Practitioner Research In Higher Education

Assessment Special Issue
Volume 15 Issue 1
October 2023

PRACTITIONER RESEARCH IN HIGHER EDUCATION

University of CUMBRIA

Contents

Learning from Wikipedia: Using digital social production as an assessment tool in higher education

Frances Edwards, University of Waikato, Hamilton, New Zealand

Abstract

Wikipedia has been viewed as a dubious resource within higher education, and students are often encouraged to avoid its use due to the questionable rigour of its entries. In this study the tables were turned when Wikipedia became the central element in an assessment task. The task experimented with a new approach to summative assessment within a university-taught pre-service teacher education programme. This paper describes an authentic task that allowed students to demonstrate their learning by editing Wikipedia entries and providing justifications for their edits. Findings showed that this assessment task was enjoyed and valued by the students as they saw they could make a positive contribution to the socially produced Wikipedia resource. Additionally, they described the development of their own digital literacy skills gained through engaging with the task.

Keywords

Learning-oriented assessment; authentic assessment; Wikipedia; higher education; pre-service teacher; social production.

Introduction

This article describes a project that was designed to explore the potential benefits and challenges of using Wikipedia as a teaching and learning resource within higher education, and in particular its use as an authentic assessment tool. It describes an assessment task used within the final year of a three-year pre-service teacher education programme in which students had to edit Wikipedia entries to improve the content. The project involved an investigation of the effectiveness of the task and invited responses from students within this programme.

Use of Wikipedia in higher education

Socially generated information is widely available and a source of information accepted by the general public. Wikipedia is one such source of socially generated information. It is a free, multilingual online encyclopaedia which has been generated by a wide range of volunteer contributors (experts and generalists) through a model of open collaboration that is updated in real time (Williams, 2008). This publicly edited, open educational resource is available worldwide, and has been ranked as the 9th most popular website globally by Similarweb, an American web traffic analysis company (Similarweb, 2022). Because Wikipedia entries consist of collaboratively produced knowledge that can be generated and edited by anyone, these entries run the risk of inaccuracy, misrepresentation of facts, and lack of rigour, and can be used for spreading misinformation (Carberry, 2009). As observed by Leuf and Cunningham (2001) "a wiki is not a carefully crafted site created by experts and professional writers and designed for casual visitors. Instead, it seeks to involve the typical visitor/user in an ongoing process of creation and collaboration that constantly changes the website landscape." (p. 16).

As a consequence of the risk of low quality entries (McDowell and Vetter, 2020), students in higher education are inevitably warned not to rely on Wikipedia content, but rather to focus their energies on high quality peer-reviewed literature (Di Lauro and Johinke, 2017). Despite this, research shows that students and indeed academics regularly use it in their everyday lives for matters that range from trivial knowledge checking to important information about their health (Di Lauro and Shetler, 2013; Menchen-Trevino and Hargittai, 2011; Olutola et al., 2016; Selwyn and Gorard, 2016). Wikipedia entries have the potential to be improved in quality as people edit the information on the website,

Citation

and numerous practices and policies have been put into place to encourage this (McDowell and Vetter, 2020). However, university students are not necessarily good at critiquing Wikipedia, with many students unaware of how Wikipedia entries are generated. Research has shown that some students have a level of concern about the credibility of information they read in Wikipedia, but many do not (Menchen-Trevino and Hargittai, 2011).

Nevertheless, Wikipedia has also been used successfully for teaching and learning in tertiary education. For example. Di Lauro (2020) illustrates the positive contribution of Wikipedia editing in her tertiary writing class, and describes this as "an open and sustainable platform for formative and summative assessment" (p. 1). Moreover, Di Lauro's paper emphasises the benefits of texts created by collaboration. Zou et al (2020) develop these ideas when investigating the benefits of flipped learning with Wikipedia, as flipped learning was found to lead to more time and space for collaboration, interaction and active learning. As described below, Wikipedia has also been used as a context for student assessment in recent studies.

Assessment in higher education

Assessment in the tertiary sector is moving from the use of traditional testing, examinations and essays towards utilising more innovative tasks that allow the assessors to make judgements about students' work, based on more "real world" tasks. This means that students are now being given opportunities to demonstrate their learning in a larger variety of ways (Hay and Mathers, 2012; Holgate and Sambell, 2020; Katz and Gorin, 2016; Struyven and Devesa, 2016). Given the time pressure and workload issues within higher education, it makes sense that students are able to learn through the assessment tasks they engage in. This has been discussed as sustainable assessment (Boud, 2000; Boud and Soler, 2016), as learning oriented assessment (Carless, 2015) and as authentic assessment (Sokhanvar et al., 2021; Villarroel et al, 2018). Sustainable assessment means that students will learn things through the assessment process that will inform their thinking after the assessment is complete, that is, assessment "that meets the needs of the present and [also] prepares students to meet their own future learning needs" (Boud 2000, p. 151). By doing this assessment serves "double duty" (Boud, 2000). Authentic assessment is assessment that "focuses on students using and applying knowledge and skills attained in their course to real-life settings" as opposed to assessing their theoretical knowledge without an application component (Miles and Foggett, 2019). It is comprised of three components: realism, cognitive challenge and feedback (Villarroel et al, 2018). The design of such assessment tasks can be intimidating for some tertiary teachers, and to this end Villarroel et al. (2018) have proposed a four-step model to build authentic assessment that they hope will guide those reluctant to adopt authentic assessment. The merits of authentic and learning-oriented assessment through the use of cases, portfolios, action research and the generation of Wikipedia entries have been highlighted as they include providing value for graduates by enriching their skills that may contribute to their employability (Darling-Hammond and Snyder, 2000; Johinke, 2020; Sotiriadou et al., 2019). These researchers position workplaces as places in which creativity, credibility and agility are all expected, and argue that authentic assessment should mirror the use of problem solving strategies that are needed in the workplace. In this way graduate students have experience "using and applying knowledge and skills in real-life settings" (Sotiriadou et al., 2019, p. 3).

Wikipedia editing has been trialled within assessment tasks in some disciplines. Fraser (2020) describes its use by students' editing gene stubs within the context of improving the representation of biomedical scientific information on Wikipedia. This task was found to both improve students' digital literacy skills, and also to contribute to improving the quality of scientific representation in Wikipedia, hence contributing to wider society. Participants in Fraser's (2020) study did encounter challenges, namely choice paralysis when trying to choose a gene stub to work on, the writing style required for Wikipedia, and editor wars with others who would potentially majorly edit student work.

However student feedback indicates that they enjoyed and engaged with the task well. In their study with undergraduate information technology students in Hong Kong, Zou et al. (2020) found that using a flipped classroom approach facilitated student learning, allowing for more interaction between students and in-class collaboration on Wikipedia entries. At postgraduate level Vetter (2020) described the use of a Wikipedia Edit assignment by doctoral students, and again findings highlighted participant learning about collaboration, writing and digital awareness, as well as the challenges of identifying a gap, which Vetter (2020) described as an exceptionally complex task.

Digital literacy
Digital literacy is a skill required by all people engaging with digital material, and particularly important for teachers, given that they will need to employ it themselves in their work, and work with pupils to develop their skill and knowledge. Digital literacy includes:

- critical thinking — questioning how authentic, valid and useful digital information is
- communicating and collaborating with others in the digital space
- using digital tools to design and create compelling original content
- using digital tools to access, use and share information. (National Library of New Zealand, 2020)

To develop these skills it makes sense that students work in contexts where they need to critically evaluate the digital information they access (Patch, 2010). Patch suggested using focus questions to help students critique the information they are accessing, in order to be more critical and ascertain whether the entry they are reading is reliable or unreliable. This sort of activity allows students to be "smarter consumers of online information and more responsible researchers" (Patch, 2010, p. 281). Research has found that students working with Wikipedia can demonstrate increased development of digital awareness and literacy, and levels of criticality (Di Lauro and Johinke, 2017; McDowell and Vetter, 2020). Additionally, Soler-Adillon et al (2018) found that by working on Wikipedia entries, students within two universities changed their perceptions of this platform with respect to reliability, trustworthiness and robustness of the process of posting information.

There has been an acknowledgement in recent literature that platforms such as Wikipedia can be used constructively in higher education, although there is still reticence to do so. Student assignments that include the generation and communication of knowledge using such platforms have been shown to help with learning and the development of critical thinking, as well as contribute to the social production of a resource available freely to society. Research to this point has not encompassed the possible use of Wikipedia edits for students in pre-service teacher education programmes. This study contributes to the literature by focussing on ways that the production of annotated edits of Wikipedia can be used for assessment for students, and as a learning process.

Specifically the research aims to answer two questions:
- Is the use of digital social production as an assessment tool an effective strategy?
- What factors contribute to the effectiveness of using social production as an assessment tool?

Research methods
Participants in this study were members of a third year university course which contributed to a Bachelor of Teaching degree designed for primary/elementary school teachers, and the lecturer who convened the course (the author). The course focussed on educational assessment. Ethical approval and clearance for the reporting of these findings was sought for this study and granted through a formal university ethics committee. All student participants enrolled in the course were provided with an invitation to participate and thirty students agreed to participate in this study.

The first assessment task in the course is the focus of this study. This task was designed to assess the students' knowledge of formative assessment. During the preceding learning period, the students were given the opportunity to explore educational research and teaching practice related to formative assessment, such as peer- and self- assessment, questioning, use of learning objectives and so on. For the assessment task they were asked to choose one of these elements as their focus.

The research was designed to involve the student participants in five stages: familiarisation with Wikipedia, choice and approval of topic, editing media and submission, marking and feedback, post-assessment interview. Firstly participants were provided with information about Wikipedia and how it was constructed and edited. To ensure students understood this process a short video clip was recorded in Zoom, in which the process was demonstrated using the Shared Screen function. Then an assessment task was presented to the students as follows:

Steps involved by students were:
1. Identify a Wikipedia entry that described one aspect of formative assessment.
2. Gain approval by a lecturer who checked to ensure there was enough scope in their choice to allow them to meet the criteria of the assessment task.
3. Cut and paste the selected Wikipedia entry into a Microsoft Word document and turn on track changes
4. Improve the entry by editing the text. Use the comments function to provide rationale for the edits and provide links to literature and commentary from practicum experiences that support the changes.
5. Upload the Word document to an online platform for marking and feedback.

The marking of this assessment was carried out by the lecturer, and focussed on the corrections and additions that students made to their selected entry as well as the reasoning provided in the comments. The marker could see the original Wikipedia entry in black font and all of the edits made in blue font, alongside the justification for these changes. This colour coding made their changes obvious and the comments section made their understanding of formative assessment visible. The completed assignments were collected as data (see Figures 1 and 2 for examples of the edited entries with tracked changes and comments).
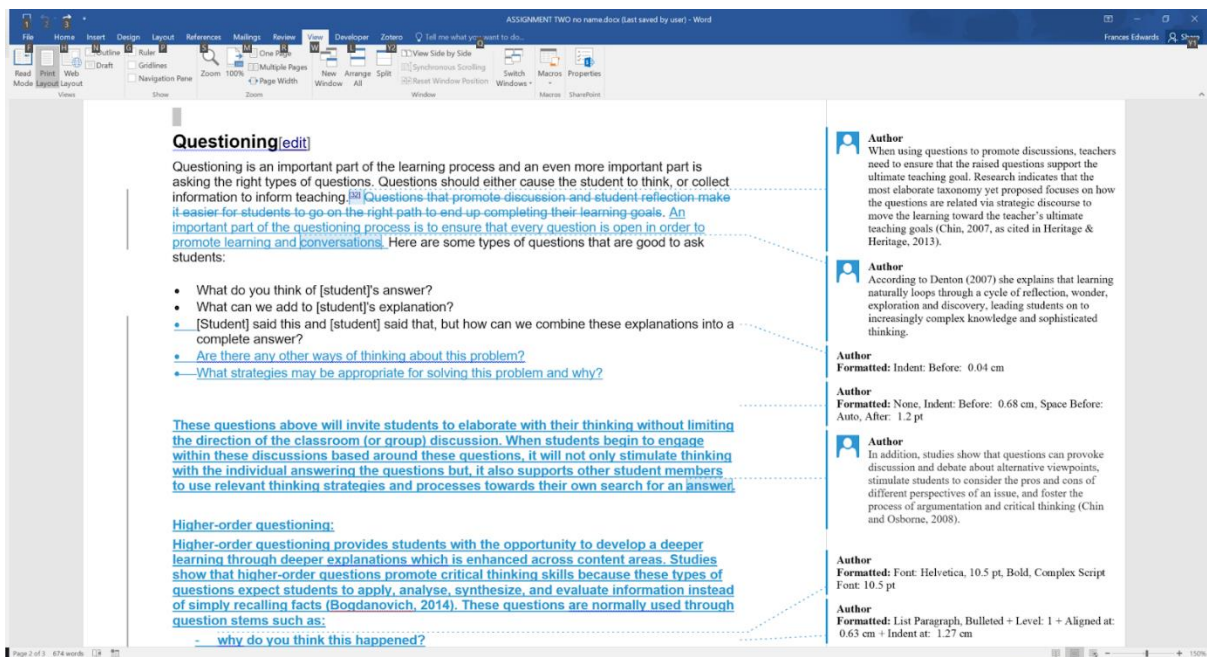

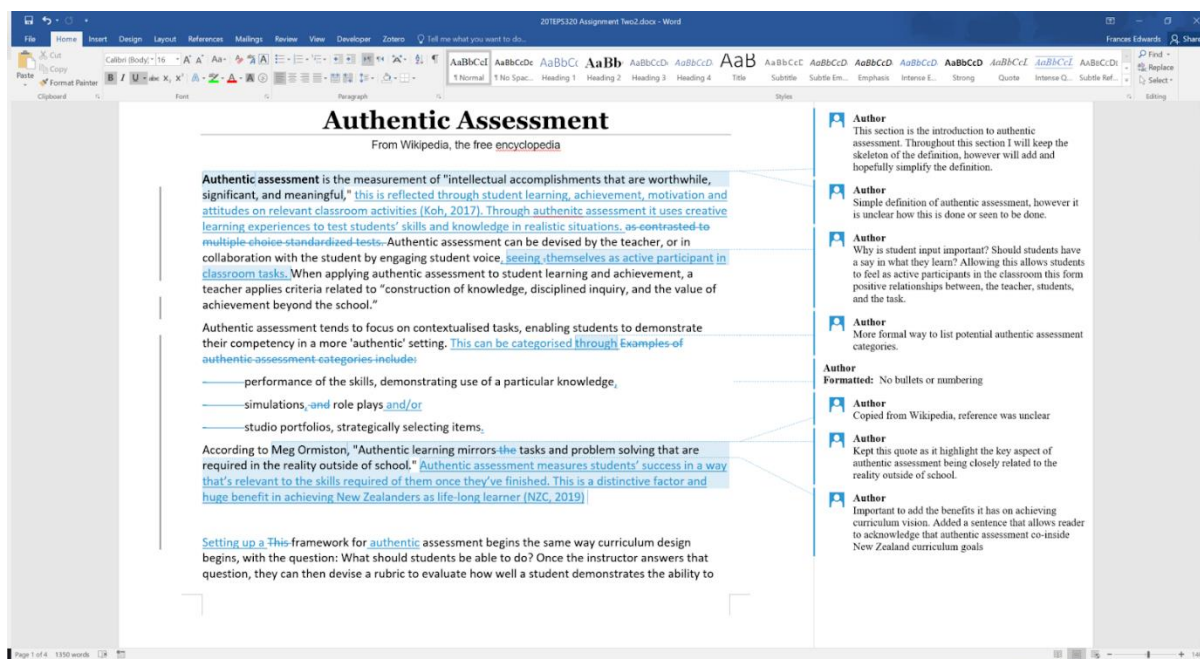**Figure 1.** An example of an edited entry with a significant re-write.

Figure 2. An example of an edited Wikipedia entry with less new text.

After the assignments had been marked and final grades uploaded to the university database, participants were interviewed using a semi-structured interview format (see Appendix for interview prompts). The students were encouraged to talk freely about their experience of the assignment and to comment on its effectiveness or otherwise. Interviews were conducted and recorded using Zoom. Files were then transcribed and the transcriptions returned to student participants to check. All student participants agreed that the transcripts were accurate and represented their ideas. The lecturer involved in the marking of the assignment wrote journal entries to record reflections on the format and content of the work that was submitted by students, and practical elements of using the task. Lecturer reflections included views on the clarity of information, ease of marking, and whether the tasks reflected principles of good assessment.

Data gathered in this study consisted of artefacts of student work i.e. their completed assignments, interview transcripts and lecturer reflections. Thematic analysis was used for identifying, analysing and interpreting patterns that emerged within the data (Clarke and Braun, 2017). This included reading and rereading the data and generating initial codes, then looking for themes across those codes. After this the themes were reviewed and condensed when they were very similar. Then themes were named and checked.

Findings
Findings are presented in two sections: firstly, findings that relate to student experience, and secondly those that relate to the implementation of the task.

Student experience
Importantly and overwhelmingly all students involved in this study spoke very positively about this assessment-type and were able to identify what they learnt through the experience of editing and providing annotations on their edits on Wikipedia entries. The first and most obvious finding is that all students in this study without exception said that they enjoyed the assignment and found it interesting. They talked positively about the assessment structure and the freedom that was afforded:

> I actually really liked it because it was different…it wasn't just an essay or something that … it was assessed in a different way which was nice, refreshing… Like the freedom of what you did it on in the first place… the topic… and then also how you rewrote it – there was a lot of freedom with that.   (Student A).

They described that they found it easier to improve a piece of writing (the Wikipedia entry), than to start writing something from scratch. For those who did not like writing essays and other long forms of writing, this task proved to be more accessible, and they enjoyed the process more. Some comments in this regard were:

> One aspect we found positive was that we did not have to start with a blank page, but instead were able to work to improve something that someone else had written.   (Student B).

> Enough scope – yes, it is much easier improving something than writing something from scratch.   (Student C).

> I like that it gave us a different opportunity to present our knowledge… I feel that that is important for people that aren't fans of essays.   (Student D).

These quotes illustrate that the assessment structure was helpful for the students. Although a few students chose the same excerpts to edit, they each followed their own lines of interest so the completed entries were quite different from each other.

Most participants in this study commented on their perception that they learned about their topic of focus as they completed the assessment task. They were able to identify specific learning that occurred for them. Many explained that they chose a topic they were particularly interested in and that they wanted to learn more about, so the task was seen as a learning opportunity as well as an assessment task.

> I feel it was good for us to learn as well, like I learnt so much about the topic that I chose and I feel that a lot of people did as well.   (Student A).

> … so although it was a wordy assignment it was really well done in that it wasn't overwhelming but I still learnt so much about summative and formative forms of assessment because of the research aspects of it.   (Student E).

They explained that as they edited the Wikipedia entry they engaged with literature and other scholarly writing, and also considered links between what they read and what they experienced on practicum. The prospect of their entry potentially being loaded on Wikipedia made the task seem more real and urgent for them.

Participants felt they were able to provide genuine evidence of deep learning through the completion of the Wikipedia editing task. They valued the opportunity to critique the writing of others, both through their use of reference to the academic literature and through their own experiences and observations while on practicum.

> I found it was a really good way of going about something instead of just writing what we know…critiquing what someone else has written... which I thought was quite valuable. You know when you are taking something quite inaccurate there are a lot of opportunities to show what you know… it takes it a bit deeper. I did like it.   (Student F).

Students felt that being able to explain their reasoning for the edits confirmed and validated their learning. In their attached comments the students added detail from literature they had read, personal reflections from practicum experiences, and even recent media articles, and acknowledged these sources. At times they also added comments to sections that they did not edit, and explained why these sections were strong and accurate and could be left unchanged.

Students' understanding of how Wikipedia actually works was limited, so they appreciated learning about the way the general public were able to contribute to this website. This knowledge was "eye opening", as they explained how they became more aware of how to critically analyse the online material they read. A number commented on the range and quality of literature used to substantiate Wikipedia entry claims; something they had not considered before. For example one student explained:

> It was a really practical way of realising that what you are reading online may not be right, and I think being here at University we get that, but I think it is a really good reminder that what I'm reading on Wikipedia, even though it is referenced, is actually super outdated and missing a whole lot… I often go to Wikipedia for information, for quick information knowing it's not a super-reliable source but … oh they're using this meta study that sounds super-really-great but actually it has been updated and improved massively about 10 years later… and it doesn't mean that the study was wrong but like we've moved on from 1996, that was before I was even born… but we don't always check the references that we read.
>
> (Student F).

Through the editing of Wikipedia entries, students had the opportunity to carefully check on the references that were used in their chosen entry, and the information it conveyed. At times they found irrelevant or outdated references being used to back up content. For example, in the following quote the student alluded to the potential that existed.

> It was good to see some of the gaps between Wikipedia and what *could* be there.
>
> (Student C).

All participants explained how they gained an increased sense of understanding of how the Wikipedia resource could be improved for the benefit of all readers, and the importance of adding recent research and literature to entries. A number found the possibility of being able to do this as they contribute to a "live" online website exciting.

Students in this study described things they had learned that would inform their future teaching. For example they spoke about how the task helped them develop better digital literacy and critical thinking skills. This was identified as something they would need to be teaching students in their future classes. A number built on this idea by explaining that they could use a modified version of this class with their own classes next year, as a way of developing their students' digital literacy. As one participant commented:

> I did like it. It's something I would do in my own class, I think it is a good idea…. It did help a lot… starting with a set piece of writing it definitely helped a lot.
>
> (Student A).

The transferability of this task to teachers' further work was seen as a positive outcome from the course.

