasks. Secondly, it suggests that students will benefit from an increase in the vertical integration of successive tasks, both within and across semesters. If assessment tasks are deliberately designed to build from earlier tasks, and the links between the sequential tasks are made explicit, students may be more inclined to draw on feedback from preceding tasks.

It is also apparent from our findings that the extent to which students interact with their feedback impacts significantly on their performance. Students who had limited or no interaction with their feedback did not improve to the same extent, or as fast, as those with higher levels of interaction. However, this investigation has begun with a broad marker of academic achievement. In the future, a more detailed analysis of variations in individual's behaviour, including their interactions with feedback annotations of varying modalities, and extent of changes in performance within specific criteria, will help to identify those areas students learn easily, those areas which are more problematic and the specific types of feedback which help students through such difficult learning transitions.

Prior to this study, research on the impact of different modalities of feedback (i.e. typed versus audio feedback) has primarily investigated student perceptions of the impact of feedback (Ice et al. 2007; Merry and Orsmond 2008; Lunt and Curran 2010; Rhind et al. 2013; Carruthers et al. 2014), but has not quantified the interactions of students with different forms of feedback, with just a few studies having employed methods to observe students actually using feedback (Dessner 1991; Dohrer 1991; Still and Koerber 2010). Consequently, there has been a clear need for empirical evidence of nature of feedback provision in large teams of markers, student use of that feedback and its impact on subsequent work. To the best of our knowledge, the analytics captured in this study have provided the very first large-scale, quantitative insights into the degree to which students are interacting with the feedback provided on their formative and summative assessment submissions. The initial data generated demonstrates the notable potential of this method to provide quantitative information to researchers, academics, markers and students on the provision of feedback by markers and students' interaction with that feedback, to guide the refinement of theoretical frameworks on effective feedback practices.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the University of Queensland Faculty of Science Teaching and Learning Committee under Grants entitled 'UQMarkUP: Effective feedback and consistent assessment for large cohorts with e-learning enabled "in situ" commenting' in 2012–2013; and the University of Queensland under an Early Career Researcher Scholarship of Teaching and Learning [grant number 2014001951].

## Notes on contributors

Kirsten Zimbardi, PhD, BSc (Hon), GCHEd is a past UQ Teaching Fellow, an early career academic with an international reputation for designing innovative inquiry-based curricula

that develop students' scientific reasoning and chief investigator for the feedback analytics project.

Kay Colthorpe, PhD, DipAppSc (Ag), GDipAgSt, GCertEd (HE) is the chair of the Education Research Unit in the School of Biomedical Science, and past UQ Teaching Fellow, with research focuses on assessment and feedback in the sciences.

Andrew Dekker, PhD, BIT Hons, BMultimedia Design, BInfEnv is a senior developer, UQx, director of Ably Digital Design, winner of the Govhack 2013 National Award. Andrew developed the feedback analytics capture system, working extensively with academic implementers in 2011–2015, and continues to support its implementation.

Craig Engstrom, PhD, MSc, BHMS(Ed)(Hon), holds a UQ Award for Teaching Excellence (2015), was awarded a UQ T&L Excellence Commendation in 2013, and initial leader in developing the audio feedback aspect of the feedback analytics capture system.

Andrea Bugarcic, PhD, MSc, GCHEd, has completed her postdoctoral work at Institute for Molecular Biosciences, and a lectureship in School of Biomedical Science, and is now the associate program leader at the Endeavour College of Natural Health.

Peter Worthy, BAppSc (Anat), LLB, GDipLegalPrac, MIDesign, is a teaching assistant undertaking a PhD on the impact of technology on quality of life and social relationships and the project officer on the feedback analytics project.

A Ruban Victor, BSc, is a postgraduate student in the School of Biomedical Science developing his interests in researching cognitive processes and educational psychology.

Prasad Chunduri, PhD, MSc, GCHEd is a Lecturer with extensive experience in coordinating large courses, and a developing SoTL profile in online teaching and learning technologies.

Lesley Lluka, associate professor, PhD, BPharm, MPharmStud has over 30 years of experience in university teaching and learning, was awarded a UQ Award for Teaching Excellence in 2010, and nominated for Australian Teaching Excellence Awards in 2011 and 2012.

Phillip Long, professor, PhD, AB Psychobiology (Hon) is the Associate Vice Provost for Learning Sciences and deputy director of the Centre for Teaching and Learning at the University of Texas, Austin. He is a co-founder and executive member of the Society for Learning Analytics Research (SOLAR) and international leader in the use of emerging technologies to enhance student engagement and learning.