Implementation of the task
From the lecturer's perspective this assignment was easy to set up and the expectations were

effectively communicated to the students by two means: written description of the assignment and a recorded video clip which briefly described the assignment and provided a shared-screen demonstration of how to access Wikipedia, select a topic, set up their Word document, use track changes and so on. A few questions were fielded from the students before they began but they all seemed to find the instructions clear. As one student commented:

> I thought the instructions on how to change it on Word were also really straight forward, so I had no problems with the assignment.   (Student E).

There was value in the students being offered the opportunity to have their entry selection checked by the lecturer. This approval step was included to ensure that their choice provided them with enough scope to demonstrate their understanding of an aspect of formative assessment in some detail. The lecturer found that most students chose appropriate entries or sections of entries to work on. One student chose a very long entry, and she was advised to work on just one section. The lecturer found that interaction with the students about their choice during tutorial times was productive as some discussion ensued about what they planned to do, enabling some formative feedback at this stage.

The lecturer found this assignment to be a particularly interesting one to mark, given that students chose a wide range of topics within the broad field of formative assessment. There was obvious deep engagement in the activity of editing. It was easy to see what changes had been made (they were highlighted in blue in the document). The comments that students added to justify their edits were easy to follow and many were quite sophisticated, revealing their depth of understanding and the literature that informed their edits. The content that the students added and in the comments used to justify these changes were "to the point" with no fillers. As one student commented:

> You have to think about and write in not so many words, so you have to be really specific and know what you are talking about. You don't have to fill it with garbage.
> (Student G).

In further commentary he admitted that he would often be very verbose in essays just to get to the word limit, whereas in this assignment he was encouraged to be concise with his language. When comparing this assignment to previous assignments this one was quicker to mark and there appeared to be a deeper engagement from the students.

A challenge in the assessment task design was the "length" of the entry. Usually, university assignments provide guidelines for length. Although 800-1000 words in length was stipulated, it soon became obvious that this was not a useful limit to put on the assignment, as some entries needed to be completely rewritten, given the inaccurate or out of date material they presented, whereas others were improved through more judicious edits. The participants did comment on this as a difficulty, as mentioned here:

> You can't really put a word count on it… With the people I talked to anyway, most people figured it out on their own. They may be some people that would struggle with that.   (Student A).

The students also suggested that the provision of exemplars would have been helpful, and this is an additional resource that could easily be added to the assessment in the future.

Discussion
This study demonstrates that the editing of Wikipedia can be used effectively within a pre-service teacher education programme as a means to assess learning and as a means to develop students' critical thinking and digital awareness and literacy. Wikipedia has been found to be useful in tertiary

education in other contexts (Di Lauro, 2020; Fraser, 2020; Vetter, 2020; Zou et al., 2020) but no literature has been located regarding its use in assessment within teacher education.

Through the editing of a Wikipedia entry, students provided evidence of their knowledge and understanding of an aspect of formative assessment, which is one of the key learning outcomes in a university course. For all participants this was their first involvement in digital social production, and they all spoke very positively about the experience. Johinke (2017) describes digital social production as authentic assessment, and, although the task used in this study did not require the students to demonstrate their understanding of formative assessment in a practical sense (i.e. actually using it in a classroom), the task can be considered authentic given that the work potentially contributed to a resource accessible world-wide. This purpose provided the students a heightened sense of the value of their contribution, and in this sense they saw their work as a contribution to the real world.

The task design aligned with Villarroel et al.'s (2018) model to build authentic assessment in that it provided a rich and worthwhile task requiring higher order skills, and it engaged students in judgement and layers of feedback, while reflecting the skills required in the graduate profile of teachers. This type of assessment task can also be described as sustainable learning-oriented assessment as its contribution to learning is evident and extends beyond the timeframe of the course (Boud and Soler, 2015). Sustainable assessment has been defined as assessment "that meets the needs of the present and [also] prepares students to meet their own future learning needs" (Boud, 2000, p. 151). Students commented that their active engagement in the assessment task caused them to consider their future use of formative assessment as well as their ongoing developing digital awareness, so it served "double duty" described by Boud (2000). For example some students spoke of using the same sort of tasks with their future classes. So although the assessment task was not positioned as part of the learning activities, it was found to fill that role.

The development of critical thinking skills and upskilling in digital literacy were two outcomes from the use of Wikipedia for assessment that were a surprise to the students themselves. These findings are similar to those found by Di Lauro and Johinke (2017) and McDowell and Vetter (2020), where both teams found that criticality and digital awareness were outcomes in other contexts. Knowing about these outcomes is particularly useful when considering the participants in this study. Students are charged with development of critical thinking skills and digital literacy in their future students. Their awareness of this development in themselves through their completion of the Wikipedia editing task has the potential to effectively equip them for their work in future years. When comparing what the students learnt to the four aspects of digital literacy (National Library of NZ, 2020) it was evident that this assessment task lead to development in each of these skills: critical thinking, communicating and collaborating, using digital tools to design and create compelling content, and using digital tools to access, use and share information. The activity allowed students to ascertain more readily the reliability of online content and challenged their perceptions about the trustworthiness on online information. This is testament to the power of a single assessment task.

At an affective level, students in this study were overwhelmingly positive about the Wikipedia assessment task and the process of contributing to something bigger than an assessment which would generate a grade. Numerous studies have shown links between students own emotional experiences and how this shapes their views of teaching practice (e.g. Antilla, Pyhältö, Soini and Pietarinen, 2017; Brown, Gebril, Michaelides and Remesal, 2018; Edwards, 2020; Stuart and Thurlow, 2000). This bodes well for the Wikipedia task's potential to engender effective and longer lasting learning for the students.

From this study a number of factors were seen to contribute to the effectiveness of using social production as an assessment tool in education. Firstly the task provided the students with a new,

interesting and different way to present their learning. Being offered opportunities to demonstrate their learning in a larger variety of ways is an increasingly common feature of assessment in higher education (Holgate and Sambell, 2020; Katz and Gorin, 2016; Struyven and Devesa, 2016), and can be motivating for students. This task provided the students with a real purpose (beyond the need to generate a grade) and allowed them choice and a level of freedom they had not often experienced, which added to their positive experience. Secondly the assessment strategy allowed students to clearly demonstrate their learning in ways that they felt were easier and more accessible. This benefit of accessibility to those who, for example, did not enjoy essay writing or found it difficult to get started on writing, was obvious in the student commentary. A third factor that contributed to the overall effectiveness of this assessment task was the spin-off learning that happened for the students. They identified that both their digital literacy and critical thinking skills developed as a result of engaging with the assignment, highlighting the move in purpose of assessment from assessment of learning to "future practice for which courses are a precursor" (Boud and Soler, 2016, p. 401). Given that they were pre-service teachers, this assessment task was particularly authentic for the students. Not only will they be using their professional judgement to decide what and how to teach in the post truth world (Boyd, 2022), but they will also be teaching their future pupils to be critical consumers of media and information.

From the course lecturer's perspective the work proved easy and interesting to mark. An added bonus were tangible learning benefits of students acquiring digital literacy, and potentially teaching strategies for their future classes which can only be of benefit to the students. These factors are likely pragmatic push factors affecting academic staff future decisions around the design of assessment tasks, and in particular using a framework of digital social production as both assessment and learning purposes. The main challenge of the task from an implementation perspective was the difficulty in quantifying how much content was required from the students. As their chosen topics varied considerably so did the editing required. Additionally some students did find it hard to choose a focus topic for the task. These sorts of challenges are difficult to mitigate if the task to remain authentic, using actual online sources. However students could receive individual feedback on their chosen material, including advice about the scope of editing required.

Conclusion

The process of utilising Wikipedia as a resource within an assessment task provides insight into the possible range of benefits as experienced by students. The assessment experience was found to be positive, interesting and useful by the participants. Additionally they identified the benefits of a deeper understanding of socially produced online resources and their growth in digital literacy. Through a fairly simple process of preparing and annotating edits in Wikipedia, these positive contributions will have ongoing impact on the pre-service teachers as they commence working with their own students in schools. The sustained outcome from a learning-oriented assessment as its contribution to learning is evident and extends beyond the timeframe of the course.

References

Anttila, H., Pyhältö, K., Soini, T., and Pietarinen, J. (2016) How does it feel to become a teacher? Emotions in teacher education. *Social Psychology of Education*, 19(3), pp.: 451–473. https://doi.org/10.1007/s11218-016-9335-0

Boud, D. (2000) Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education, 22*(2), pp. 151-167. https://doi.org/10.1080/713695728

Boud, D., and Soler, R. (2016) Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), pp. 400-413. https://doi.org/10.1080/02602938.2015.1018133

Boyd, P. (2022) Teachers developing research-informed practice in the post-truth world. *Research in Teacher Education* 21(1), pp. 47-52. https://www.uel.ac.uk/sites/default/files/rite-may-2022-guest-author-pete-boyd.pdf

Brown, G. T., Gebril, A., Michaelides, M. P., and Remesal, A. (2018) Assessment as an emotional practice: Emotional challenges faced by L2 teachers within assessment. In Agudo, D. (Ed.) *Emotions in second language teaching.* SpringerCham. pp. 205-222.

Carberry, G. (2009, May 6) Student's Wikipedia hoax quote used worldwide in newspaper obituaries. *The Irish Times*. Available at: https://www.irishtimes.com/news/student-s-wikipedia-hoax-quote-used-worldwide-in-newspaper-obituaries-1.759391 (Accessed: August 2022).

Carless, D. (2015) *Excellence in university assessment: Learning from award-winning practice.* London: Routledge.

Darling-Hammond, L., and Snyder, J. (2000) Authentic assessment of teaching in context. *Teaching and teacher education,* 16(5-6), pp. 523-545. https://doi.org/10.1016/S0742-051X(00)00015-9

Di Lauro, F. D. (2020) 'If it is not in Wikipedia, blame yourself' edit-a-thons as vehicles for computer supported collaborative learning in higher education. *Studies in Higher Education,* 45(5), pp. 1003-1014. https://doi.org/10.1080/03075079.2020.1750191

Di Lauro, F. and Johinke, R. (2017) Employing Wikipedia for good not evil: innovative approaches to collaborative writing assessment. *Assessment and Evaluation in Higher Education,* 42(3), pp. 478–491. https://doi.org/10.1080/02602938.2015.1127322

Edwards, F. (2020) The effect of the lens of the teacher on summative assessment decision making: the role of amplifiers and filters. *The Curriculum Journal,* 31(3), pp. 379-397. https://doi.org/10.1002/curj.4

Fraser, S. T. (2020) A new frontier: developing an undergraduate assessment task aimed at improving the representation of biomedical scientific information on Wikipedia*. Studies in Higher Education,* 45(5), pp. 972-983. https://doi.org/10.1080/03075079.2020.1749794

Hay, M., and Mathers, L. (2012) Designing assessment for autonomous learning. *Practitioner Research in Higher Education,* 6(2), pp. 95-106. http://insight.cumbria.ac.uk/id/eprint/1425/

Holgate P., and Sambell, K. (2020) The professional doctorate by portfolio: alternative assessment for advanced practitioner-led scholarship? *Practitioner Research in Higher Education Journal,* 13(1), pp. 57-69.

Johinke, R. (2020) Social production as authentic assessment: Wikipedia, digital writing, and hope labour. *Studies in Higher Education,* 45(5), pp. 1015-1025. https://doi.org/10.1080/03075079.2020.1750192

Katz, I., and Gorin, J. (2016) Computerising assessment. in Brown, G. T. L. and Harris, L. R. (ed.), *Handbook of Human and Social Conditions in Assessment.* New York: Routledge. Pp. 472-479.

Leuf, B. and Cunningham, W. (2001) *The wiki way: quick collaboration on the web*. Addison-Wesley.

Similarweb (2022). Available at: https://www.similarweb.com/website/wikipedia.org/#overview (Accessed: August 2022).

McDowell, Z. J. and Vetter, M. A. (2020) It takes a village to combat a fake news army: Wikipedia's community and policies for information literacy *Social Media + Society,* 6(3), pp.e 1-13. https://doi.org/10.1177/2056305120937309

Menchen-Trevino, E. and Hargittai, E. (2011) Young adults' credibility assessment of Wikipedia. *Information, Communication & Society,* 14(1), pp.24-51. https://doi.org/10.1080/13691181003695173

Miles, C. A. and Foggett, K. (2019) A shift from discouraging academic dishonesty to authentic assessment. https://www.acode.edu.au/mod/resource/view.php?id=675 (Accessed: June 2022).

National Library of New Zealand. (2020) Why digital literacy matters. Available at: https://natlib.govt.nz/schools/digital-literacy/understanding-digital-literacy/why-digital-literacy-matters (Accessed: June 2022).

Olutola, A. T., Olatoye, O. O., and Olatoye, R. A. (2016) Assessment of social media utilization and study habit of students of tertiary institutions in Katsina State*. Journal of Education and practice,* 7(3), pp. 178-188.

Patch, P. (2010) Meeting student writers where they are: Using Wikipedia to teach responsible scholarship. *Teaching English in the Two Year College,* 37(3), pp. 278.

Selwyn, N. and Gorardb, S. (2016) Students' use of Wikipedia as an academic resource — Patterns of use and perceptions of usefulness. *The Internet and Higher Education,* 28: 28-34. http://dx.doi.org/10.1016/j.iheduc.2015.08.004 (Accessed: June 2022).

Shetler, A. M. and Di Lauro, F. (2013) Writing with Wikipedia: Building ethos through collaborative academic research. in K. E. Pytask, R. E. Ferdig and T.V. Rasinski (eds.) P*reparing teachers to teach writing using technology.* Pittsburgh, PA: ETC Press. pp. 201 – 222.

Soler-Adillon, J., Pavlovic, D., and Freixa, P. (2018) Wikipedia in higher education: Changes in perceived value through content contribution. [Wikipedia in the University: changes in the perception of value with the creation of contents]. *Communicate,* 54, pp. 39-48. https://doi.org/10.3916/C54-2018-04

Sokhanvar, Z., Salehi, K., and Sokhanvar, F. (2021) Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation, 70*: 101030. https://doi.org/10.1016/j.stueduc.2021.101030 (Accessed: June 2022).

Sotiriadou, P., Logan, D., Daly, A., and Guest, R. (2019) The Role of Authentic Assessment to Preserve Academic Integrity and Promote Skill Development and Employability. *Studies in Higher Education*, 45(11), pp. 2132-2148*.* https://doi.org/10.1080/03075079.2019.1582015

Stuart, C., and Thurlow, D. (2000) Making it their own: Pre-service teachers' experiences, beliefs and classroom practices. *Journal of Teacher Education,* 51(1), pp. 113–121. https://doi.org/http://dx.doi.org/10.1177/002248710005100205

Struyven, K., and Devesa, J. (2016) Students' perceptions of novel forms of assessment. in Brown, G. T. L. and Harris, L. R. (Eds.), *Handbook of Human and Social Conditions in Assessment*. London: Routledge. pp.129-144.

Vetter, M. A. (2020) Broadening representations of rhetoric in Wikipedia: disciplinary praxis as graduate pedagogy and research. *Studies in Higher Education,* 45*(*5), pp. 990-1002. https://doi.org/10.1080/03075079.2020.1749798

Villarroel, V., Bloxham, S.,Bruna, D., Bruna, C., and Herrera-Seda, C. (2018) Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), pp. 840-854, https://doi.org/10.1080/02602938.2017.1412396

Williams, P. (2008) Assessing context-based learning: not only rigorous but also relevant. *Assessment and Evaluation in Higher Education,* 33(4), pp. 395–408. https://doi.org/10.1080/02602930701562890

Zou, D., Xie, H., Wang, F., and Kwan, R. (2020). Flipped learning with Wikipedia in higher education. *Studies in Higher Education,* 45(5), pp. 1026. https://doi-org.ezproxy.waikato.ac.nz/10.1080/03075079.2020.1750195

Appendix

Interview prompts for semi-structured interview with student participants:

1. Tell me about your experiences in completing Assignment 2?

2. What in particular do you see as the strengths of the assignment? Why?

3. What in particular do you see as the weaknesses of the assignment? Why?

4. How well did the assignment give you opportunity to provide evidence of what you knew and could do with respect to Formative assessment?

5. If you were asked to improve the assignment task what would you suggest?

David Bell, Vikki O'Neill and Vivienne Crawford, Queen's University Belfast, Northern Ireland

Abstract

We compared the influence of open-book extended duration versus closed book time-limited format on reliability and validity of written assessments of pharmacology learning outcomes within our medical and dental courses. Our dental cohort undertake a mid-year test (30xfree-response short answer to a question, SAQ) and end-of-year paper (4xSAQ, 1xessay, 1xcase) in pharmacology. For our first year medical cohort, pharmacology is integrated within a larger course, contributing 20xclinical vignette questions (to select the single best answer (SBA) to each question from a choice of 5 plausible answers) to a mid-year test and 3-5xSAQ to an end-of-year paper. Our experience indicates that SAQ are as reliable as SBA for closed-book time-limited assessments; reliability correlates with number of questions employed. We have found good correlation between mid-year and end-of-year performance (predictive validity), between questions (factorial validity) and between pharmacology and other subjects within the assessment (concurrent validity). Adoption of open-book extended duration assessments resulted in only modest reduction in reliability and validity.

Key words
Validity; reliability; single best answer; short notes questions; Covid-19 pandemic.

Introduction

A range of strategies exist to assess the academic performance of healthcare students enrolled on professional courses using traditional closed-book, time-limited written examinations including: single best answer (SBA) or only correct answer selected by the candidate from a choice of multiple answers (MCQ) to a question, (very) short answer free written response to a question (SAQ), longer essay-style questions and structured problems and clinical case studies (Fallatah et al., 2015; Hift, 2014; Sam et al., 2016; Wilkinson and Shaw, 2015). There is a lack of consensus whether candidates tend to score more highly in SBA / MCQ assessments than in free written response assessments (Preston et al, 2020; Sam et al, 2016; Sullivan, 2011; Wilkinson and Shaw, 2015); the former have been criticised for encouraging cueing and superficial learning (Holzinger et al., 2020), reflecting student perception that SBA/MCQ assessments are easier and require less effort to be invested in learning (Holzinger et al., 2020; Jaenicke et al., 2020; Preston et al., 2020).

Reliability is defined as the extent to which an assessment method or instrument measures consistently the performance of the candidate (Andreatta and Gruppen, 2009; Downing, 2003; Fallatah et al.; 2015; Sullivan, 2011). Reliability is measured by analysing the correlation between the answers to multiple questions. For this purpose, Cronbach's alpha can be applied to short free-response answers, whereas Kuder-Richardson 20 [KR20] represents a special case of Cronbach's alpha as applied to dichotomous answers and is more suitable for measuring reliability of SBA assessments. Values tending to 1 indicate high reliability.

Validity defines how well the assessment tool employed actually measures the underlying outcome of

Citation

interest. There are various facets to validity (Andreatta and Gruppen, 2009; Downing, 2003; Fallatah et al.; 2015; Patil et al, 2015; Sam et al 2016; Sullivan, 2011). Content validity establishes that assessment strategies sample the breadth of the curriculum, often employing sampling grids. Construct validity ensures that assessment strategies are mapped to professional learning outcomes at an appropriate level (for example, blueprinting to General Medical or General Dental Curriculum outcomes for graduates- GDC 2015; GMC, 2018). Predictive (Criterion) validity determines the extent to which an assessment outcome predicts performance in another future assessment. Concurrent validity addresses the extent to which performance in the assessment correlates with performance in another assessment by the same cohort of candidates. Finally, factorial validity measures the extent of correlation of different discrete factors within the whole assessment.

UK medical schools are increasingly drawing on SBA/MCQ style-assessments. Experience indicates that the assessment strategy adopted will likely impact how students approach their learning (Preston et al, 2020; Witt et al., 2022). Although students may prefer SBA/MCQ format and find such assessments easier these might not necessarily test deeper understanding and facilitate acquisition of long-term knowledge to the same degree as other strategies (Holzinger et al., 2020; Witt et al., 2022). There is a lack of consensus in the educational literature however as to whether SBA/MCQ assessments are more reliable and convenient than free-response SAQ, but at the cost of validity (Hift, 2014; Holzinger et al., 2020; Patil et al., 2015). The first objective therefore was to analyse the reliability and validity of our approaches to assess achievement of learning outcomes for basic and clinical pharmacology content within the undergraduate medical and dental curricula.

It has been argued that open-book extended duration assessments encourage deeper engagement and application of knowledge relative to simple recall of memorised factual information in closed book time-limited assessments. Open-book assessments may also generate less student anxiety but are more susceptible to cheating (Spiegel and Nivette, 2023). The second objective therefore was to compare the impact of open-book, extended duration format versus traditional time-limited closed-book format on the reliability and validity of our assessment strategies. The COVID-19 pandemic necessitated that many assessments adapted to open-book, extended duration format, providing further incentive and opportunity to address this objective.

Methods
Organisational context and study cohorts
Medicine at QUB
At The Queen's University of Belfast has an intake of >260 students annually [261+5.0, mean + sd, n=8 years]. Prior to 2017, an introduction to pharmacology and therapeutics was incorporated together with pathology into a single semester module in spring of Year 1 of 5 (20 CATS): Principles of Disease and Treatment. Assessment comprised an end of semester written examination (weighted 100%, 2h, short answers x10, of which 5 were pharmacology and 5 were pathology). Since 2017, introduction to pharmacology and therapeutics has been part of larger full-year first year module (40 CATS). Assessment comprises 2x 40 SBA mid-year class tests each weighted 15% (2h, 20 SBA pharmacology and therapeutics and 20 SBA pathology; 40 SBA genetics and biochemistry) together with an end of year examination weighted 70% (2h, short answers x10, of which pharmacology x3, pathology x3, genetics and biochemistry x4).