## References

Boud, D., and N. Falchikov. 2006. "Aligning Assessment with Long-Term Learning." *Assessment & Evaluation in Higher Education* 31 (4): 399–413. doi:10.1080/02602930600679050.

Boud, D., and E. Molloy. 2013. *Feedback in Higher and Professional Education*. London: Routledge.

Boud, D., and R. Soler. 2015. "Sustainable Assessment Revisited." *Assessment & Evaluation in Higher Education* 41 (3): 400–413. doi:10.1080/02602938.2015.1018133.

Brearley, F. Q., and W. R. Cullen. 2012. "Providing Students with Formative Audio Feedback." *Bioscience Education* 20: 22–36. doi:10.11120/beej.2012.20000022.

Carruthers, C., B. McCarron, P. Bolan, A. Devine, and U. McMahon-Beattie. 2014. "Listening and Learning: Reflections on the Use of Audio Feedback. an Excellence in Teaching and Learning Note." *Business and Management Education in Higher Education* 1 (1): 4–11. doi:10.11120/bmhe.2013.00001.

2015

Colthorpe, K. L., K. A. Bugarcic, and A. Smith. Forthcoming. "Progressive Development of Scientific Literacy through Assessment in Inquiry-Based Biomedical Science Curricula."

Colthorpe, K. L., K. Zimbardi, A. Bugarcic, and A. Smith.

*International Journal of Innovation in Science and Mathematics Education* 23 (5): 52–64.

Colthorpe, K. L., S. Rowland, and J. Leach. 2013. *Threshold Learning Outcome 4: Communication*. Sydney: Office for Learning and Teaching.

Dessner, L. E. 1991. "English as a Second Language College Writers' Revision Responses to Teacher-Written Comments." Dissertation available from ProQuest (Paper AAI9125629). Philadelphia.

Dohrer, G. 1991. "Do Teachers' Comments on Students' Papers Help?" *College Teaching* 39 (2): 48–54.

Ellis, C. 2013. "Broadening the Scope and Increasing the Usefulness of Learning Analytics: The Case for Assessment Analytics." *British Journal of Educational Technology* 44 (4): 662–664. doi:10.1111/bjet.12028.

Furnborough, C., and M. Truman. 2009. "Adult Beginner Distance Language Learner Perceptions and Use of Assignment Feedback." *Distance Education* 30 (3): 399–418.

Good, J., K. L. Colthorpe, K. Zimbardi, and G. Kafer. 2015. "The Roles of Mentoring and Motivation in Student Teaching Assistant Interactions and in Improving Experience in First-Year Biology Laboratory Classes." *Journal of College Science Teaching* 44 (4): 88–98.

Gould, J., and P. Day. 2013. "Hearing You Loud and Clear: Student Perspectives of Audio Feedback in Higher Education." *Assessment & Evaluation in Higher Education* 38 (5): 554–566.

Hattie, J. 2008. *Visible Learning*. London: Routledge.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112.

Hepplestone, S., G. Holden, B. Irwin, H. J. Parkin, and L. Thorpe. 2011. "Using Technology to Encourage Student Engagement with Feedback: A Literature Review." *Research in Learning Technology* 19 (2): 89–100. doi:10.3402/rlt.v19i2.10347.

Ice, P., R. Curtis, P. Phillips, and J. Wells. 2007. "Using Asynchronous Audio Feedback to Enhance Teaching Presence and Students' Sense of Community." *Journal of Asynchronous Learning Networks* 11 (2): 3–25.

Jonsson, A. 2012. "Facilitating Productive Use of Feedback in Higher Education." *Active Learning in Higher Education* 14 (1): 63–76. doi:10.1177/1469787412467125.

Kluger, A. N., and A. DeNisi. 1996. "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin* 119 (2): 254–284. doi:10.1037/0033-2909.119.2.254.

Knauf, H. 2015. "Reading, Listening and Feeling: Audio Feedback as a Component of an Inclusive Learning Culture at Universities." *Assessment & Evaluation in Higher Education* 41 (3): 442–449.

Kolb, D. A. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Vol. 1. Englewood Cliffs, NJ: Prentice-Hall.

Long, P., and G. Siemens. 2011. "Penetrating the Fog: Analytics in Learning and Education." *Educause Review* 46 (5): 30–32.