Dentistry at QUB
There is an intake of 55-60 students annually [58+5.6, mean + sd, n=7 years]. Pharmacology for Dentistry is a full year module delivered in Year 2 of 5. Assessment of this comprises a mid-year class test (1h, open-ended very short answer questions x30, weighted 10%), end of year examination (2h,

short answers x4 , essay x1, case study x1, weighted 80%) and a group presentation on pharmacological management of a dental condition (15min, weighted 10%) .

Our assessment strategies map to respective professional learning outcomes (General Dental Council, 2015; General Medical Council, 2018), ensuring construct validity. Free response answers are marked by one internal examiner but 25% of scripts are double-marked by a second internal examiner prior to review by the external examiner: inter-marker variability is <5%. Detailed model answers and marking schemes indicating allocation of marks are provided for all questions for use by the examiners to facilitate standard setting and reduce potential for inter-marker variability. A representative selection of these drawn from past papers is also provided to students for the purposes of self-assessment and formative feedback in advance of the summative examination.

Example questions

Dental class test: List TWO advantages of adding a vasoconstrictor to a local anaesthetic preparation
Dental short notes question: Write short notes on mode of action, therapeutic use and adverse effects of (1) ibuprofen; (2) nystatin
Dental essay question: Discuss the pharmacological management of diabetes mellitus and its implications for dental practice

---

Dental case study: in the case below a series of questions is based around the pharmacology of the drugs taken, management of the dental condition and any acute medical emergencies arising during dental practice. Answer <u>all parts</u> of the following case history:

Mr Brown is a 68 year old life-long smoker. His medical history includes chronic obstructive pulmonary disease. He had been using a salbutamol inhaler as required for relief of his breathlessness, particularly on exertion.

(a) Describe the mechanism of action by which salbutamol provides symptomatic relief.

This has helped, but more recently his symptoms had been getting worse and more frequent and his doctor had commenced him on a regular (preventer) Seretide inhaler containing salmeterol plus fluticasone.

(b) Explain how the properties of salmeterol differ from salbutamol.

(c) Name the class of drug that fluticasone belongs to and describe the mechanism of action.

(d) Name a side-effect affecting the oral cavity that is associated with inhaled fluticasone.

(e) Describe two practical measures that Mr Brown could take when using his Seretide inhaler to minimise the risk of the side-effect identified in (d) from occurring.

(f) If despite the measures recommended in (e) the side-effect identified in (d) did occur, what drug could be prescribed to treat the condition?

(g) Describe the mechanism of action of drug identified in (f)

A routine dental examination reveals halitosis and periodontitis, with red, swollen and recessed gums, significant calculus and some loosening of several teeth. During the root surface instrumentation to clean below the gum-line, Mr Black becomes breathless.

(h) How would you manage the acute breathless episode?

As an adjunct to mechanical debridement to remove calculus Mr Brown's dentist decided to prescribe an antibiotic therapy.

(i) Name an antibiotic that is indicated for the treatment of periodontitis and justify your choice.

(j) Describe the mechanism of action of the drug identified in (i)

(k) List any TWO steps Mr Brown could take to improve his oral hygiene and prevent progression of his periodontitis?

Despite the use of the regular preventer inhaler, Mr Brown is still experiencing persistent breathlessness and his doctor decided to step up his treatment with addition of a third preventer drug, tiotropium, prior to his next dental appointment.

(l) What class of drug does tiotropium belong to and what is the mechanism of action?

(m) Name a side-effect affecting the oral cavity associated with inhaled tiotropium?

Medicine short notes question: Compare and contrast the mode of action, clinical indications and adverse effects of apixaban and warfarin.

Medicine sample SBA/MCQ question:

> *A 73 year old woman is taking atorvastatin. What is the key mode of action of this drug?*
> (a) Activation of HMG Co-A reductase enzyme
> (b) Activation of PPARa receptors
> (c) Inhibition of HMG Co-A reductase enzyme
> (d) Reduced absorption of cholesterol from the intestine
> (e) Reduced LDL receptor expression

Adaptation to open book extended duration format during the Covid-19 pandemic 2020-2022
Dentistry: the mid-year class test was held in January 2020 in the traditional closed book format before restrictions were introduced. The end of year paper held in May 2020 was delivered remotely in open-book format with extended duration of 24 hours for completion. The mid-year class test in January 2021 was also delivered remotely, in open-book format and was of 90 minutes duration rather than the usual 60 minutes. The end of year paper held in May 2021 was delivered remotely in open-book format, but time available for completion was scaled back from 24 hours to 3 hours. In January and May 2022 the class test and paper both reverted to closed-book in-person completion within a traditional examination venue and returned to pre-covid duration of 1 hour and 2 hours, respectively.

Medicine: the first mid-year class test was held in December 2019 in the traditional closed-book format before restrictions were introduced; the second mid-year class test scheduled for March 2020 was cancelled. Questions pertaining to Biochemistry and Genetics were omitted from the end of year short notes examination as the related learning outcomes had been examined in the December class test. Candidates received instead a paper in May 2020 which contained 5x pathology and 5x pharmacology and therapeutics short notes questions, which was delivered remotely in open-book format of an extended (24 hour) duration.

Data-sets and Analysis:

In this study, we analysed data on the performance of our dental students in closed book assessments (held prior to Covid-19, January 2015-January 2020), open-book assessments (delivered during the pandemic, May 2020-May 2021) and on return to closed book assessments (post-pandemic, January 2022-May 2022). We have also included data on the performance of our medical students in closed book assessments (held prior to Covid-19, May 2013-December 2019) and open book assessments (held during the pandemic, May 2020). [In September 2020, QUB Medical School introduced a new case-based learning curriculum and a strategy of assessing all subjects and specialties in an integrated fashion through regular progress testing (Heeneman et al, 2015) so we have not been able to include any comparable data beyond May 2020 in the present study]. Statistical analysis was undertaken using GraphPad Prism (Version 5) and SPSS (Version 23) to generate reliability and correlation coefficients. This research study relates to taught courses at The Queen's University of Belfast and received ethical clearance within the arrangements provided by the University for taught programmes.

Results
Dental Curriculum
In closed book, time-limited assessments held between January 2015 and January 2020 performance was 20.39+1.16 out of 30 (68%), mean + sd., n=6 in the mid-year class test and 51.46+2.8 out of 80 (64%), mean + sd., n=6 in the end of year examination (standard set cut score was 38.5+2.2 out of 80, mean +sd., n=6). Cronbach's alpha was 0.82+0.06 and 0.74+0.09, mean + sd., n=6 for the mid-year class test and end of year examination, respectively, evidencing reliability. No significant differences were noted between subsections /question type within the end of year examination in regards to reliability.

There were positive correlations between subsections of the paper indicating factorial validity: short notes v essay Pearson's r= 0.42+0.16; short notes v case study Pearson's r=0.44+0.22; essay v case study Pearson's r=0.42+0.09; mean + sd., n=6, each p<0.001.   Performance in the mid-year test correlated (p<0.001) with performance in the end of year paper (Pearson's r=0.63+0.05, mean + sd., n=6) indicating predictive validity. Neither correlated with performance in the group presentation. As illustrated by data obtained for the 2019 cohort, the mark for the Pharmacology for Dentistry module correlated (p<0.001) with those for the Physiology for Dentistry (Pearson's r= 0.65) and Disease Mechanisms for Dentistry (Pearson's r= 0.71) modules, evidencing concurrent validity.

In 2021, in which the mid-year class test was delivered remotely in extended open-book format, performance was higher relative to that observed in previous years: 25.74+5.3 out of 30 (86%), mean + sd., 60 students. In 2022, the mid-year class test reverted to pre-Covid-19 in-person, closed-book short-duration format and performance reverted to pre-pandemic levels (19.74+5.3 out of 30 (66%) mean + sd., 56 students). Performance in the remotely-delivered open book end of year examination in 2020, which was of 24 hours duration, was also inflated: 68.97+6.44 out of 80 (86%), mean + sd., n=57 students. Grade inflation was less evident in 2021 when duration of the open book remotely delivered examination was reduced from 24 to 3 hours: 60.45+7.07 out of 80 (76%), mean + sd., n =60 students. In 2022, on reverting to a closed-book in-person time-limited examination, performance returned to (and indeed was lower than) pre-pandemic levels: 44.95+9.04 out of 80 (56%), mean + sd., n= 56 students.

Reliability of the class test was reduced on switching to remote open-book extended duration assessment in 2021 (Cronbach's alpha = 0.63) but returned to pre-pandemic levels with reinstatement of the in-person closed-book time-limited assessment in 2022 (Cronbach's alpha = 0.88). Similarly, reliability of the end of year paper was reduced by open-book remote delivery (24 hour assessment 2020, Cronbach's alpha = 0.65; 3 hour assessment 2021, Cronbach's alpha =0.68) but returned to pre-

pandemic levels with reintroduction of a time-limited in-person closed-book examination in 2022 (Cronbach's alpha= 0.77).

There was still evidence for correlation between sections of the remotely-delivered open-book written paper (during the pandemic in 2020 and 2021) evidencing factorial validity but correlations were less positive and less significant than when using closed-book time-limited format: short notes v essay, Pearson's r=0.37+0.11, p<0.01 mean +sd., n=2; short notes v case study: Pearson's r=0.35+0.14 p<0.05 mean + sd., n=2; case study v essay: Pearson's r=0.42+0.14 p<0.01 mean + sd., n=2. Performance in the mid-year class test still correlated with that in the end of year paper but less strongly than before indicating a reduction in predictive validity: 2020 Pearson's r= 0.25, p=0.06; 2021 Pearson's r=0.33, p<0.05. Correlation improved markedly in 2022 on return to closed-book, time-limited in-person assessment: Pearson's r=0.77, p<0.0001. with use of open-book assessments of extended duration during the pandemic, correlation between the Pharmacology for Dentistry module mark and those for the Physiology for Dentistry (Pearson's r= 0.45, p<0.001) and Disease Mechanisms for Dentistry (Pearson's r=0.57, p<0.0001) modules for the same cohort of 60 second year dental students (2021) was less positive but still significant. Correlation of marks achieved across modules improved markedly in 2022 on return to closed-book, time-limited in-person assessment: correlation between the Pharmacology for Dentistry module mark and Physiology for Dentistry module mark: Pearson's r= 0.73, p<0.0001; correlation between the Pharmacology for Dentistry module mark and Disease Mechanisms for Dentistry module mark: Pearson's r=0.75, p<0.0001.

Medical Curriculum
Reliability of the short notes paper increased with the number of questions included: during 2013-2016, in which candidates completed 5 questions in 60 minutes in closed-book format, the mean performance was 35.39 out of 50 (71%) and Cronbach's alpha was 0.79+0.06, mean + sd., n=4; from 2017-2019 during which candidates completed 3 questions in 36 minutes in closed-book format, the mean performance was 17.92 out of 30 (60%) and Cronbach's alpha was 0.66+0.10, mean + sd., n=3. The mean performance in the closed-book SBA class test introduced from 2017 onwards was 13.5 out of 20 (68%) for the 20 pharmacology questions completed in 30 minutes and KR20 was 0.67+0.03, mean + sd., n=3. Predictive validity was evidenced by correlation between pharmacology questions in the SBA class test and end of year short notes pharmacology questions in 2019 (Pearson's r=0.53, p<0.001). Factorial validity was shown by correlation between the pharmacology question styles within the short notes examination: basic principles of pharmacology v drug comparison question Pearson's r=0.46, n=7, p<0.0001.; basic principles of pharmacology v clinical case study Pearson's r=0.40, n=7, p<0.0001; drug comparison question v clinical case study Pearson's r=0.52, n=7, p<0.0001. Concurrent validity was evidenced by correlation between the pharmacology SBA questions and the pathology SBA questions within the same class test undertaken in 2019: Pearson's r=0.42, p<0.001. There was also correlation between sections of the closed-book, time-limited short notes assessment: pharmacology v pathology (10 questions in total 2013-2016): Pearson's r =0.61+0.06, mean + sd., n=4, p<0.001; pharmacology v pathology (6 questions in total 2017-2019) Pearson's r= 0.54+0.18, mean + sd., n=4, p<0.001; pharmacology v biochemistry and genetics (7 questions in total 2017-2019) Pearson's r= 0.53+0.16, mean + sd., n=3, p<0.001.

Performance in an 2020 open-book extended duration (24 hours) short notes examination delivered remotely during the pandemic was markedly greater than in the pre-pandemic closed-book time-limited examination; 86.94%+2.73, mean + sd., n=267. Reliability was reduced but not significantly different to that of the closed-book time limited examination in previous years; Cronbach's alpha was 0.70 (5 questions). Factorial validity between different question types within the paper was also reduced but remained statistically significant: basic principles of pharmacology v drug comparison question: Pearson's r= 0.33, p<0.0001; basic principles of pharmacology v clinical case study: Pearson's

r= 0.20, p<0.001; drug comparison question v clinical case study: Pearson's r= 0.36, p<0.0001. Concurrent validity was also reduced but remained significant: correlation with the pathology section of the paper: Pearson's r=0.30, p<0.001 .Similarly, predictive (criterion) validity was reduced but remained significant: correlation with the time-limited closed book SBA biochemistry and genetics class test earlier in the year: Pearson's r= 0.213, p<0.001.

Discussion

There is a lack of consensus regarding whether candidates tend to score more highly in SBA /MCQ assessments than in free response SAQ assessments (Sam et al 2016; Sullivan, 2011; Wilkinson and Shaw, 2015) and whether the former enhance reliability at the expense of validity (Hift, 2014; Patil et al., 2015). Preston et al (2020) have proposed that students perform more highly in SBA/MCQ assessments in part because these are most frequently employed and students become more familiar with this format as they progress through the course. SBA were incorporated into the profile for assessment of basic and clinical pharmacology learning outcomes within our undergraduate medical curriculum in 2017; we have found no evidence that students score consistently higher in such assessments than in traditional free response short notes examinations. However, our study was restricted to students in the early years of the 5 year medical and dental programmes during which students may still be becoming familiar with this format.

Furthermore, in our experience, appropriately constructed free response short notes papers can be as reliable as SBA papers for assessing basic and clinical pharmacology learning outcomes in closed-book, time-limited assessments: utilising approximately 30-35 minutes allocated to the pharmacology section of a larger assessment with the medical curriculum permitted the use of either 3 short notes questions or 20 SBA: Cronbach's alpha was similar regardless of which strategy was employed. Reliability of each assessment method would be expected to increase if more time was available enabling an increased number of questions to be used; for example we found Cronbach's alpha was higher when 5 short notes questions were set rather than 3. Student perception in regards to subjectivity in marking of free response answers (Holzinger et al, 2020) can be mitigated by construction of detailed model answers and internal (and external) quality assurance of the paper. In our experience such measures have resulted in an inter-maker variability of less than 5% particularly when internal examiners are subject experts of many years' standing and experienced markers. SBA arguably offer the convenience of automated marking, but short notes questions in our experience afford more opportunity for meaningful feedback at cohort and individual level which students consider to be very important to allow improvements (Preston et al, 2020), whereas protection of the security of the question bank of quality assured discriminating SBA questions is often prioritised over provision of specific and sufficiently detailed feedback to candidates and can encourage dependence by students on external SBA question banks of variable quality. Furthermore, Preston et al (2020) have reported that students perceive that performance in short answer questions more accurately reflects the effort they put into learning and their knowledge of the content material than that afforded by SBA/MCQ-based assessment.

For both forms of assessment, content and construct validity increase when a greater number of questions is used. The argument that SBA papers by virtue of containing a greater number of questions and allocating 1-2 minutes to answer each enhances content validity relative to a much smaller number of free response short notes questions allocated 10-15 minutes each for completion is misleading: appropriately constructed short notes questions (potentially with multiple parts) together with detailed model answers and marking schemes can permit assessment of multiple learning outcomes and aspects of the subject material within the same question. In both our medical and dental short notes assessments, factorial validity was evident when comparing across question types, for example comparing case studies and drug comparison questions, which supports the construction

of short notes papers that include a range of question formats. Positive correlation between SBA and short notes assessments within our medical curriculum also advocates for inclusion of a range of written assessment methods rather than dependence on one assessment type. The performance of our dental students in the written assessments did not however correlate with performance in the group presentation; this can be attributed to the latter assessing a different set of subject specific and generic learning outcomes relative to the written papers and also group size diluting individual performance.

The Covid-19 pandemic led to a dependence on remotely-delivered open-book assessments often of extended duration. A desire to protect the security of SBA question banks so that robust quality-assured SBA questions could be re-used at a future date also encouraged renewed interest in the use of short notes papers as an assessment strategy during the pandemic. Not unexpectedly, open-book, remotely-delivered assessments encouraged grade inflation although this could be mitigated in part by limiting the extent to which the duration of the assessment was extended to the absolute minimum necessary to facilitate downloading and uploading of the paper and typing of answers by candidates. The decision was taken not to alter the standard set cut score (which is based on assessor judgement of the performance of a borderline student by the modified-Angoff method, George et al., 2006) to account for the Covid-19 related change to open-book assessment delivery. Despite grade inflation, there was still opportunity to discriminate between stronger and weaker candidates, as evidenced by the spread of marks. Dependence on open-book extended duration assessments resulted in reduction in assessment reliability; this was unavoidable but the impact was more modest than anticipated and could be mitigated by inclusion of an increased number of questions. Open-book remotely delivered assessments did not reduce content or construct validity since sampling strategy and blueprinting was unchanged. There were modest reductions however in factorial, predictive and concurrent validity of assessments employed; however although less positive, correlations remained statistically significant. Such reductions may be accounted for by the open-book format and extended duration; access to learning resources and increased time available to construct answers often of greater length may have increased variation in a candidate's answering of differing styles and types of question within pharmacology and also more widely, across all subjects assessed concurrently.

A possible limitation of the current study is that analysis was undertaken by the investigators using historical spreadsheets provided by the Progress and Assessment Office Administrative Manager of anonymised data so it was not possible to undertake subgroup analysis of possible influencing variables: only the final ratified overall module mark was available to the investigators in de-anonymised form. The majority of students enrolled in our medical (60-75%) and dental (65-80%) programmes in the last decade are female. Routine quality assurance by our Progress and Assessment Office of the summative assessments undertaken within our medical and dental courses has not however found any statistical difference between male and female candidates in terms of their academic performance in either MCQ/SBA or free response short notes assessments providing reassurance that the assessment strategy employed would not confer a gender-related advantage. Small subgroup size has limited opportunity for analysis of other potential influences such as those with registered disability, or home /international status. Our medical and dental students enter university with a strong foundation in the sciences, acquired during secondary education, and/or a primary degree in a relevant subject. Caution is necessary when extrapolating to other disciplines for whom the SBA and/or SAQ assessment format may be less suitable: it would be interesting to explore the wider application of the study findings by comparing the reliability and validity of SBA and SAQ assessments in measuring performance of students undertaking courses in the arts and humanities which may wish to assess achievement of different generic and subject-specific learning outcomes.

In conclusion, appropriately constructed papers employing short notes free response answers are reliable and valid for assessing basic and clinical pharmacology learning outcomes in closed-book, time-limited written assessments. We would argue for their retention within the assessment strategy, alone or possibly used in combination with other forms of written assessment (such as SBA) to increase assessment variety, discourage cueing and foster deeper learning and critical understanding (Witt et al., 2022) as they afford opportunity for provision of detailed constructive feedback to candidates. Students also acknowledge the benefit afforded by inclusion of a variety of assessment types to accommodate for a range of learning styles and needs (Holzinger et al., 2020; Preston et al, 2020). Dependence on short notes papers for remote delivery of open-book extended duration assessments during the Covid-19 pandemic did not reduce content or construct validity but did cause modest reduction in factorial, predictive and concurrent validity and reliability. With the benefit of experience, we propose that this could potentially be mitigated in future by careful consideration of the optimum number of questions employed and restriction to assessment duration. Thorough item analysis could discard ambiguous or poorly performing questions to enhance reliability. Adaptation of questions to place greater emphasis on critical understanding and problem-solving rather than simple factual recall would also compensate for inability to rigorously enforce closed-book conditions on a remotely-delivered assessment without the significant logistical challenge of online proctoring of a large cohort of candidates.

References

Andreatta, P.B., Gruppen, L.D. (2009) 'Conceptualising and classifying validity evidence for simulation', *Medical Education 2009*;43, pp. 1028-1035.

Downing, S.M. (2003) On the meaningful in'terpretation of assessment data', *Medical Education 2003*; 37, pp. 830-837.

Fallatah, H.I., Tekian, A., Park, Y.S., Al Shawa, L. (2015) 'The validity and reliability of the 6th Year internal medicine examination', *BMC Medical Education 2015*; 15, pp.10.

General Dental Council (2015) Preparing for Practice 2015. https://www.gdc-uk.org/docs/default-source/quality-assurance/preparing-for-practice-%28revised-2015%29.pdf (Accessed: September 2022).

General Medical Council (2018) Outcomes for Graduates 2018. https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/outcomes-for-graduates/outcomes-for-graduates (Accessed 21 September 2022).

George, S, Haque, M.S., Oyebode, F. (2006) 'Standard setting: comparison of two methods', *BMC Medical Education* 2006; 6, pp. 46.

Heeneman, S., Schut, S., Donkers, J., van der Vleuten C.P.M., Muijtiens, A.J.J.M. (2016) 'Embedding of the progress test in an assessment program d,esigned according to the principles of programmatic assessment', Medical Teacher; *DOI: 10.1080/0142159X.2016.1230183*

Hift, R.J. (2014) 'Should essays and other open-ended type questions retain a place in written summative assessment I clinical medicine?,' *BMC Medical Education 2014*; 14, pp.249.

Holzinger, A., Lettner, S., Steiner-Hofbauer, V., Melser, M.C. (2020) 'How to assess? Perceptions and preferences of undergraduate medical students concerning traditional assessment methods', *BMC Medical Education 2020*; 20, pp. 312.

Jaenicke, J., Jhet, M. T, Dave D, Makeev, V., Nasim, S.A. (2020) 'Student reflections on clinical prioritisation questions and their assessment', *Med Teach 2020*; 42. pp. 1193-1194.

Patil, S.Y., Gosavi, M., Bannur, H.B., Ratnakar, A. (2015) 'Blueprinting in assessment: a tool to increase validity of undergraduate written examinations in pathology', *Int J App Basic Med Res 2015*; 5 Suppl 1.

Preston, R., Gratani, M., Owens, K., Roche, P., Zimany, M., Malau-Aduli, B.(2020) 'Exploring the impact of assessment on medical students'', *Learning. Assessment & Evaluation in Higher Education 2020*; 1, pp. 109-124.

Sam, A.H., Hameed, S., Harris, J., Meeran, K. (2016) 'Validity of very short answer versus single best answer questions for undergraduate assessment', *BMC Medical Education 2016*; 16, pp.266.

Sullivan, G.M. (2011) A primer on the validity of assessment instruments', *J Grad Med Ed 2011*; 3, pp.120-121.

Wilkinson, T., Shaw, H. (2015) 'Are spot test multiple choice questions easier to answer than short answer questions?', *FASEB J 2015*; 29. Pp.344 (1).

Witt, E.E., Onorato, S.E., Schwartzstein, R.M. (2022) 'Medical students and the drive for a single right answer', *ATS Scholar 2022*; 3, pp. 27-37.

Effects of adaptive comparative judgement on student engagement with peer formative feedback

Claire Ellison, University of Salford, Manchester UK.

Abstract

Developing assessment and feedback strategies to assist students with progression and graduation is a key focus for many higher education institutions. However, student engagement with feedback is often poor and they can find it difficult to act upon; often stating the feedback is generic or of insufficient quality for improvement. Here, I present the outcomes of integration of adaptive comparative judgement as a strategy of peer formative feedback amongst a small cohort of students. Adaptive comparative judgement a process that allows work to be marked by making comparisons between pieces of work, rather than assessing work against a mark scheme or rubric. Student opinions on the access to examples of work, and personalised feedback through online tools are discussed. Engagement and self-reflection were measured through collection of qualitive data obtained from questionnaires. Positive outcomes included improved self-awareness and regulation by students as they were more active and engaged with formative feedback. The study also demonstrated that running comparative judgement is possible with a small cohort of students. However, engagement of students can be variable and is improved with dedicated timetabled sessions. Further work is required to assess whether increased engagement with feedback translated to an improvement in the standard of work students produced.

Keywords

Assessment; Feedback; Peer comparison; Self-efficacy.

Introduction:

Developing assessment and feedback strategies to assist students with their learning to allow progression through their degrees and eventual graduation is a key focus for many higher education institutions in England. This is for both fundamental intrinsic reasons related to the benefits to the students themselves, but also due to external pressures such as the focus on "continuation and completion rates" as a lead indictor in institutional monitoring by the Office for Students (Office for Students, 2022). However, from my personal experience, student engagement with feedback is often poor and they can find it difficult to act upon; often stating the feedback was generic or of insufficient quality for improvement. Thus, it is important to find assessment and feedback strategies that support student learning. Here I present one strategy (adaptive comparative judgement) that had positive impacts on students' learning, including improved self-awareness and regulation by students as they were more active and engaged with formative feedback.

There is a large variety of assessment and feedback strategies used in higher education institutions across the world, especially if one considers both the formative and summative strategies used to assist and assess student learning respectively. These are well summarised by Evans (2013), who provides a broad review of assessment in the Higher Education environment, discussing the outputs of 490 papers on the topic of assessment. Haughney et al. (2020) also provide a thorough look back over the last 20 years of research into assessment and feedback and examines some of the wider issues around successful feedback strategies. For example, the authors highlight the impact of well-known factors of good feedback including student and staff buy-in to the process, quality of feedback in terms of specificity and positivity, and time for students to action any feedback they receive. In

addition, the authors also highlight less well-known features that have a positive impact, including peer feedback, novel forms of feedback and novel feedback tools. Vattøy et al. (2020) examines student opinion of feedback at the Higher Education level and discusses the problems they can face when engaging with feedback. Indeed, students are often less satisfied with feedback than any other aspect of their course (Carless & Boud, 2018; Nicol et al., 2014). However, a critical point to note is that students are ultimately responsible for actioning feedback they receive, and therefore improvement in feedback practices will not simply come from academics producing more detailed feedback for an ever-increasing number of students (Carless & Boud, 2018). It is critical that students self-regulate their feedback into actionable outcomes by being able to accurately judge the quality of their own work (Evans, 2021). Thus, although there is a huge choice of assessment and feedback strategies available to us in Higher Education, it is important that we deeply consider what works best for the students in front of us and what best practice would be in any given scenario.

As mentioned previously, one emerging assessment and feedback strategy that has demonstrable positive impacts on student learning is the use of peers (Haughney et al., 2020; Venables & Summit, 2003). Simonsmeier et al. (2020) highlight a difference between peer feedback and peer assessment that is worth clarifying at this stage. The authors state the term 'peer assessment' should only be used if peers were to award a final summative grade to the work of a peer, whereas any other feedback which helps students improve without a summative grade, can be considered peer-feedback. Numerous studies have demonstrated the value of such feedback on student learning, with the feedback often being more specific, easier to understand, and more actionable than that given by academic staff (see references in (Simonsmeier et al., 2020)). Also, students benefit when exposed to the work of peers as they make internal comparisons between their own work and that of their peers (McConlogue, 2015; Nicol et al., 2014; Thompson & Meer, 2021). Thus, the process of generating peer feedback is a valuable tool to facilitate student learning in Higher Education; the comparison process is of intrinsic value and the benefit does not rely solely on comments being left (Nicol & McCallum, 2021). However, successfully implementing marking criteria requires implicit judgement calls on the standard of work that staff often do not make explicit to students (Sadler, 2005). Thus it is no wonder that students lack the skill required to make suitable judgements (Ibarra-Sáiz et al., 2020) and instead prefer to compare work to an exemplar piece (Carless, 2015; Kean, 2012).

Comparative judgement in an assessment context refers to the process of marking work by comparing it to other pieces of work, rather than assessing against a mark scheme or rubric. The idea that humans are more able to pick the better of two options than they are ranking a list of items is a long standing one (Thurstone, 1927). In an assessment context, markers find it easier to state if a piece work is 'better' than a comparator, rather than state if it is an inherently 'good' piece of work (Potter et al., 2017). Multiple studies have demonstrated that comparative judgement can produce outcomes as consistent as traditional marking practices (Hardy et al., 2016; Jones & Alcock, 2014; Seery et al., 2012). Repeating the comparison process multiple times across an array of submissions allows a rank of work to be formed. This could then be used in a norm-referencing manner to assign grades to work (Wheadon et al., 2020) or as a formative feedback mechanism to students (Hardy et al., 2016). Theoretically, to get a full ranking of all work, each piece would have to be compared to each other, which could lead to huge number of comparisons being required depending on cohort size. For example, a small group of 10 submissions would require a minimum of 45 comparisons, whereas a cohort of 100 would require 4950. Thus, the number of comparisons required can quickly become unmanageable. However, studies have shown that a reliable ranking can be achieved with fewer iterations if the process is optimised (Potter et al., 2017; Verhavert et al., 2019). This process is referred to as Adaptive Comparative Judgement (Pollitt, 2012) and is achieved with software that calculates the statistical probabilities of a submission being better or worse than others, and optimises the comparisons given to the user to provide the most reliable ranking with the fewest number of comparisons (Hardy et al., 2016; Potter et al., 2017).

Previous studies have demonstrated the ability of adaptive comparative judgement to help standardise marking across academic teams (Barber, 2018), reduce the total amount of time required for marking (Barber, 2018), and facilitate improved student outcomes when used in a formative manner (Bartholomew et al., 2020; Potter et al., 2017). In addition, benefits can also be seen when adaptive comparative judgement is completed by peers (Barber, 2018; Hardy et al., 2016; Potter et al., 2017; Seery et al., 2012), where the benefits of comparative judgement and peer feedback are combined. Nicol (2020) argues that students are constantly using comparisons to action any feedback they receive; "how does this feedback compare to previous work?", and thus combining adaptive comparative judgement with peer assessment allows students to perform even more critical comparisons, and hence action improvements accordingly.

Aims of the study:

The research aims of this investigation all relate to the question of 'can adaptive comparative judgement be applied in the University of Salford context?'. This is because the benefits of adaptive comparative judgement are clearly available in the literature, but the individual nature of each institution (and even each programme in Higher Education) means that results in one scenario may not be transferable to another. The studies available in the literature often focus on large cohorts of students, where peer comparisons are often used to reduce the marking workload of staff. The cohort sizes on some of the programmes at Salford are significantly smaller than those discussed in the literature, and thus the need for peer comparison from a workload perspective is not present. However, there are other benefits to peer feedback (as described above) that could still be achieved with small cohorts. Therefore, the work published here aims to answer the following questions specifically in relation to the undergraduate pharmaceutical science degree at the University of Salford:

- Question 1: Are students willing and able to effectively engage with the adaptive comparative judgement process when part of a small cohort?
- Question 2: Does adaptive comparative judgement offer advantages over traditional forms of formative feedback (from the student perspective)?

Methods

One cohort of final year undergraduate students were asked to participate in this study. The group comprised 19 students on the Pharmaceutical Science BSc (Hons) programme, studying the 'Frontiers in Medicine Design' module. The students of this module were chosen because I designed and ran the module, and thus I had complete control over the curriculum and assessment design. This allowed me to implement the feedback schedule into the course without the need for consultation with other academic staff. All students were asked to consent to taking part in the study during an in person class at the start of the module. They were asked to read a participant information sheet and then sign a corresponding consent form. Before signing, they were given the opportunity to ask questions directly, but also post questions anonymously to a projected chat board run through Mentimeter (https://mentimeter.com). It was also explained that if they did not want to take part in the study, they would still need to take part in the submission of drafts and peer comparisons as this was an integral part of the learning strategy on the course. However, their data would be removed from the analysis.

In total, 18 students agreed to take part in the study and hence the outcomes of their work are presented below. The one student who did not consent was absent for key sessions on the course, and thus was a passive rather than active decliner. The focus here was on final year students as it was presumed these would be the most able to identify a 'good' piece of work through the comparative process because they have the greatest experience of submitting and receiving feedback on assessments in Higher Education. In addition, they would also be in the strongest position to compare

adaptive comparative judgement with the full breadth of formative feedback strategies that they had experienced at university. However, it is noted, that if successful, the formative feedback could be more beneficial to students if they experienced it earlier in their university career.

The adaptive comparative judgement process was applied to four pieces of work that students submitted while studying for the module. These four pieces of work (sub-components) formed the basis of the summative coursework assessment which they submitted at the end of the module. This strategy is described as an 'Integrated multi-component' assignment by Gibbs (2010) and aims to ensure that students act upon the feedback they are given. Students submitted each piece of work through the standard online submission process used at the University of Salford to the timeline described in table 1. There was specific time allocated for peer comparisons, release of results, and time for students to action any feedback they received (the final submission of coursework was the 13th May). Of note is the first round of comparisons which took place on a single day when all students were shown how the process works in a dedicated computer class. The rest of the comparisons were scheduled over a week as students were expected to complete these in their own time.

Table 1. Timelines for draft submissions and peer comparisons.

Piece of work	Draft Submission	Peer review	Feedback release
Sub-component 1 (Database)	18th Feb	24th Feb – In class	25th Feb
Sub-component 2 (Model)	18th March	21st – 25th March	28th March
Sub-component 3 (Lab Report)	1st April	25th – 29th April	3rd May
Sub-component 4 (Regulatory assessment)	29th April	2nd – 6th May	9th May

The adaptive comparative judgements were carried out with the RM Compare software (https://compare.rm.com/). The software runs as an online system that can be accessed from any internet enabled device with the use of a username and password. Student work was added to the system by the module leader for each draft submission, ensuring that all identifying text (i.e. names and students IDs) were removed from the work. After the work had been uploaded, the judges were then defined by generating a specified number of usernames and passwords that were then distributed to the students. Students accessed the software through these automatically generated usernames and passwords and kept the same ones throughout the whole process. The anonymity of the owners of the work and judges was maintained throughout.

When logged into the system, the students were presented with two pieces of work to compare and a holistic statement on which to judge them. Students are asked to compare using a holistic statement rather than a mark scheme as it makes the comparisons simpler, and it has been shown that students prefer holistic statements to analytical criteria (Gibbs, 2010). The students chose which of the two pieces of work were best described by the holistic statement, in addition to providing individual comments on each piece and also providing a brief comment on why they had made that decision (both types of comment were optional).

The holistic statements used for each comparison are given below:

- Sub-component 1: *A well organised table of data including at least 50 compounds covering a range of sub-structures and chemical properties.*
- Sub-component 2: *A clearly annotated model with all relevant nodes and statistical analyses described.*
- Sub-component 3: *A well structured report with clear links between the results obtained and the discussion in which the potential limitations of the investigation are described.*
- Sub-component 4: *Correct regulatory assessments are provided for all compounds with clear explanations of the resulting required actions.*

In summary, the holistic statements aimed to capture the essence of what an excellent piece of work would look like and were thus based on the 'excellent' level grade descriptor from the marking rubric which was used on the final summative submission.

After each round of peer review, students were sent a summary of the comments that had been written about their work in addition to the rank of their work compared to everyone else's (e.g. 4th out of 18 submissions) by the module leader. However, it was emphasised that the ranks should be considered only in a formative sense and were not in any way related to overall grades; it was possible for all the cohort to receive the highest available marks if their work met the relevant criteria standards described in the mark scheme. The 'Implications for practice' described by Hardy et al. (2016) were thus considered in this methodological design by (i) integrating the feedback timelines into the course structure, (ii) ensuring there was 'buy in' from all staff involved, and (iii) focussing student attention on the feedback they received rather than their rank of work.

At the end of the course, after submission of their final piece of coursework, but before grades were released, students were asked to complete a questionnaire comprising 10 questions administered through Microsoft Forms (questions available as supplementary information). The purpose of the questionnaire was to gather qualitive data on student perceptions of how adaptive comparative judgement compared to other formative feedback strategies they had experienced while at university. Additional qualitive data was also taken from the final assignments, where students were asked to write a 600-word reflection on their use of formative feedback to improve their work over time. Both sources of information were subjected to a thematic analysis (Clarke & Braun, 2017) to identify and interpret the key opinions of the students.

Ethical approval for this study was granted through the internal review process for staff research projects conducted in the School of Science, Engineering and Environment at the University of Salford. The application focussed on the how students were recruited, the types of information to be collected, and how data were stored. It was made clear to all participants that participation was voluntary and that not participating in the study would have no effect on their ability to complete their studies. Students were also made aware of their right to withdraw from the study at any time. No personal or identifiable data were collected, and all questionnaire responses were collected anonymously using an online Microsoft Form. The judging process was also anonymous where students were provided with auto-generated log-on information rather than creating an account using their personal data.

Results

In answer to the first research question proposed in the aims, "*Are students willing and able to effectively engage with the adaptive comparative judgement process when part of a small cohort?*", table 2 demonstrates that although initial engagement was high, this quickly disappeared towards the end of the study. Round 1 of the judging was completed in a timetabled session, where detailed instructions on how to complete the task were given, whereas for rounds 2-4, students were expected to complete these in their own time. Of note is that in addition to the number of engaged judges

steeply declining through the duration of the course, so did the number of (draft) pieces of work that were submitted. This suggests that students chose to prioritise other work, even when they knew they would get useful feedback on any drafts they submitted.

Table 2. Student engagement with both the draft submission and judging process.

Round	Number of drafts submitted	Number of engaged judges
One	18	18
Two	15	12
Three	12	4
Four	5	4

Thus to summarise the results in answer to question 1: students are able to engage with the adaptive comparative judgement process where support is provided, and are willing to do so when their competing priorities allow. Answers to the second research question, "*Does adaptive comparative judgement offer advantages over traditional forms of formative feedback (from the student perspective)?*", can be gleaned from the student questionnaire results and student reflections.

The questionnaire received five responses and the main outcomes are summarised in figure 1. As the response rate was very low, definitive conclusions cannot be drawn from these data. However, it is reassuring to see that the students who did fully engage found the feedback they received useful and there was a general preference for adaptive comparative judgement over other feedback strategies that the students had experienced. To improve the engagement with surveys such as this one in future studies, it may be better to use a paper-based exercise which has been shown to have higher return rates in other studies (Czaplinski & Fielding, 2020). However, considerations would have to made on how to assure anonymity is maintained and to ensure that the students do not feel under pressure to provide any specific response given the power dynamic between academics and students.
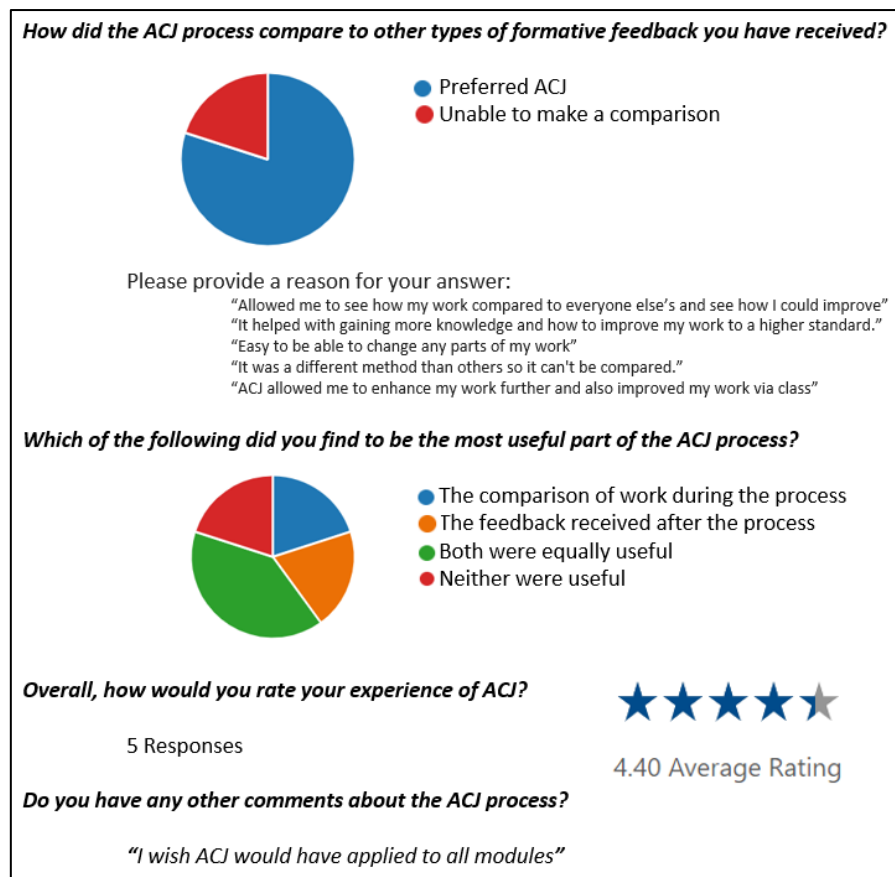
Figure 1. Key questions and answers relating to student engagement with adaptive comparative judgement.

In addition to the qualitive data from the questionnaires, students also provided useful insights through their final submission of work where they had to write a 600 word reflection on how they had used the feedback from the adaptive comparative judgement process to improve their work. It was possible to gather the opinions of the whole cohort from these assignments as it was a compulsory aspect of the work, and thus more data was gathered from these reflections than the surveys described above. The key themes from those pieces of writing a shown in Table 3.

Table 3. Key themes from student reflections on using feedback from adaptive comparative judgement process to improve their work.

Key themes	Number of assignments that discuss theme (n=18)
Discussion of rank of work	7
Improving work based on comparison to peers	10
Improving work based on direct feedback	9
Positive opinion of feedback strategies	13
Negative opinion of feedback strategies	0
Positive opinion of software used for comparison	2
Negative opinion of software used for comparison	0
Enjoyment/benefit from leaving comments on other's work	3
Desire to have adaptive comparative judgement in other modules	4
No improvement due to no direct feedback	4

Other key messages that came through from the reflections were comparisons aiding self-reflection and improvements in engagement. The direct quotes from student work are given below:
Comparisons aiding self-reflection:

> "I was able to put myself in the position of 'marking' the work of another which aided me with better insights of what to 'look' for in one's work"
> "Now, when I am completing any tasks, I always find myself questioning my own work to see what ways I can improve"
> "The feedback … has been a great advantage as I now will take this on to the next part of my education by allowing someone to read over my work before submitting"

Improvements in engagement:

> "I felt way more confident in submitting reports ….. allowed me to be more engaged within this module and simply enjoy what I was learning and submitting"
> "…it helped me in completing the work for each class on time so that I could get formative feedback"

> *"…I had asked my peers by looking at the workflow that I had created if there were any improvements I could make before submitting"*

In addition the students also provided insightful comments around the wider scope of comparative feedback and how they felt it related not only to their time at university, but also wider career as a scientist:

> *"This process is also very similar to that of the peer review process in which other scientists read your work and provide improvements and criticisms, completing this process in a simpler situation has also prepared me for taking part in this in the future."*
>
> *"Also, to note our lectures have to individually mark each submission on their own and as expected to give detailed feedback is simply unachievable with the demand lecturers have within their role so being able to get feedback from peers was very helpful."*
>
> *"Personally, I would have loved to have this adaptive comparative judgement process throughout my whole university experience as I feel like I could have gained lots of valuable feedback and then sort out my work accordingly and achieve higher marks."*

Thus in summary to answer question 2: student responses suggest they had an overall positive experience of adaptive comparative judgement, and could see the direct benefit to their studies but also in developing skills such as self-efficacy that will be useful to them throughout their career.