Loughlin, W., S. Gregory, G. Harrison, and J. Lodge. 2013. "Beyond the First Year Experience in Science: Identifying the Need for a Supportive Learning and Teaching Environment for Second Year Science Students." *International Journal of Innovation in Science and Mathematics Education* 21 (4): 13–26.

Lunt, T., and J. Curran. 2010. "'Are You Listening Please?' the Advantages of Electronic Audio Feedback Compared to Written Feedback." *Assessment & Evaluation in Higher Education* 35 (7): 759–769. doi:10.1080/02602930902977772.

MacDonald, R. B. 1991. "Developmental Students' Processing of Teacher Feedback in Composition Instruction." *Review of Research in Developmental Education* 8 (5): 3–7.

Merry, S., and P. Orsmond. 2008. "Students' Attitudes to and Usage of Academic Feedback Provided via Audio Files." *Bioscience Education* 11 (1): 1–11.

Nicol, D. J., and D. Macfarlane-Dick. 2006. "Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice." *Studies in Higher Education* 31 (2): 199–218. doi:10.1080/03075070600572090.

Rhind, S. M., G. W. Pettigrew, J. Spiller, and G. T. Pearson. 2013. "Experiences with Audio Feedback in a Veterinary Curriculum." *Journal of Veterinary Medical Education* 40 (1): 12–18. doi:10.3138/jvme.0912-081R.

Ribchester, C., D. France, and K. Wakefield. 2008. "'It Was Just like a Personal Tutorial': Using Podcasts to Provide Assessment Feedback." Paper presented at the Higher Education Academy Conference, Lancaster University, Lancaster.

Rodway-Dyer, S., J. Knight, and E. Dunne. 2010. "A Case Study on Audio Feedback with Geography Undergraduates." *Journal of Geography in Higher Education* 35 (2): 217–231. doi:10.1080/03098265.2010.524197.

Sadler, D. R. 1998. "Formative Assessment: Revisiting the Territory." *Assessment in Education: Principles, Policy & Practice* 5 (1): 77–84. doi:10.1080/0969595980050104.

Sinclair, H. K., and J. A. Cleland. 2007. "Undergraduate Medical Students: Who Seeks Formative Feedback?" *Medical Education* 41 (6): 580–582. doi:10.1111/j.1365-2923.2007.02768.x.

Still, B. 2006. "Talking to Students Embedded Voice Commenting as a Tool for Critiquing Student Writing." *Journal of Business and Technical Communication* 20 (4): 460–475.

Still, B., and A. Koerber. 2010. "Listening to Students: A Usability Evaluation of Instructor Commentary." *Journal of Business and Technical Communication* 24 (2): 206–233. doi:10.1177/1050651909353304.

Warnakulasooriya, R., D. J. Palazzo, and D. E. Pritchard. 2007. "Time to Completion of Web-Based Physics Problems with Tutoring." *Journal of the Experimental Analysis of Behavior* 88 (1): 103–113. doi:10.1901/jeab.2007.70-06.

West, J., and W. Turner. 2015. "Enhancing the Assessment Experience: Improving Student Perceptions, Engagement and Understanding Using Online Video Feedback." *Innovations in Education and Teaching International.* Ahead-of-print: 1–11. doi:10.1080/14703297.2014.1003954. http://www.tandfonline.com/doi/abs/10.1080/14703297.2014.1003954.

Zimbardi, K., A. Bugarcic, K. L. Colthorpe, J. P. Good, and L. J. Lluka. 2013. "A Set of Vertically Integrated Inquiry-Based Practical Curricula That Develop Scientific Thinking Skills for Large Cohorts of Undergraduate Students." *AJP. Advances in Physiology Education* 37 (4): 303–315. doi:10.1152/advan.00082.2012.

Zimbardi, K., K. L. Colthorpe, A. Bugarcic, A. Dekker, P. Worthy, C. Engstrom, P. Long, L. J. Lluka, and P. Chunduri. 2013. "Analytics of Student Interactions with Electronic Feedback Using UQMarkUp." Transforming Assessment, Melbourne & Brisbane, Australia. http://www.transformingassessment.com/events_4_september_2013.php.

Zimbardi, K., A. Dekker, A. Bugarcic, K. L. Colthorpe, P. Chunduri, J. Kibedi, L. J. Lluka, C. Engstrom, P. Worthy, and P. Long. 2014. "Are Students Reading My Feedback? Using a Feedback Analytics Capture System to Understand How Large Cohorts of Biomedical Science Students Use Feedback." Proceedings of the Australian Conference on Science and Mathematics Education (Formerly UniServe Science Conference). Sydney.