Discussion:

The motivations for the work presented here were to assess if the success of adaptive comparative judgement demonstrated in the literature with large cohorts could be replicated at the University of Salford with a much smaller group. Previously published work demonstrated that with cohorts of greater than 100 (Hardy et al., 2016; Ibarra-Sáiz et al., 2020; Jones & Alcock, 2014; Seery et al., 2012), the adaptive comparative judgment process can be easily facilitated because even with a dropout rate of 10%-20% there are still enough judges to allow the process to continue. However, with a smaller cohort, even a small level of drop off can be disastrous. Indeed, the results of this study demonstrate that only 25% of the cohort dedicated and prioritised the adaptive comparative process through the duration of the study. A key development of this work will be to find strategies of improving this engagement over time. One possibility would be to include more compulsory timetabled sessions, where students can work on the comparison process in a more collaborative manner. The ability to collaborate could improve students' motivations to complete the work through the formation of informal social contracts as they would feel the pressure to co-operate with the group and complete the judging process as everyone is involved. This is in contrast to completing the judging process on their own, asynchronously to everyone else, which could be an isolating and demoralising process.

For the minority of students that did complete all the comparisons in this study, they still felt the benefits of the process, even though the number or comparisons for the final subcomponents was small. The ability to see the work of others, no matter how small the sample, provided the students with enough guidance on where their work sat within the class and thus how their work compared in terms of quality. This tallies with other studies that have shown that students benefit from seeing other work, even when only a very small number of example pieces of work are provided (Kean, 2012). This observation and internalisation of quality is key for students understanding the standard of work required and is difficult to get across via other means (Bartholomew et al., 2020; Kean, 2012). Thus the benefits of adaptive comparative judgement can still be found in small cohorts.

When asking students to compare adaptive comparative judgement with other formative feedback students had previously received, the results presented here show that they felt it was very different to anything they had experienced during their three years at university. This suggests that they have

limited to access to any peer feedback processes and when examining module descriptions, informal peer feedback is frequently mentioned, but the formal process as described here is missing. Instead, students are more likely to get formative feedback on tasks specifically designed as a formative exercise from an academic member of staff. This is concerning given the benefits to learning that have been demonstrated through allowing students to mark work (Davies, 2004) and also the development of self-reflection through peer-marking (Gibbs & Simpson, 2005). In this study, students saw the advantages not only in terms of their own development, but also logistically in terms of being able to access detailed feedback without having to wait for a lecturer who may have hundreds of assignments to review. They also, insightfully, saw that peer review was a key skill they will need if they move forward into a career as a research scientist and thus saw the benefits in terms of skill development rather than just immediate work improvement.

A significant outcome of this work will be my continued use of adaptive comparative judgement across the modules on which I teach because it has proved to be an effective learning experience for my students. I also want to work with my programme team to integrate the process across all levels of the pharmaceutical science undergraduate programme. This is based on the feedback I received from students expressing their wish that they had experienced the process before their final semester. In addition to these plans, I also want to share the practice across the institution and with this in mind I presented the outcomes at the University of Salford Learning and Teaching conference in September 2022. I wanted to share the practice across the institution to find other academics who would be interested in using adaptive comparative judgement in their modules and programmes. This has led to discussions across departments on how the process can be implemented into modules, and thus providing the opportunity for more students to have access to an improved learning experience.

The limitations of this small study relate to the inability to assess if student work significantly improved through the feedback process. Although students have been able to articulate that they feel their work has improved over time, without an initial and then final assessment of quality, it is not possible to state if this feeling translated into improved work. In addition, the validity of the student judgements was not assessed; I did not examine if what the students thought was the highest quality piece of work correlated with my own judgement. This is the same as the limitations discussed by Bartholomew et al. (2020) and there the authors suggested that inserting high quality examples into the comparison would be an easy method of assessing the students' ability to spot what the lecturer considered a high-quality piece of work. However, I would argue that when the ranking of work does not lead to grades (as in this study) this 'validity' of judgements is less of a concern. When the work was finally marked for the students' summative grades, I did not look at how that work had been raked through the peer-feedback process. Thus, the ranks did not directly affect their grades. However, I do agree that there may be an indirect problem in that the students attempted to improve their work to be more like those which they perceived as better, without an indication over whether their judgement on what was the best piece of work was sound. Finally, the last limitation of the work presented here is that the questionnaire did not delve into why students stopped participating in the process, and thus I only have my hypotheses of why this might be so (i.e., competing priorities of final year dissertation). Without this information, it is impossible to suggest ways of improving engagement that are not complete speculation. I hope to investigate this problem further in future studies.

In conclusion, the results I have presented show that running comparative judgement as a method for acquiring peer feedback is possible with a small cohort of students. However, engagement of students can be variable and is improved with dedicated timetabled sessions. Students valued both the process and outcome of comparative feedback and felt that they were able to improve their self-efficacy by engaging. However, further work is required to assess whether this translated to an improvement in the standard of work students produced.

References:
Barber, J. (2018) 'Five Go Marking an Exam Question: The Use of Adaptive Comparative Judgement to Manage Subjective Bias', *Practitioner Research in Higher Education*, 11(1), pp. 94-100. https://search.proquest.com/docview/2101387668?accountid=8058

Bartholomew, S. R., Mentzer, N., Jones, M., Sherman, D., & Baniya, S. (2020) 'Learning by evaluating (LbE) through adaptive comparative judgment', *International Journal of Technology and Design Education*, *32*, pp. 1191-1205. https://doi.org/10.1007/s10798-020-09639-1

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London: Routledge.

Carless, D. and Boud, D. (2018) The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), pp. 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Clarke, V. and Braun, V. (2017) 'Thematic analysis. *The Journal of Positive Psychology'*, 12(3), pp. 297-298. https://doi.org/10.1080/17439760.2016.1262613

Czaplinski, I. and Fielding, A. L. (2020) 'Developing a contextualised blended learning framework to enhance medical physics student learning and engagement', *Physica Medica*, 72, pp. 22-29. https://doi.org/10.1016/j.ejmp.2020.03.010

Davies, P. (2004) 'Don't write, just mark: The validity of assessing student ability via their computerized peer-marking of an essay rather than their creation of an essay', *ALT-J*, 12(3), pp. 261-277. https://doi.org/10.1080/0968776042000259573

Evans, C. (2013) 'Making sense of assessment feedback in higher education. *Review of educational research'*, 83(1), pp. 70-120. https://doi.org/10.3102/0034654312474350

Evans, C. (2021) *A self-regulatory approach to assessment in higher education*. Cardiff University. Available at: https://www.researchgate.net/publication/357172330_A_self-regulatory_approach_to_assessment_in_higher_education (Accessed: October 2022).

Gibbs, G. (2010) *Using assessment to support student learning*. Leeds Met Press. Available at: https://eprints.leedsbeckett.ac.uk/id/eprint/2835/1/100317_36641_Formative_Assessment3Blue_WEB.pdf (Accessed: October 2022).

Gibbs, G. and Simpson, C. (2005) 'Conditions under which assessment supports students' learning', *Learning and teaching in higher education,* (1), pp. 3-31. https://eprints.glos.ac.uk/id/eprint/3609

Hardy, J., Galloway, R., Rhind, S., McBride, K., Hughes, K. and Donnelly, R. (2016) *Ask, answer, assess*. Available at: https://www.heacademy.ac.uk/system/files/ask_answer_assess.pdf (Accessed: October 2022).

Haughney, K., Wakeman, S. and Hart, L. (2020) 'Quality of Feedback in Higher Education: A Review of Literature', *Education Sciences*, 10(3), pp. 60. https://www.mdpi.com/2227-7102/10/3/60

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G. and Boud, D. (2020) 'Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement', *Higher Education*, 80(1), pp. 137-156. https://doi.org/10.1007/s10734-019-00469-2

Jones, I. and Alcock, L. (2014) 'Peer assessment without assessment criteria', *Studies in Higher Education'*, 39(10), pp. 1774-1787. https://doi.org/10.1080/03075079.2013.821974

Kean, J. (2012) 'Show AND tell: using peer assessment and exemplars to help students understand quality in assessment', *Practitioner Research in Higher Education*, 6(2), pp. 83-94. https://ojs.cumbria.ac.uk/index.php/prhe/article/view/126

McConlogue, T. (2015) 'Making judgements: Investigating the process of composing and receiving peer feedback', *Studies in Higher Education*, 40(9), pp. 1495-1506. https://doi.org/10.1080/03075079.2013.868878

Nicol, D. (2020) 'The power of internal feedback: exploiting natural comparison processes', *Assessment & Evaluation in Higher Education*, 46(5), pp. 1-23. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D. and McCallum, S. (2021) 'Making internal feedback explicit: exploiting the multiple comparisons that occur during peer review', *Assessment & Evaluation in Higher Education*, 47(3), pp. 1-19. https://doi.org/10.1080/02602938.2021.1924620

Nicol, D., Thomson, A. and Breslin, C. (2014) 'Rethinking feedback practices in higher education: a peer review perspective', *Assessment & Evaluation in Higher Education*, *39*(1), pp. 102-122. https://doi.org/10.1080/02602938.2013.795518

Office for Students (2022) *Securing student success: Regulatory framework for higher education in England*. Availble at https://www.officeforstudents.org.uk/publications/securing-student-success-regulatory-framework-for-higher-education-in-england/ (Accessed: July 2023).

Pollitt, A. (2012) 'The method of Adaptive Comparative Judgement', *Assessment in Education: Principles, Policy & Practice*, 19(3), pp. 281-300. https://doi.org/10.1080/0969594X.2012.665354

Potter, T., Englund, L., Charbonneau, J., MacLean, M. T., Newell, J. and Roll, I. (2017) 'ComPAIR: A New Online Tool Using Adaptive Comparative Judgement to Support Learning with Peer Feedback', *Teaching & Learning Inquiry*, 5(2), pp. 89-113. https://doi.org/10.20343/teachlearninqu.5.2.8

Sadler, D. R. (2005) 'Interpretations of criteria-based assessment and grading in higher education', *Assessment & Evaluation in Higher Education*, 30(2), pp. 175-194. https://doi.org/10.1080/0260293042000264262

Seery, N., Canty, D. and Phelan, P. (2012) 'The validity and value of peer assessment using adaptive comparative judgement in design driven practical education', *International Journal of Technology and Design Education*, 22(2), pp. 205-226. https://doi.org/10.1007/s10798-011-9194-0

Simonsmeier, B. A., Peiffer, H., Flaig, M. and Schneider, M. (2020) 'Peer Feedback Improves Students' Academic Self-Concept in Higher Education', *Research in Higher Education*, 61(6), pp. 706-724. https://doi.org/10.1007/s11162-020-09591-y

Thompson, D. and Meer, N. (2021) 'Blurring the boundaries of formative and summative assessment for impact on learning', *Practitioner Research in Higher Education*, 14(1), pp. 28-40. https://ojs.cumbria.ac.uk/index.php/prhe/article/view/676

Thurstone, L. L. (1927) 'Law of comparative judgment', *Psychological review*, 34(4), pp. 273. https://psycnet.apa.org/doi/10.1037/h0070288

Vattøy, K.-D., Gamlem, S. M. and Rogne, W. M. (2020) 'Examining students' feedback engagement and assessment experiences: a mixed study', *Studies in Higher Education*, 46(11), pp. 1-13. https://doi.org/10.1080/03075079.2020.1723523

Venables, A. and Summit, R. (2003) 'Enhancing Scientific Essay Writing Using Peer Assessment: Journal of the Association for Programmed Learning', *Innovations in Education and Teaching International*, 40(3), pp. 281-290. https://doi.org/10.1080/1470329032000103816

Verhavert, S., Bouwer, R., Donche, V. and De Maeyer, S. (2019) 'A meta-analysis on the reliability of comparative judgement', *Assessment in Education*, 26(5), pp. 541-562. https://doi.org/http://dx.doi.org/10.1080/0969594X.2019.1602027

Wheadon, C., Barmby, P., Christodoulou, D. and Henderson, B. (2020) 'A comparative judgement approach to the large-scale assessment of primary writing in England', *Assessment in Education: Principles, Policy & Practice*, 27(1), pp. 46-64. https://doi.org/10.1080/0969594X.2019.1700212

The constrained formative feedback in doctoral examiner reports

Deborah Chetcuti, Joseph Cacciottolo, Nicholas Vella
University of Malta

Abstract

The award of a PhD degree is based on a rigorous examination process which is unique in that examiners can recommend that the degree is awarded subject to minor or major amendments to the thesis. In their examiner reports, examiners also include formative feedback that is intended to help the doctoral researchers improve their work. Using data from examiner reports from two faculties at the University of Malta (Arts and Science) for the years 2017-2018, the study looks at the formative feedback comments provided by doctoral examiners. Results suggest that the feedback comments provided can be described as editorial, instructional and reflective, with the most common type of feedback comments being editorial and instructional. This indicates that the focus of doctoral examiner reports is on giving advice to doctoral researchers so that it reaches the expected doctoral standards. This study points to: the need for a rethinking of the way in which feedback is provided in doctoral examiner reports so that the focus is shifted to reflective comments that lead to in-depth learning; and for more professional development for doctoral examiners to that they can make this shift in their provision of formative feedback.

Keywords:
Doctoral examination; formative feedback; doctoral examiner reports.

Introduction

The award of a PhD degree is an important academic milestone for doctoral researchers, providing them with international recognition as independent scholars in their field, access to an academic career and certification that they have acquired skills and competencies to work in industry and other professions (Boud and Lee, 2009; Nerad, 2020). In most universities, the first step in being awarded a doctorate, is the rigorous examination of the doctoral thesis which is the product of a research journey in a specific area of expertise that extends over a number of years, followed in most universities by an oral examination or viva-voce (Trafford and Leshem, 2008). This examination is carried out by a board of examiners set up by the awarding institution and usually includes both internal as well as external examiners. The examiners are required to make a recommendation as to whether the PhD degree should be awarded or not, depending on whether the doctoral thesis has met a set of explicit criteria. Within European universities, these criteria are mostly adopted from the Framework for Qualifications of the European Education Area (EHEA) and are based on the expectation that doctoral researchers: make an original contribution to knowledge in their research field which can be published and have developed skills that allow them to conduct research independently (Denicolo, Duke, and Reeves, 2020).

For purposes of accountability and transparency, examiners are required to record this recommendation in a written individual report (Trafford and Leshem, 2008). In most universities, the recommendation is not final but subject to minor or major modifications (Kumar and Stracke, 2011). In their report examiners are therefore expected to include summative feedback or information regarding the quality of the doctoral work against expected criteria and standards and when modifications are required formative feedback about what still needs to be done to the thesis so that

Citation

it meets the doctoral standards (Dally et al., 2019). In universities, that include a viva as part of the doctoral examination these individual reports are used by examiners to confirm their initial impressions of the work and to prepare their questions for the viva (Pearce, 2005). Following the viva, the individual reports are used to write a final negotiated report that includes both the recommendation for the award of the PhD degree and feedback regarding changes that is passed on to doctoral candidates and their supervisors (Denicolo, Duke, and Reeves, 2020). The individual examiner reports are therefore important documents that are expected to follow guidelines and protocol of the awarding institutions and can be reviewed by university panels and regulatory bodies, especially in cases of failure (Erwee and Perry, 2018). However, despite the attempts of universities to standardise doctoral examiner reports, in many cases, they are usually individualistic documents that are written in different styles and formats and draw on a mixture of official expectations of standards and the personal views and experiences of the examiners (Mullins and Kiley, 2002).

As examiners of doctoral work at the University of Malta we were aware of the huge responsibility in writing examiner reports. In a previous study (Chetcuti, Cacciottolo, and Vella, 2022) we explored the summative component of the doctoral examiner report, and how doctoral examiners used criteria to make a recommendation for the award of the PhD degree. In the current study we wanted to look at the type of formative feedback comments that examiners provide for doctoral researchers in their examiner reports. To try and explore this question we looked at examiner reports presented by examiners from two different faculties at the University of Malta. We are aware that the context in which the study was carried out is highly specific, and that any generalisations can only be made with due caution. However, we hoped that our study would enable us to gain a better understanding of the role of formative feedback in the doctoral examination process.

Defining formative feedback in doctoral examiner reports
The doctoral examiner report serves a dual function and addresses different audiences. The first function of the examiner report is a 'gatekeeping' one, in that examiners need to ensure that the standards of the discipline and the doctoral degree are maintained (Denicolo, Duke and Reeves, 2020). In their summative comments, examiners are accountable to the awarding institution that needs to ensure that examiners have followed institutional guidelines and assessment criteria (Erwee and Perry, 2018). In their formative comments, examiners address the doctoral candidates who are repositioned as learners. The comments are usually in the form of instructions and advice to help the doctoral candidates improve both their submitted thesis, as well as future publications (Golding et al., 2014). Research studies (Holbrook et al., 2004; Dally et al., 2020) suggest that examiners place a great deal of emphasis on the formative comments that they provide for students in their report. For doctoral examiners, the thesis is 'work-in-progress' and in their feedback they provide the necessary impetus for the doctoral candidates to continue to develop as scholars (Stracke and Kumar, 2010), and build their "capacities as autonomous self-regulating learners" (Carless, 2013, p. 113). This is especially important within the context of doctoral education, since one of the main aims of doctoral education is for doctoral researchers to develop the academic and research skills that enable them to become independent scholars, researchers, and professionals (Nerad, 2020).

Types of formative feedback comments in doctoral examiner reports
A number of research studies have been carried out to investigate the type of formative feedback information provided by doctoral examiners in their reports. In their study using doctoral examiner reports from Australian universities, Holbrook et al. (2014) suggest that the comments provided by doctoral examiners in their reports are mainly prescriptive and written in an authoritative voice that provide instructions and advice that are related to the editing and formatting of the text, to methodological issues and conceptual arguments. The comments focus on correcting the product of the research, the thesis, with little emphasis on learning and reflection. Similarly, in an attempt to classify feedback comments, Kumar & Stracke (2007) suggest that feedback in doctoral examiner

reports can be referential, directive or expressive. They describe referential comments as focusing on "editorial, organisational and content matter"; directive comments that include "suggestions, questions and instructions" and expressive comments that offer "praise, criticism or an opinion" (p. 467). Kumar and Stracke (2007) suggest that the majority of the comments made by the supervisors in their study were in fact referential, and fix-it type comments that focused on the technical aspect of the thesis. However, they also point out that it was the expressive comments that invited the doctoral researchers "to reflect upon the issues raised from a different perspective, which resulted in a process of further changes, and thus, more discoveries" (p. 467).

In another study, Pryor and Crossouard (2010), use examples of feedback from a professional doctoral context to develop a classification of feedback that is enacted by moving constantly through what they describe as "a concrete-procedural-reflective-discursive-existential continuum" (p. 274). In their classification concrete/procedural comments focused on the practical aspects of the task; reflective/discursive comments encouraged deeper reflections on conceptual arguments and methodological issues in the text; and discursive/existential comments motivated the doctoral researchers to identify with their work and give voice to their own views. Again, Pryor and Crossouard (2010), state that feedback comments given to the doctoral researchers were mainly of the concrete/procedural type with discursive/existential comments being very infrequent. However, they too suggest that, when possible, examiners should provide feedback that encourages the exploration of ideas and a reformulation of learner identity.

These research studies suggest that in practice within the context of the doctoral examination, formative feedback is still viewed by examiners as the giving of advice to the students (Ajjawi & Boud, 2018; Yang & Carless, 2013). While there is the expectation for action to be taken on the feedback comments provided by examiners, research suggests that there is very little dialogue between examiners and the doctoral researchers, who are expected to engage with the feedback comments on their own (Holbrook et al., 2004). This view of feedback focuses on the educator as the information giver and the student as a passive recipient of this information (Boud & Malloy, 2013).

The study
At the University of Malta, doctoral work is examined by a board of examiners, which usually includes two examiners from the University of Malta and an external examiner, all appointed by Senate, and chosen for their expertise in the research area. The examination of the doctoral work takes place in two stages. The first stage involves the assessment of the written thesis. Using criteria outlined in the UoM PhD regulations (2008), the examiners assess whether the doctoral thesis shows evidence of: a significant and original contribution to knowledge; an extensive review of existing knowledge in the field of study; appropriate research methods and a satisfactory literary presentation. Based on this initial summative assessment examiners make a recommendation as to whether they believe that the PhD should be awarded, subject to the outcome of the second stage of the examination, the oral defense or viva.

In most cases, the recommendation for the award of the PhD degree is made subject to minor or major modifications and in this case, the examiners also include in their individual report a list of feedback comments on specific amendments or modifications required to improve the work. This feedback is then discussed with the doctoral researchers during the viva. Following the viva, the chairperson of the board of examiners writes a joint report that is negotiated and is usually a collation of comments from the individual examiner reports and a list of requested changes. Doctoral researchers are expected to make these amendments prior to final submission of the doctoral thesis and the award of the PhD degree (see PhD regulations, University of Malta, 2008).

Research methods: the data and data analysis
The data that we used in our study was obtained from the examiner reports presented to two faculties at the University of Malta, the Faculty of Arts and the Faculty of Science. These two faculties were chosen since they are two of the largest and more traditional faculties, and represent two different areas of knowledge. Examiner reports were chosen as the main source of data since they provided detailed information and valuable insights about examiners' views and expectations of the quality of work at doctoral level and included feedback for the doctoral researchers (Holbrook et al, 2004). In a previous study, we looked at the summative aspect of the doctoral examiner reports, and the explicit and implicit criteria used by examiners to award a PhD degree (Chetcuti, Cacciotolo and Vella, 2022). In the current paper, we limit our focus to the formative feedback comments, since gaining a better understanding of formative feedback in the doctoral examination process was the main aim of our research.

We analysed a total of 95 individual examiner reports (50 from the Faculty of Arts and 45 from the Faculty of Science). These were presented to the University of Malta during the years 2017 and 2018. The reports were written by both internal examiners from the University of Malta as well as external examiners from various international universities. Since the examiner reports are not readily available to third parties due to data protection policies, permission to use the reports for our study was obtained from the ethics committee of the University of Malta (UREC). Throughout, the study, the examiner reports were anonymised and handled with strict confidentiality at all times. To ensure this confidentiality, prior to the analysis examiners were identified by a code name such as *Ex1 Arts UoM* or *Ex1 Sci Ext,* to indicate the faculty, and whether the examiner was an internal or external examiner. Care was also taken to ensure that in the subsequent write-up of the study quotations were screened and, in some cases, edited so that no individual, whether doctoral researcher or examiner, could be identified.

The examiner reports were analysed using a thematic analysis (Boyzatzis, 1998; Braun & Clarke, 2006). This involved reading and re-reading the examiner reports in order to "derive themes and concepts from the raw data through detailed reading and analysis" (Sharmini et al., 2015, p. 92). We focused on the formative feedback comments and used an inductive approach to our analysis. First, we identified feedback comments in the examiner reports as being formative if they were written in terms of specific advice or guidance for doctoral researchers to improve their work. We then looked at the formative feedback comments and classified them according to the specific purpose and tasks that the examiners wanted the doctoral candidates to address.

The Results: Feedback in doctoral examiner reports
The first thing that we observed when analysing the individual examiner reports was that out of the 22 doctoral candidates (11 from the Faculty of Arts and 11 from the Faculty of Science), who submitted their thesis for examination between 2017 and 2018, all were awarded the PhD degree, except for one student from the Faculty of Arts. This high success rate indicates that like what has been reported in international studies (Holbrook et al., 2014), doctoral examiners go into the doctoral examination, expecting the doctoral candidates to be awarded a Pass. However, in most cases, this Pass was awarded subject to minor modifications (for 17 out of the 22 doctoral candidates) or requiring major modifications (for 3 out of the 22 doctoral candidates). Only one doctoral candidate from the Faculty of Science was awarded the PhD degree without requiring any changes. Minor changes are usually considered to be editorial changes, or simple clarification of ideas, and are expected to be carried out within three months from the viva. Major changes involve significant re-writing and further development of conceptual ideas, that are expected to be carried out within six months from the viva. Since most of the examiners requested changes before the PhD could be awarded, most of them, 47 out of the 50 examiners (94%) in the Faculty of Arts and 41 out of the 45 examiners (91%) in the Faculty

of Science, included formative feedback comments in their report. These feedback comments can be described as formative since as stated by the examiners themselves, they were intended to help the doctoral candidates: "to improve the final version of the dissertation" (*Ex1 Sci UoM*); and lead to learning which would help the doctoral candidates "with writing up articles based on their study for publication in scientific journals" (*Ex5 Sci UoM*).

Our initial analysis of the feedback comments, suggested that the main purposes of the formative feedback comments could fit into three distinct categories that included:

1. **Editorial comments:** where the examiners give clear instructions about the editing of the main text such as spelling, grammar and organisation of chapters.
2. **Direct instructions:** where the examiners give clear prescriptive instructions about what the doctoral researchers need to do to improve their work.
3. **Reflective questions:** where the examiners ask questions that are intended to motivate the doctoral researchers to explore alternatives, and reflect in greater depth on the implications of their research work.

These three types of comments are captured by an examiner who in his report states that the changes he is asking for are:

> ...mainly editorial in nature. Some deal with methodological and ethical issues and require the inclusion of short paragraphs to the work. Some of the changes are in the form of questions intended for the candidate to reflect on her work...   (Ex35 Arts Ext).

The most common type of comments made by examiners in both faculties (Figure 1) were in fact, editorial comments to help the doctoral candidates clean up their work before final submission. Examiners in both faculties also gave clear instructions and made suggestions regarding various aspects of the research work such as the theoretical framework, methodology, results and discussions and conclusions. Similar to the comments made by examiners in the study by Kumar and Stracke (2007), these comments were: fix-it type comments that were mainly procedural. The least common form of feedback provided by the examiners were reflective questions, that encourage exploration of ideas and lead to learning. Although in theory and as reported in various research studies (Kumar and Stracke, 2011; Pryor and Crossouard, 2008; Dally et al., 2019) reflective questions are the most important form of feedback leading to learning and the development of an identity as a researcher, in practice reflective questions were less frequent in the examiner reports analysed in the study.
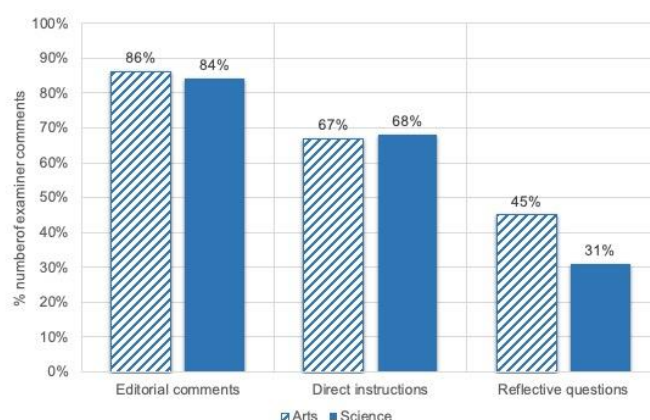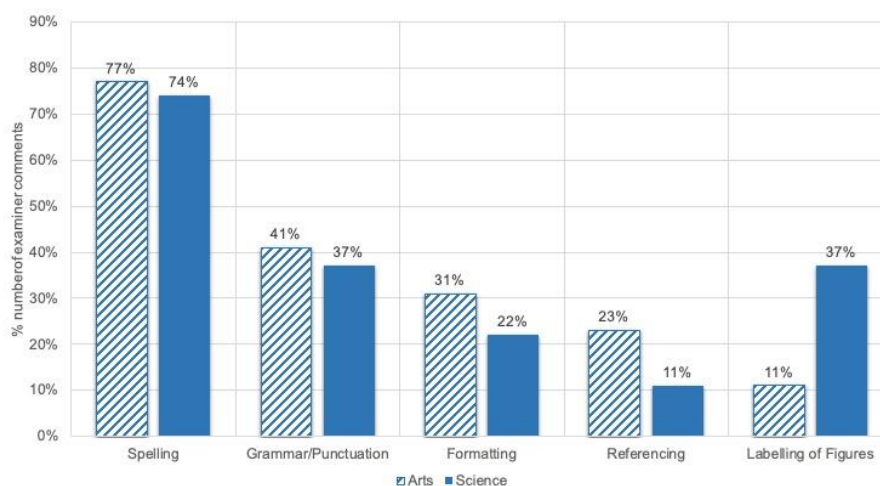

Figure 1, Categories of feedback comments.
Note: Percentage number of comments were calculated from the total number of examiner reports (N=50 for Faculty of Arts and N=45 for Faculty of Science). Examiners made more than one type of comment.

These categories consistent with existing theoretical frameworks and are very similar to the types of feedback comments outlined in previous studies such as those of Kumar and Stracke (2007) and Pryor and Crossouard (2008). On further analysis a few subtle differences emerged mainly in the category of direct instructions. While direct instructions are described in previous studies (Kumar and Stracke, 2007) as being mainly task-oriented and dealing with technical issues, in our analysis of the examiner reports, we also identified direct instructions which challenged the doctoral researchers and led to critical thinking.

Editorial comments

In the examiners' reports that we analysed, editorial comments were the most common type of feedback comments provided by the examiners (86% of examiners in the Faculty of Arts and 84% of examiners in the Faculty of Science). Editorial comments were either listed at the end of the individual examiner report, or in some cases they were in the form of "corrections, deletions and suggestions throughout the hard copy of the dissertation" (*Ex 19 Sci Ext)*, which was passed on to the doctoral candidates for their consideration. The examiners' reports (Figure 2) included editorial comments that referred to: language (mainly typographical and grammatical errors); formatting of the text (such as paragraph alignment and labelling of figures); and incorrect referencing (both in text citations and reference lists).

Figure 2. Types of editorial comments.

Note: Percentage number of comments were calculated from the total number of editorial comments made by examiners (N=50 in Faculty of Arts; N=45 in Faculty of Science).

As shown in the work of Kumar and Stracke (2007), the feedback in the form of editorial comments is a very low-level type of feedback, what are described by Holbrook et al. (2004), as the "correction of mechanical, typographical or referencing errors" (p. 143). The feedback is in the form of what Carless (2019) describes as "a teacher monologue about performance" (p. 56), and does not involve any complex learning (Sadler, 2010). Examples of typical comments made by examiners are shown in Table 1.

Table 1. Typical editorial comments made by examiners.

Language	The presentation is satisfactory. There are however a number of typographical errors to be corrected. There are 100 minor stylistic errors marked in the text - mainly limited to missing punctuation, missing words or letters and the odd change in font size etc. (*Ex8 Arts Ext*)
Formatting of the text	… it is better if the figure is enlarged and extended to the whole page…the figure should also be moved closer to the point of its first reference in the text…the reference in the caption of the figure is different from the corresponding one in the main text… (*Ex 21 Sci Ext)*
Referencing	There are a number of inaccuracies in the references…It is also important to note that there is no section entitled References and instead there is only one section, Bibliography. *(Ex21 Arts UoM)*

There is not much difference in the type of editorial comments provided by examiners in the Faculty of Arts and the Faculty of Science, although examiners in the Faculty of Arts commented slightly more about language and referencing, while examiners in the Faculty of Science were more concerned with presentation of figures and results. This can be related to the nature of the doctoral work where science theses focus on the presentation of numerical results usually in the form of figures and tables. Although editorial feedback comments are the most common type of corrective feedback in individual examiner reports, they are not necessarily formative in nature. Within the literature on formative feedback, feedback is described as formative, if it leads to improvement of the work and learning (Sadler, 2010; Carless and Boud, 2018). As noted by Holbrook et al. (2014), doctoral candidates can go through these lists of 'fix-it' corrections without any reflection or learning taking place.

Direct Instructions
Following editorial feedback, instructional feedback was the most common type of feedback provided by examiners in their reports. Instructional feedback is described by Holbrook et al. (2004), as formative feedback that is given when "examiners feel that there is a need to inform, instruct, question or extend a candidate's understanding of their topic or application of their methodology" (p. 113). In the examiner reports that we analysed, we observed that examiners very often gave direct instructions which as shown in Figure 3., asked for clarifications, made suggestions and gave direct instructions with regards to improving discussion and arguments in the text; appropriate referencing and presentation of figures and in some cases even made comments about rewriting of paragraphs to make the thesis more readable.
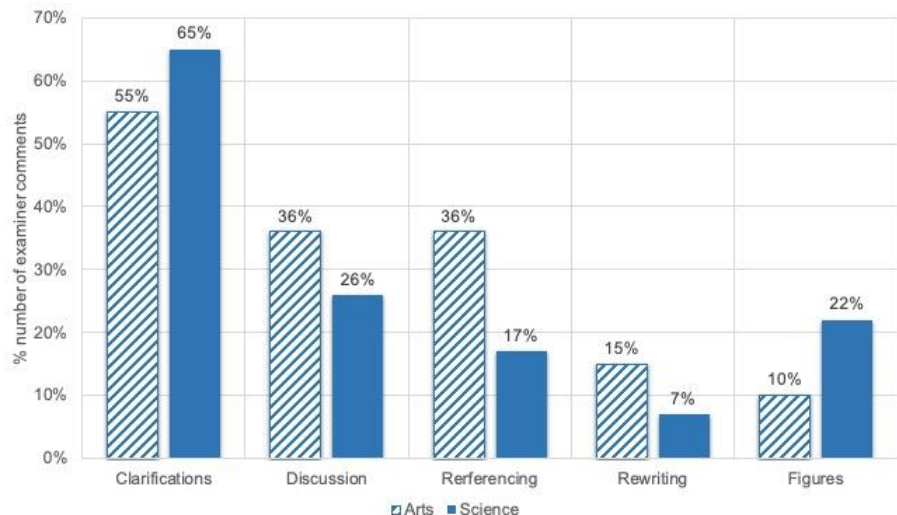
Figure 3. Types of direct instructions.

Note: Percentage number of comments were calculated from the total number of instructional comments made by examiners (N=50 in Faculty of Arts; N=45 in Faculty of Science).

What we observed in our analysis of the instructional feedback provided by the examiners in their report, is that the instructions provided by the examiners seemed to lie on a continuum. At one end of the continuum the examiners gave direct instructions that were mainly task-specific and provided the doctoral researchers with detailed suggestions on what could be done to bridge what Sadler (1989) describes as the gap between the current performance and the desired achievement. As shown in Table 2., the feedback is mainly fix-it comments and written in a prescriptive and authoritative voice. The focus is on getting the task finished in as linear a manner as possible and getting closure, that is the submission of a thesis that merits the award of a doctorate (Holbrook et al., 2014). In the Faculty of Arts the examiners gave more direct instructions in relation to the language of the written thesis and accurate referencing: "there needs to be some rewriting of the discussion to include some references from the literature to make the arguments clearer…referencing needs to follow the appropriate format" (*Ex16 Arts UoM*). The science examiners on the other hand, gave more direct instructions asking the doctoral researchers to clarify research methods used and results: "You need to clarify certain methodologies used to analyse the data and discuss sources of error and bias…include figures to highlight your results" (*Ex5 Sci UoM*).

Table 2. Task-specific Instructional feedback (Examiner 1 Science UoM).

Include justifications for claims being made.
Include references to substantiate assertions made.
Include more detail on choices regarding research methods.
Give a better analysis of conclusions.
Make editorial changes.
Explain equations used.
Include all text citations in reference list.

Yet at the same time, we observed that although the instructional comments were written in a top-down format, with examiners indicating what the doctoral candidates needed to do in specific terms, there was also a different type of instructional comment. These comments can be classified at the other end of the instructional feedback continuum. While they were written in a top-down format and very prescriptive, they were not simply technical and corrective but motivated the doctoral candidates to reflect on their work and think critically about it. Examples of such challenging instructional

feedback can be seen in Table 3. Although they challenge the doctoral candidates to engage in more depth with their academic work, these feedback comments are still considered to be mainly instructional since in most cases they reflect what Pryor and Crossouard (2010) describe as the doctoral examiner's own perspectives. While they can be formative in that they result in improvement of work and learning, they are still formulated within the disciplinary context of the examiner's expertise.

Table 3. Challenging Instructional feedback.

The candidate needs to include more recent literature and use this to elaborate on his theoretical framework. He then needs to use this literature to beef up his arguments and conclusions. (Ex 16 Arts UoM) I would like to recommend that the candidate: • Gives a better explanation of the research methods used. • Explains how equations were developed. • Discusses the variation of results. • Gives ideas for further research. (Ex 1 Sci UoM)

The use of such instructional, prescriptive feedback comments has given rise to considerable debate within the context of the literature on feedback in doctoral education (Dally et al., 2019). Some educators suggest that instructional feedback does not lead to complex learning and they advocate for more process-oriented feedback that focuses on skill development, encourages dialogue and leads to more self-regulated learning (Sadler, 2010; Carless, 2015). However, at the same time other educators believe that when the feedback provided is not highly specific, it can result in the doctoral researchers becoming confused and unsure of how to respond to feedback that did not indicate very precisely what changes needed to be carried out (Dally et al., 2019). The ideal situation would be for examiners to find a balance between task-specific and process-oriented feedback (Sadler, 2010), so that doctoral researchers understand and engage with the feedback comments in a practical manner but at the same time are encouraged to think and reflect on their work. This will enable the doctoral candidates to develop new identities as scholars and researchers (Pryor and Crossouard, 2010).

Reflective questions
The final type of feedback comment that was identified in the examiner reports that were analysed is being described as reflective questions. The difference between challenging instructional feedback comments and reflective questions is mainly in the way in which the feedback comments are presented to the doctoral candidates. While challenging instructional comments encourage critical thinking from a top-down examiner perspective, reflective questions are more open-ended. Reflective questions are exploratory and try to actively involve the doctoral candidates in a dialogue (albeit through written comments) to prompt "further engagement rather than correcting mistakes" (Pryor and Crossouard, 2010, p. 4). They allow for the development of a researcher identity that does not necessarily reflect that of the examiner (Dally et al., 2019). Although most of the examiner reports that we analysed, included mostly comments that were corrective in nature, 45% of examiners in the Faculty of Arts and 31% of examiners in the Faculty of Science also included reflective questions in their report. Some examples of reflective questions include:

> I am very surprised by this result and would like the [name of student] to think about how this result can be related to any previous studies. Would it be possible to think of comparisons?
> (Ex24 Sci Ext).

> I wish that [the candidate] thinks a bit more about how she interpreted the participants' views in her interviews? Can these views be open to a different interpretation? How did she reach her own conclusions? Can she make a stronger conclusion more in line with the work she has been doing? (Ex31 Arts Ext).

The main characteristic of reflective questions is that they do not include direct, ready-made instructions or fix-it comments, but rather this type of feedback focuses on "nurturing emerging ideas and does not attempt to be dictatorial" (Stracke and Kumar, 2010, p. 28). In posing questions that generate reflection, the examiners try to encourage a deeper engagement with the research work leading to changed understandings and the development of researchers capable of independent critical thinking (Holbrook et al., 2014).

In their reports, examiners also included indications of questions that they would like to ask the doctoral researchers in the viva. The reports that we analysed were written prior to the viva, but it was clear that the examiners in both faculties viewed the viva, as the place where they could interact with the doctoral researchers, and engage in dialogue with them about areas of the research that they thought were interesting and could be further developed by the doctoral researchers. As stated by two examiners:

> There are a number of areas I wish to explore further with the candidate in the oral viva… (Ex31 Arts Ext).

> In the viva, I will ask […] and clarity some details about how the final estimated corrections are applied to the data in practice. In the viva, I would like to discuss the relative benefits of […] as compared to other research (Ex10 Sci Ext).

The reflective questions are used by examiners to help the doctoral candidates engage more deeply with the issues and arguments raised in their research work, not only for the completion of their thesis but also as future learners, and for the writing of future publications. As stated by an examiner:

> In my report I have made a number of suggestions and raised some questions for [the candidate] with a view to assisting not only in finalising the thesis but also to help her when writing up articles for publication based on her study (Ex23 Arts Ext).

The reflective questions help to create what Golding, Sharmini and Lazarovitch (2014) describe as a learning opportunity for the doctoral researchers that contributes to the development of their identity as independent researchers. The use of reflective questions in examiner reports, suggests that examiners wish to engage in some form of dialogue with the doctoral candidates, which to some extent they can achieve during the viva. However, following the viva, in most universities, there is no further contact between the examiners and the doctoral candidates. There is very little information about how the doctoral researchers engaged with the feedback provided by the examiners, and whether any learning took place. This is an area which requires further research.

Discussion and Conclusions

This paper investigated the use of formative feedback comments in doctoral examiner reports presented by examiners in the Faculty of Arts and the Faculty of Science at the University of Malta. Although we only analysed a small sample of examiner reports, the results provided us with valuable insights into the doctoral examination process at the University of Malta. The study also helped us to identify areas which need to be researched in greater depth to make the doctoral examination a more meaningful assessment practice for both doctoral researchers and doctoral examiners.

We started out this study since as examiners of doctoral work at the University of Malta, the recommendation for the award of a PhD degree can be made subject to minor or major modifications. We wanted to gain a better understanding of the type of feedback comments provided by examiners to improve our own practice as doctoral examiners. What the results of our study have shown, is that within the context of the examiner reports presented to the Faculty of Arts and the Faculty of Science at the University of Malta and as reported in other research studies (Kumar and Stracke, 2007; Ajjawi and Boud, 2018), the feedback comments are still very prescriptive, instructional and of the fix-it type. Although some of the instructional feedback comments can be described as comments that challenge the doctoral candidates to think further about their work, the comments are written from an examiner perspective. The focus is on the transmission of information and advice by the doctoral examiner and the doctoral researcher passively receiving this information (Boud and Malloy, 2013). Some examiners in both faculties did include reflective questions, what Pryor and Crossouard (2010) describe as 'discursive/existential' (p. 266) intended to bring about changes in the researcher identity of the doctoral candidates. Some follow up of these reflective questions was carried out during the viva, but there is no further information about how doctoral researchers engage with this reflective feedback.

This raises the question of whether the feedback comments provided by doctoral examiners in their individual reports can be described as formative. According to Sadler (2010), for feedback to be considered formative, "it has to be both specific (referring, as it necessarily does to the work being appraised) and general (identifying a broader principle that could be applied to later works" (p. 538). In our view, most of the feedback comments provided by examiners is mostly specific and intended to provide the doctoral researchers with information that they can "use to improve the quality of their work" (Winstone et al., 2022, p. 224), to be awarded their doctorate, as mandated by the university. This prescriptive feedback appears to be "directed to what is known rather than what is new and focuses on closure and convergence – getting the task finished in known ways" (Holbrook et al., 2014, p. 997). The feedback is constrained formative feedback since it does lead to an improved dissertation but might not completely satisfy Sadler's (2010) condition that stipulates that formative feedback should be more general and aim to develop researchers who as described by Nerad (2020) are creative problem-solvers, critical thinkers, intellectual risk-takers, and autonomous professionals (Nerad, 2020). In our view, within the context of the doctoral examination process, this discursive formative feedback is more important, since such reflective feedback comments, encourage the doctoral researchers to explore their research from different perspectives, and help them develop an individual identity as independent researchers (Kumar and Stracke, 2011; Pryor and Crossouard, 2010). Reflective feedback comments also lead to in-depth learning that the doctoral researchers can carry forward in their future academic and professional roles. What the current study fails to identify is what happens in the timeframe between the passing on of the feedback information to the doctoral researchers and the re-submission of their final thesis. More information and research are needed about how the doctoral researchers engage with the feedback and use it as a learning tool.

We would therefore argue that the role of formative feedback in doctoral examiner reports needs to be re-examined both at the individual examiner level as well as at the institutional level. On an individual examiner level, there is the need for more professional development for examiners so that they can develop what Carless and Boud (2018) describe as 'feedback literacy', or a better understanding of the impact of the feedback that they provide in their examiner reports. On an institutional level, there needs to be a re-thinking of the way in which doctoral examiners and doctoral researchers interact, and space created in the doctoral examination process for doctoral examiners to be able to dialogue with the doctoral researchers to ensure that their feedback is being understood and leads to learning. In line, with changing purposes of doctoral education (Nerad, 2020), and the move towards a more learner-centered dialogic approach towards feedback in Higher Education (Carless, 2015), there is the need to review the way in which examiners provide formative feedback for doctoral researchers. This requires more dialogue between university administrators, doctoral

examiners, and doctoral researchers to allow for the emergence of a 'feedback literacy' that makes the doctoral examination process a truly learning experience that is the start of a new researcher's journey.

References

Ajjawi, R. and Boud, D. (2018) 'Examining the nature and effects of feedback dialogue', *Assessment and Evaluation in Higher Education*, 43(7), pp. 1106-1119.

Boud, D. and Malloy, E. (2013) 'Rethinking models of feedback for learning: the challenge of design', As*sessment and Evaluation in Higher Education,* (38)6, pp/ 698-712.

Boud, D. and Lee, A. (2009) Introduction, in Boud, D. and Lee, A. (eds). *Changing Practices of Doctoral.* London: Routledge: 1-9.

Boyatzis, R. E. (1998) *Transforming qualitative information: thematic analysis and code development*. Thousand Oaks: Sage.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology,* 3(2), pp. 77-101.

Carless, D. (2013) 'Sustainable feedback and the development of student self-evaluative capacities', in Merry, S., Price, M., Carless, D. and Taras, M. (eds.) *Reconceptualising feedback in Higher Education: developing dialogue with students.* London: Routledge. pp.113-122.

Carless, D. (2015) *Excellence in University Assessment.* London: Routledge.

Carless, D. (2019) 'Feedback loops and the longer-term: towards feedback spirals', *Assessment and Evaluation in Higher Education*, (44)5, pp. 705-714.

Carless, D. and Boud, D. (2018) 'The development of student feedback literacy: enabling uptake of feedback', As*sessment & Evaluation in Higher Education*, 43(8), pp. 1315-1325.

Chetcuti, D., Cacciottolo, J. and Vella, N. (2022) 'Achieving a PhD degree: What examiner reports tell us about the doctoral examination process', *International Journal of Higher Education Pedagogies,* 3(1), pp.1–12.

Dally, K., Holbrook, A., Lovat, T. and Budd, J. (2019) 'Examiner feedback and Australian doctoral examination processes', *The Australian Universities' review,* 61(2). pp. 31-41.

Denicolo, P., Duke, D. and Reeves, J. (2020) *Delivering Inspiring Doctoral Assessment.* London: Sage.

Erwee, R. and Perry, C. (2018) *Examination of doctoral theses: Research about the process and proposed procedures.* DOI: 10.1007/978-981-10-5249-1_4.

Golding, C., Sharmini, S. and Lazarovitch, A. (2014) 'What examiners do: what thesis students should know;, *Assessment and Evaluation in Higher Education,* 39(5), pp. 563-576.

Holbrook, A., Bourke, S., Lovat, T. and Dally, K. (2004) 'Investigating PhD thesis examination reports', *International Journal of Educational Research,* 41(2), pp. 98-120.

Holbrook, A., Bourke, S., Fairbairn, H. and Lovat, T. (2014) 'The focus and substance of formative comment provided by PhD examiners', *Studies in Higher Education,* 39(6), pp. 983-1000.

Kumar, V. and Stracke, E. (2007) 'An analysis of written feedback on a PhD thesis', *Teaching in Higher Education*, 12(4), pp. 461-470.

Kumar, V. and Stracke, E. (2011) 'Examiners' reports on theses: Feedback or assessment?', *Journal of English for Academic Purposes,* 10(4), pp. 211-222.

Mullins, G. and Kiley, M. (2002) '"It's a PhD, not a Nobel Prize": how experienced examiners assess research thesis', *Studies in Higher Education,* 27(4), pp. 369-386.

Nerad, M. (2020) 'Governmental innovation policies, globalisation, and change in doctoral education worldwide: Are doctoral programmes converging? Trends and tensions', in Cardoso, S., Tavares, O., Sin, C., and Carvalho, T. (eds.), *Structural and institutional transformations in doctoral education: Social, political and student expectations*, London: Springer Nature. pp. 43-84.

Pearce, L. (2005) *How to examine a thesis.* Maidenhead: Open University Press.

Pryor, J. and Crossouard, B. (2010) 'Challenging formative assessment: disciplinary spaces and identities', As*sessment and Evaluation in Higher Education*, 35(3), pp. 265-276.

Sadler, D.R. (1989) 'Formative assessment and the design of instructional systems, *Instructional Science',* 18(*2*), pp.535-550.

Sadler, D.R. (2010) 'Beyond feedback: developing student capability in complex appraisal', *Assessment and Evaluation in Higher Education*, 35(5), pp. 535-550.

Sharmini, S., Spronken-Smith, S., Golding, C. and Harland, T. (2015) 'Assessing the doctoral thesis when it includes published work', *Assessment and Evaluation in Higher Education,* 40(1), pp. 89-102.

Stracke, E. and Kumar, V. (2010) 'Feedback and self-regulated learning: insights from supervisors' and PhD examiners' reports', *Reflective Practice*, 11(1), 19-32.

Trafford, V. and Leshem, S. (2008) *Stepping Stones to Achieving your Doctorate.* Berkshire: Open University Press.

University of Malta (2008) *PhD degree Regulations.* Education Act, 2008.

Winstone, N.E., Boud, D., Dawson, P., and Heron, M. (2022) 'From feedback-as-information to feedback-as-process: a linguistic analysis of the feedback literature,' *Assessment and Evaluation in Higher Education,* 47(*2*), pp. 213-230.

Yang, M. and Carless, D. (2013) 'The feedback triangle and the enhancement of dialogic feedback process', *Teaching in Higher Education,* 18(3), pp. 285-297.

Students' use of online rubrics: Unexpected digital barriers to feedback literacy development

Joy Robbins and Milena Marinkova
University of Leeds

Abstract

While studies have extolled the value of using online rubrics, the benefits have usually been presented in terms of enhancing marking or delivery of teacher feedback. These benefits are welcome, but they nonetheless couch digital as simply an improved way for "old paradigm" transmission approaches to feedback that do little to help students develop feedback literacy. This study therefore investigates whether the affordances of online rubrics might also enhance students' metacognitive engagement with feedback. Five qualitative case studies followed students over 1-2 semesters as they submitted multiple pieces of work and received online feedback, including rubrics, via Turnitin Feedback Studio. Student perceptions were investigated through interviews and student-recorded screencasts in which students followed a think-aloud protocol as they engaged with their online feedback. The findings indicate that counter to our hopes for digital enhancement, the online rubrics in these cases tended to actually inhibit feedback literacy development. At the same time, participants' online behaviours showed a range of useful strategies for making sense of and acting on online feedback, even when the online rubrics themselves are lacking. This is something that programme and assessment teams should draw on in order to maximise learners' engagement with and learning from online rubric feedback.

Keywords
Online rubrics, digital rubrics, feedback literacy, formative feedback, TurnItIn.

Introduction

Assessment rubrics and feedback literacy

The growing use of rubrics as feedback and assessment tools that can enhance students' learning has prompted an accompanying growth of rubric research in higher education. While there have been strong criticisms of rubrics (e.g. Torrance, 2007; Sadler, 2009, 2014); the general trend in the research is positive, and the critiques thus far have not proven robust (Panadero and Jonsson, 2020). The various learning benefits of (considered, well-implemented) rubric use continue to be refined in the literature. For example, their noted effectiveness as a means of developing self-assessment judgement (Andrade and Du, 2005; Andrade, 2010) could be questioned as such benefits may well be due to instructional intervention – in these cases the teaching of self-assessment and rubric use skills – rather than showing evidence that rubrics *themselves* can help (Jonsson, 2014 p.841). However recent research has found that it is indeed the rubrics that are responsible for beneficial self-assessment development (Krebs, Rothstein and Roelle, 2022). Less common in the literature is investigation of whether the many learning benefits of rubric use carry over into the online or digital space, which, it need hardly be said, is increasingly where rubric use and all manner of assessment and feedback occur.

Online rubrics have been found to improve students' judgement of peer work in a MOOC environment (Ashton and Davies, 2015) though this was down to whether or not students received guidance, showing the same instructional intervention issue as previously noted rubric studies. Online rubrics have also been found to reduce the amount of time teachers spend creating feedback (Anglin *et al.*, 2008; Atkinson and Lim, 2013), allow analysis of consistency across a marking team (Reed, Watmough and Duvall, 2015), and increase teacher and student satisfaction with the feedback (McKinney, 2018).

While these benefits are beyond what paper-based rubrics could bring, they seem to pull online rubric use back into a so-called 'old paradigm' feedback approach (Winstone and Carless, 2022) where the emphasis is on one-way delivery of teacher comments and grades to students, rather than a more evolved feedback literacy understanding wherein students learn to appreciate the purpose and process of feedback, deal with its emotional aspects, and appropriately take action on it (pp.26-29). Being literate in this sense will have far more long-term beneficial impacts for a student than simply getting rubric feedback more quickly, more clearly, or more standardized. However, it has not yet been investigated whether the affordances of online rubrics can positively impact feedback literacy rather than just feedback deployment. The current study therefore examined how students used online rubrics, in this case in the Turnitin Feedback Studio platform, to see what feedback literacy benefits from rubrics in the online space could be observed.

Methodology and participants

Given the learner-centric nature of feedback literacy, we viewed this study as inherently qualitative, choosing to follow a few students in depth longitudinally via a case study approach to generate rich data related to each individual's feedback literacy development and how online rubrics were or were not present in that journey. Ethical approval for this study was granted via our Faculty Research Ethics Committee (reference LTSLCS-114) and signed consent was obtained from each participant. To be sure that all participants would have online rubrics as part of their feedback, we recruited students from our own School as school-wide online rubric use for both formative and summative assessment had been rolled out the previous year. Students were recruited via short promotional talks at the start of colleague's classes and through advertising in our self-access centre; engagement was entirely voluntary and none of our own students were recruited. Three participants engaged with the study across two semesters for academic year 2019-20, and as others had had to drop out due to Covid pressures, we recruited a further two students for the second semester for a total of five case studies.

To understand how students interacted with their online rubrics, while at the same time not guiding or coaching them in ways that would affect the validity of the study, we used a fusion data collection technique of screencasts with a think-aloud protocol. Students were asked to record a screencast of themselves accessing their feedback (just the screen, not their face) at whatever time and place they would normally do this, while speaking aloud what they were doing, thinking, or feeling. No time limits were suggested: length and focus were left entirely up to each student. We therefore hoped to diminish performativity and other impacts that might have occurred if a researcher had been present during student feedback use or if stricter guidance had been given. Students received 1-to-1 training in how to record and securely share the screencasts, and in how to follow the think-aloud protocol. For the latter, we asked students to engage in non-metacognitive verbalisation (Bowles, 2010), that is, they were asked to report what they were doing without necessarily interpreting their actions, thoughts, or feelings, thus reducing the cognitive load of thinking aloud. This likely made the recordings more accurate, and we further coached students to conduct their think-alouds concurrently, while they were performing the task of accessing their feedback, rather than retrospectively, thus additionally heightening accuracy (Ericsson and Simon, 1993). Each student brought their own laptop to the initial 1-to-1 training session and did a practice screencast think-aloud with support and encouragement from one of the researchers. This first meeting was also used to discuss how the study would work, make sure the participant was happy to sign the consent form, and discuss any questions.

Students were asked to make a screencast for each time they received online feedback in Turnitin. As students were English for Academic Purposes (EAP) students, studying advanced language use before entering their postgraduate studies, there were multiple formative and summative writing assignments throughout the semesters, a common feature in many EAP curricula, and therefore multiple possibilities for students to create screencasts. While not all assessments in the screencasts

had an online rubric as part of the feedback, we did not stipulate that a rubric had to be present, nor even share that we were only interested in rubrics, as this would likely have impacted the students' feedback use and what they chose to focus on while recording. As much as possible we wanted these narrated feedback-use videos to provide a window into the occluded practice of what students actually do with online feedback, allowing their engagement with the online rubric in particular to surface naturally. While we gave no rubric training, students had been shown the rubrics by their tutors and could access the rubrics in Word document form on the Virtual Learning Environment (VLE).

Each screencast was followed up (as scheduling allowed) with an open-ended, semi-structured interview to discuss the feedback interactions observed in the think-aloud video, and feedback beliefs and practices more generally. The interview schedule began with questions about the writing task the student had completed, then asked what they thought about the feedback and what they had done with it, then finally asked explicitly about the rubric. Where relevant, we also used stimulated recall (Gass and Mackey, 2016), showing students an image or snippet from their screencast as a prompt for them to comment on or further explain. As an additional data point, we wrote contact summary sheets for each student which were added to after each interview. While interviews were initially conducted face-to-face, the pressures of Covid moved them online during the study.

In total we had 18 screencasts and 12 interviews, with variance in the number from each student: on the highest end, S1 provided us with 6 screencasts and 3 interviews, while on the lowest end, S5 gave 2 screencasts and 2 interviews. (The difference is due in part to S4 and S5 being in the study for only one semester.) All data were transcribed, anonymised, and compiled into chronological cases, one per student. Each case was then iteratively reread and analysed in conjunction with viewings of the original screencasts to include more nuanced data elements that were beyond transcription, such as tone of voice and the actual content of the students' written work and teacher feedback. We employed a collaborative, reflexive coding approach to identify themes, or patterns of shared meaning (Braun and Clarke, 2019) across the cases, but also sought to preserve in our analysis each student's individual developing narrative (see Marinkova and Robbins, 2023 for more detail and a case-based rather than thematic analysis presentation of the data). Our analysis was inductive; we looked for repetition in observable online behaviours as well as explicit verbal comments to do with online rubrics, which led to the creation of three 'data domains' (Braun and Clarke, 2019): accessing online rubrics, making sense of online rubrics, and engaging with online rubrics. We built on these domains to arrive at the themes presented below.

Findings and discussion
What emerged from the data is five intriguing student journeys, all of which showed growth of feedback literacy to varying extents. However, as we wish to consider only where the online rubrics affected this growth, the findings here are presented by themes that emerged across the case studies.

Theme 1: Access not found
Perhaps surprisingly, the majority of participants (4/5) did not use the "View Rubric" button for the TurnItIn rubric (see Figure 1) in their first screencast with rubric feedback available. All students were motivated enough to take part in this voluntary study for which the only reward was the possibility to talk about feedback, and all had engaged with the in-text and overall feedback comments on their work in their screencasts. In other words, all were keen but the digital rubric presentation, requiring a further click to access, was less effective than an old-fashioned paper rubric would have been.
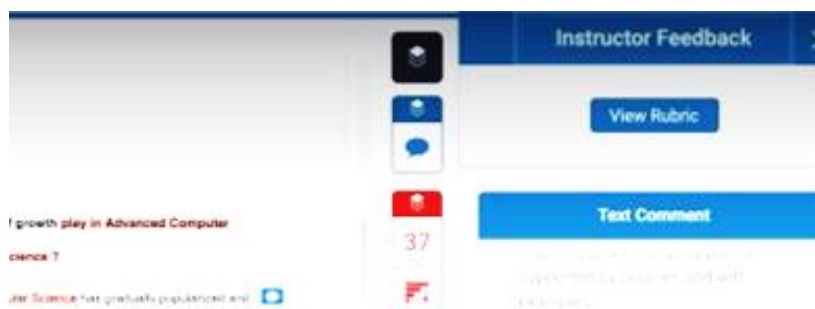
Figure 1. "View Rubric" button on TurnItIn.

This barrier cannot be ascribed to lack of knowledge about rubrics, since follow-up interviews revealed that participants had been made aware of rubrics by their tutors and did know where to find static rubrics (e.g., in Word file on the VLE). Nor can it be ascribed to low technical skills as these students all easily managed the many-stepped tech demands of the current study, and students were also able to explore the Turnitin interface. For instance, participant S3, a tech-savvy student completing a degree in Computing, found the Similarity check features herself but not the rubric. In the follow-up screencasts and interviews, participants S2, S3, and S5 – who had not found the online rubric initially – were eventually able to make (various) uses of the online rubric by the end of their course. However, participant S4 never did.

Theme 2: Interface is difficult to perceive, awkward to operate and hard to understand
Even once the online rubric was accessed, participants struggled with its inaccessible interface. As participant S1 comments in her screencast, once opened, the TurnItIn rubric appears as an overlay and covers the entire screen (see Figure 2), including tutor commentary and assignment script, preventing her from engaging with the online rubric feedback while also looking at her writing. This had obvious negative impact on her ability to actively engage with the rubric feedback.

Figure 2. A full view of a TurnItIn rubric, blocking tutor comment and script.

As a motivated student who wanted to use the online rubric together with her writing and other feedback, she attempted digital workarounds. Initially she tried screenshotting the rubric but found "it's not possible to make screenshots because you see the whole text is not too clear and it's disappear." She then hit on a much more cumbersome method wherein she downloaded the static rubric document from the VLE and highlighted it according to what her teacher had selected in the online rubric. This meant that to use the online rubric as a learning tool, i.e., to make sense of her

feedback, S1 had to forsake its interactive affordances, such as hyperlinks to specific annotations in the script.

The easier approach might appear to be printing out the Turnitin rubric with teacher highlights already included, however, as participant S3 notes, the print view of the TurnItIn rubric is unwieldy. "I think the rubric is OK. But I think the print ....it's not easy. And I didn't see how to print it. It's…not simple to print it." What S3 explains is that when downloaded for printing, TurnItIn rubrics no longer appear as a succinct, 1- or 2-page table. Instead, the format is changed beyond recognition: criteria and associated band descriptors appear as a multi-page list below the assignment script, which is followed by another long list of in-text annotations (see Figure 3.). Students have to scroll up and down these lists in order to make sense of the different elements of online feedback: grade and tutor commentary, rubric descriptors and in-text annotations.

Figure 3. Print/Download view of a TurnItIn rubric.

Even when project participants managed to manipulate the rubric interface by minimising or moving screens around, they struggled seeing the band descriptors (i.e., rubric cells) that their tutors had selected and thus missed important learning opportunities. For instance, S4 was aware that a rubric shows the criteria against which students' work is assessed; however, she hadn't realised that the online rubrics can offer a breakdown of the grade for each criterion: "I didn't know about each part can be a grade". This gap in her assessment literacy was down to not being able to see the tutor highlights because of the online rubric design. In this case, the online rubric was used for categorical marking of a summative; in other words, within each band 3 defined points were used for mark allocation (e.g., 42 – 45 – 48). Categorical marking allows markers to use the online rubric to calculate the overall grade with considerable ease, precision and consistency (Isbell and Goomas, 2014). As a result, however, the number of descriptors (i.e., cells) per criterion can increase exponentially, making the rubric too crowded and descriptors hard to tell apart. Moreover, when opened in TurnItIn, the rubric view only shows up to 5 band levels (either at the top or bottom end of the grading scales). Given the sheer number of band descriptors used for categorical marking (over 10 in this case), and the fact that S4's grade was somewhere in the Merit range, all she could see when she opened the online rubric was the top end descriptors, which hadn't been selected (see Figure 4). Consequently, the online rubric did not have any impact on S4's evolving feedback literacy; in fact, the student had been much more engaged with the static version of the rubric.

Figure 4. Rubric view without any highlights visible.

When project participants succeeded in navigating the multiple band descriptors, and spotted what performance level they'd been awarded, the metalanguage used in online rubric descriptors could present a challenge. Participant S5, for instance, noticed the differences in the language of descriptors *between* different bands. However, the identical wording of descriptors *within* the same band – which is how categorical marking is operationalised in TurnItIn – proved confusing: "My 'Whole text' is Distinction, but I have some problem about the scores for 68 and 65 because their details are same with each other. So which type of writing belong to 68?" (S5). The student could see that the cell for 65 was selected by his marker, but the descriptor was identical to that for 68. This seems to have raised questions about the precision of grade calculation with online rubrics. For unlike paper and static rubrics in Word, where markers could be more granular and select multiple cells (i.e., sub-descriptors) for the same criterion to signal a marginal or uneven performance levels, in *TurnItIn* this is not possible and only one cell can be selected. Thus, even though online rubrics seem to be efficient

marking tools (Anglin *et al*., 2008; Atkinson and Lim, 2013), they can less effective as learning tools including in guiding students to understand grade calculation.

This is summed up in the following exchange with participant S2, following the receipt of her summative feedback:

> S2: I think this one, the rubric, is not useful for me…. I always find for printed or read to see I'm "Distinction" or I'm "Merit", like this, because I have to pass […], I have to be more than another this level, a little more.
> *Interviewer*: Right. Yeah, so just to sum up, the most important thing about the rubric is the grading so you can see where you are. You don't use it for learning.
> *S2*: Yeah, sorry for that. […] Yeah because I think it's so many sentence making for me they are confused. So many sentence and boxes […], same levels or same "Distinction" or same "Merit", like this. So many.

A highly motivated student of Medical Imaging, S2 repeatedly pointed out the grade as her extrinsic motivation for engaging with the rubric and acting on feedback. Not surprisingly, this participant made use of the rubric for summative purposes only, monitoring her performance level rather than developing her understanding of criteria: "I saw where is the line for 'Distinction', 'Merit', like this. And sometimes I have to read quickly." (S2) This doesn't mean that S2 didn't recognise the importance of feedback. As a matter of fact, this participant was particularly invested in the social- affective dimensions of feedback (Yang and Carless, 2013), and appreciated the confidence-boosting messages in her tutor's commentary, "with marks or without marks, the positive words, it's important" (S2). S2 also engaged with the cognitive dimensions of feedback, by drawing on what she called "negative feedback": "[I] Like to see the positive side and ignore the negative side, but I have to learn from the negative side." (S2) However, the cluttered layout and replicated descriptors of the TurnItIn rubric, as key elements of the structural dimension of online feedback (Yang and Carless, 2013), seem to have acted as a barrier to the development of S2's feedback literacy.

Theme 3. Teacher-owned tools that can't be used for self or peer review
An important aspect of feedback literacy is student ownership of feedback: from feedback sought, generated and negotiated by students as "active constructors" (Nicol, 2010, p.503) rather than being simply "delivered" by tutors in a one-off transaction (Carless and Boud, 2018). When it comes to TurnItIn rubrics, however, these seem to be exclusively tutor-owned marking tools, which students cannot use for dynamic online self- or peer- assessment purposes at the formative, pre-submission stage. And even though there is a PeerMark functionality in TurnItIn, which can lend certain degree of student agency to generate and review peer feedback (Nicol et al., 2014), this functionality doesn't allow for a marking rubric to be attached and used in peer marking activities; rubrics can only be attached to tutor-created assignments and used for tutor-initiated assessment purposes. Neither can online rubrics sustain or generate dialogic feedback, either with peers or with tutors, post-submission. Instead, as project participant S4 points out, a static rubric shared as a Word file seems to have given students the flexibility to use the tool for self-monitoring: "I use [electronic rubric in Word] on my computer whenever I need it, like before the writing or after receiving the feedback, I would open it to check where I can improve and where I have done great" (S4).

This paradoxical situation of online rubrics lacking in interactive affordances to the point of limiting learner autonomy and dialogue was particularly relevant to participant S3, a pre-Masters student in Advanced Computer Design, who demonstrated high levels of learner autonomy and willingness to co-construct knowledge with her peers. Initially unaware of the TurnItIn rubric, S3 only realised the "View rubric" button was there "when friends told" her (S3). Open to dialogic feedback practices, S3 was proactive in seeking further clarification of her feedback from her tutor or the advice of

classmates: "I also to talk about my question, my problems with my friend. And he also received the feedback, and we compared others' feedback and we will also talk about how to write this article better." (S3). In an attempt to exert this agency, S3 wanted to share the online rubric with her peers (pre- and post- submission) so that they could engage in a meaningful conversation: "I want to ... add or share links with others... Maybe… this rubric cannot share now." (S3) The student, therefore, found herself constrained by the online medium and her alternative strategy to engage in *dynamic* dialogue was, just like with S4, to deploy the *static* rubric on the institutional VLE.

Theme 4. In-task feedback synthesis was facilitated
Given the interactive affordances of online rubrics, we assumed that in-task feedback synthesis can be easily facilitated. Enabling students to see the links between various comments and criteria can help them make sense of the diverse feedback elements online, move beyond reliance on direct corrective feedback and "visualise 'the bigger picture'" of their learning (Winstone and Carless, 2020, p.65). This could be achieved through the hyperlinking facility in Feedback Studio, which links individual in-text annotations to a relevant criterion in the rubric. If used consistently, this tool should provide learners with more detail about which specific instances of their writing meet the marking criteria, how and why. In addition, the number of hyperlinked annotations to a specific criterion can also give students and markers an indication of whether there is an area of learning that might need further developing.

What we observed in project participants' online behaviour in their screencast recordings, however, was that they often did not draw on the interactive affordances of TurnItIn rubrics. As Figure 5 shows, learners tended to prioritise the overall tutor commentary ("Text Comment") and then move onto in-text annotations. Depending on the student's levels of self-efficacy and the nature of the task (summative or formative), there could be some synthesis (moving back and forth) between overall feedback and in-text annotations. Engagement with the third element of Feedback Studio, the online rubric, tended to be limited, usually at the end of the in-task synthesis, with only participant S5 making the most of these interactive affordances.
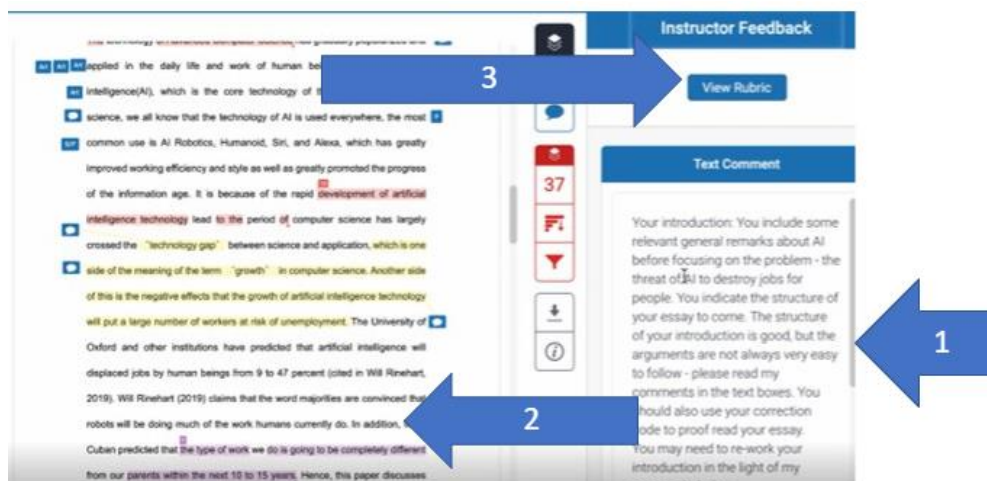
Figure 5. Preferred route for in-task synthesis of online feedback elements.

S5 was a highly driven pre-Masters student of Electrical Engineering and Renewable Energy Systems. From the very start, this participant demonstrated high metacognitive awareness and feedback literacy: in his screencasts, S5 would summarise key feedback messages, relate to relevant in-text annotations, and reflect on his prior learning and next steps rather than simply reading out loud tutor comments. S5's approach to the online space of Feedback Studio was integrated as well, flexibly

synthesising and navigating between different elements. In addition to making summative use of the rubric as a way of monitoring own performance in assessments, this participant deployed the online rubric formatively in order to make sense of the overwhelming number of annotations in his script (see Figure 6, left) and categorise these into specific areas for improvement. "You can see in my assignment draft, there a lot of blue details [speech bubbles used for in-text annotations]; so, you know, it looked very complex at first. But after I go to the rubric, I found that there are … only four or five information because some information can be classified into the same topic, such as 'Structure' and other criteria, yeah. And that rubric table helped me get a better understanding my tutor's advice." (S5)

Figure 6. Assignment script with annotations and rubric with hyperlinked annotations (first column on the left with the speech bubbles).

Nonetheless, for all the advantages of hyperlinking, the decontextualized list of comments in a drop-down menu that appears over the online rubric was confusing to S5 (see Figure 6, right):

> *S5*: I think the network [hyperlinking] is useful for me because I know clearly which part had more feedbacks about my contents, but the details on one sentence I think I can't understand them because as I'm reading the sentence, I think I need to [go] back to the contents to … find the details.
>
> *Interviewer*: So, here they're a bit out of context, aren't they? You need to go to the text?
> *S5*: Yes, but I think the overview comments in the [right] window is very important to me because after reading every sentence I find the details in the left window, not in the rubric part.

In the extract above, the participant highlights the need to make sense of the tutor comment in the immediate context of his writing. Thus, while the interactive affordances of online rubric can certainly support the cognitive processing of individual in-text comments and their relevance to marking criteria through handy hyperlinks, the experience of S5 reiterates the conclusion reached in feedback literature about the meaningfulness of feedback and rubric-use in context, with exemplars or other types of authentic student writing (Tierney and Simon, 2004; Sadler, 1989).

Theme 5. Cross-task synthesis of feedback must be in a different platform
Beyond in-task synthesis, making it possible for students to consolidate their knowledge and transfer learning to other contexts can also support feedback literacy development (Hattie and Timperley, 2007). And it seems that online rubrics via TurnItIn make this cross-task synthesis hard, if not impossible – the feedback lives within a single Turnitin assignment with no way to see other assignment and feedback at the same time. As already discussed, if rubric criteria, in-text annotations or tutor commentary are to be exported, the TurnItIn downloads are so confusing and the process so

awkward, that our project participants simply opted for static, often paper-based, alternatives. Others came up with alternative online solutions; for instance, participant S4 created her own multipage learning portfolio in OneNote (see Figure 7), which gave her the online affordances, flexibility and agency that Feedback Studio couldn't:

> I summarised some of my problems here and in order to avoid the same mistakes next
> time. So firstly, it's about my academic language... And second problem is about my paraphrasing 'coz it's the first time for me to do this kind of writing task…. So, after this, I summarised some tips... And lastly, about my references here …I like to write a list, a checklist, for example, something I should avoid and something I should improve. So, next time during the process of writing other assessment tasks, I can use this checklist then to examine which I need to improve in my future task. (S4)
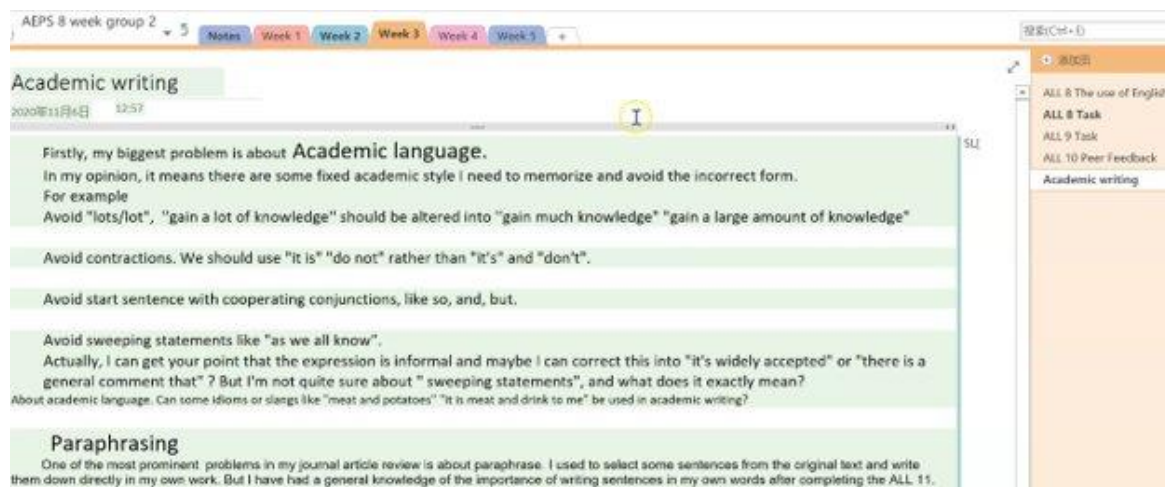
Figure 7. A student's feedback portfolio in OneNote.

The OneNote platform allowed S4 to generate her own comments, ask her tutor questions, draft action plans and thus avoid an instrumental and passive engagement with rubric, feedback and learning (Sadler 2014; Torrance, 2007).

A different, low-tech strategy was adopted by another project participant, S1. A pre-Masters student in Education, S1 was invested in teaching and learning as an educator herself and was the only participant who made use of the online rubrics from the very first data collection point (see above), which shows her advanced feedback literacy. Therefore, it wasn't surprising that she was keen on transferring her learning to different tasks. What was surprising to her was the fact that in order to do so she had to choose a different medium: S1 opted for creating a hard copy folder of print rubrics and tutor feedback, cross-referenced with a checklist:

> And what I do also I made the folder for all feedback from level two until level four, and collect all these together because this is helpful to see the progress how we are considered the feedback and take action to develop…I put two column - positive and negative comment that mean I have to consider and to improve […]
> What I'm do just I note for each criteria, and … I just highlight … "Distinction" or "Merit", and then I will highlight the Word document and I will highlight… my final score. (S1)

Keen to use the rubrics for purposes other than summative and to extend their formative impact to learning beyond her current assignment, S1 found that she could export neither rubric nor tutor feedback (nor in-text annotations for that matter) in a format that is operable. Rather, she had to

recreate the feedback space in a completely different medium, that of paper. Thus, in order to engage actively and meaningfully with the feedback from TurnItIn rubrics, act on it and develop their learning, both S1 and S4 had to …leave TurnItIn altogether!

Implications

While we had high hopes that the digital affordances of online rubrics would bring advantages to feedback literacy development beyond what their paper cousins could do, our investigation revealed instead that students developed feedback literacy in spite of the many barriers that online rubrics presented. There are of course limitations to this finding, specifically the small number of participants and the use of only a single online rubric platform. It would appear though that the operational barriers to rubric use would likely be similar in larger cohorts given their basic nature.

The innovations of the students in these case studies teach us as educators that we should let rubrics be in student-owned, dialogic spaces so that any learning potential they have can actually be realised. Such spaces might be in Word online/ Microsoft tools or in Google, and interesting steps on dialogic feedback in more open tools have already been taken (Wood, 2021a, 2021b). Additionally, we should perhaps consider scaffolding students to find an individual approach to feedback engagement that suits their preferences, as S1 and S4 did, rather than requiring a single online solution for all. While wholesale removal of Turnitin and similar restrictive platforms may not be possible in our contexts, they can be used only where necessary, e.g. mandated similarity checking, storing work for Quality Assurance compliance, while true feedback-driven learning happens engagingly elsewhere. Finally, we hope the case studies presented here make a clear call for programme and assessment teams to investigate the myriad ways their own students do (and don't!) engage with online rubrics and feedback so that learning can be enhanced.

References

Andrade, H. and Du, Y. (2005) 'Student perspectives on rubric-referenced assessment. *Practical Assessment', Research & Evaluation*, 10(3), pp.1–11.

Andrade, H. L. (2010) Students as the Definitive Source of Formative Assessment: Academic Self-Assessment and the Self-Regulation of Learning in *NERA Conference Proceedings*. https://opencommons.uconn.edu/nera_2010/25, p. 25.

Anglin, L., Anglin, K., Schumann, P. L. and Kaliski, J. A. (2008) 'Improving the efficiency and effectiveness of grading through the use of computer-assisted grading rubrics', *Decision Sciences Journal of Innovative Education*, 6(1), pp. 51-73.

Ashton, S. and Davies, R. S. (2015) 'Using scaffolded rubrics to improve peer assessment in a MOOC writing course', *Distance Education*, 36(3), pp. 312–334.

Atkinson, D. and Lim, S. L. (2013) 'Improving assessment processes in Higher Education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS', *Australasian Journal of Educational Technology,* 29(5), pp. 651-666.

Braun, V. and Clarke, V. (2019) 'Reflecting on reflexive thematic analysis. *Qualitative Research in Sport', Exercise and Health,* 11(4), pp.589–597.

Bowles, M. (2010) *The Think-Aloud Controversy*. New York: Routledge.

Carless, D. and Boud, D. (2018) 'The development of student feedback literacy: Enabling uptake of feedback', *Assessment & Evaluation in Higher Education,* 43(8), pp.1315-1325.

Ericsson, K. and Simon, H. (1993) *Protocol Analysis: Verbal Reports as Data*. 2nd ed. Cambridge, MA: MIT Press.

Gass, S. M. and Mackey, A. (2016) *Simulated Recall Methodology in Applied Linguistics and L2 Research*. 2nd ed. New York: Routledge.

Hattie, J., and Timperley, H. (2007) 'The Power of feedback. *Review of Educational Research',* 77(1), pp. 81-112.

Isbell, T. and Goomas, D. (2014) 'Computer assisted rubric evaluation: Enhancing outcomes and assessment quality', *Community College Journal of Research and Practice*, 38(12), pp. 1193-1197.

Krebs, R., Rothstein, B. and Roelle, J. (2022) 'Rubrics enhance accuracy and reduce cognitive load in self-assessment', *Metacognition and Learning*, 17, pp.627–650.

Marinkova, M. and Robbins, J. (2023) 'Can Online Rubrics Develop Learners' Metacognition?: A Qualitative Case Study Analysis, in Gonsalves, C. and Pearson, J. (eds.) *Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn.* Hershey, PA: IGI Global. pp. 248-279.

McKinney, B. (2018) 'The impact of program-wide discussion board grading rubrics on students and faculty satisfaction', *Online Learning Journal*, 22(2), pp.289–300.

Nicol, D. (2010) 'From monologue to dialogue: Improving written feedback in mass higher education', *Assessment & Evaluation in Higher Education,* 35(5), pp.501-517.

Nicol, D., Thomson, A. and Breslin, C. (2014) 'Rethinking feedback practices in Higher Education: A Peer Review Perspective', *Assessment & Evaluation in Higher Education*, 39(1), pp. 102-122.

Panadero, E. and Jonsson, A. (2013) 'The use of scoring rubrics for formative assessment purposes revisited: A review', *Education Research and Reviews*, 9, pp. 129-144.

Reed, P., Watmough, S. and Duvall, P. (2015) 'Assessment analytics using Turnitin and Grademark in an undergraduate medical curriculum' *Journal of Perspectives in Applied Academic Practice*, 3(2), pp. 92–108.

Sadler, D. R. (2014) 'The futility of attempting to codify academic achievement standards', *Higher Education*, 67, pp.273–288.

Sadler, D. R. (2009) 'Indeterminacy in the use of preset criteria for assessment and grading', *Assessment & Evaluation in Higher Education*, 34(2), pp.159–179.

Sadler, D. R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science,* 18, pp.119-144.

Tierney, R. and Simon, M. (2004) 'What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels', *Practical Assessment, Research and Evaluation,* 9(2).

Torrance, H. (2007) 'Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning', *Assessment in Education*, 14(3), pp. 281–294.

Winstone, N. and Carless, D. (2022) *Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach*. Abingdon: Routledge.

Wood, J. (2021a) 'A dialogic technology-mediated model of feedback uptake and literacy', *Assessment & Evaluation in Higher Education*, 46(8), pp. 1173–1190.

Wood, J. (2021b) 'Making peer feedback work: the contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy', *Assessment & Evaluation in Higher Education*, 47(3), pp. 1–20.

Yang, M. and Carless, D. (2013) 'The feedback triangle and the enhancement of dialogic feedback processes', *Teaching in Higher Education*, 18(3), pp.85–297.

Academic feedback and performance of students in an institution of higher education: How does our feedback impact our students?

Dave Darwent and Amy Musgrove
Sheffield Hallam University

Abstract
Attending higher education institutions and achieving academic success are associated with positive outcomes, valued at individual and societal levels, such as reducing unemployment and poverty and increasing civic participation. Thus, many studies have focused on enhancing learning by examining the factors that affect students' performance, including the teaching behaviour most strongly related to academic success. Within this research framework, in a population of social sciences students on different degree routes, the relationship between academic performance and different components of academic feedback is explored. The feedback students received in the dissertation module, which had two assessment points, was examined. The first assessment point feedback was statistically analysed to find changes in the grade point average of students between their first submission (T1) and the final dissertation (T2). Analysis to assess the impact of students' performance on the structure and content of feedback, as well as the extent to which the content of the feedback, can affect T2 grade increases.

Keywords
Assessment; Attainment; Feedback; Performance; Stanford-WISE.

Introduction
The aim of this paper is to present some preliminary findings of a mixed methods research project analysing feedback and the impact of this on attainment. This research was initially conducted across two departments in one Higher Education Institution in the UK. Both authors were working in their respective departments trying to seek ways of improving feedback for their students. One project was exploring whether Higher Education was training teachers to give praise and was based on qualitative focus groups with three cohorts; trainee teachers, Higher Education lecturers in teacher education and a random sample of Higher Education students at the institution. Two key findings from this project were that teachers stated that they were giving praise to students but the students received this as feedback and that the destructive influence of poorly constructed feedback far outweighed any developmental benefits (Darwent, 2016). The disjuncture between students' and teachers' understanding of what 'praise' is was quite startling. Dictionary definitions of praise (e.g. "*to express admiration or approval of the achievements or characteristics of a person or thing*" (Cambridge Dictionary, 2019)) bore no resemblance to what teachers were providing, and it was established that most educators and learners do not define praise as the dictionary does, nor as each other do. It was further established that many teachers believed that the purpose of praise (as they called it) was to encourage or stimulate learners to improve their work and thus achieve more highly: this later formed a crucial link with a second research project (see below). The researcher sought to develop a model for praising achievement whilst providing constructive feedback and feed-forward without the latter negating the former.

A separate project running in another department in the same institution was focussing on assessment guidance and feedback received. The first scoping exercise was conducted with second year undergraduate students and demonstrated that many of the practices were at modular level and there was very little programme approach (Jessop, et al., 2014) taking place. Feedback was perceived

as unhelpful as it was focussed on the task completed, rather than on wider learning objectives, and it was inconsistent across modules. Focus groups were then conducted with second year undergraduate students to explore some of the issues in greater detail and it was found that standardised assessment information would be helpful, level approaches to assessments were of benefit to students. Standardisation of feedback mechanisms were also something which the students welcomed.

Through the dissemination of this work within the institution the projects were brought together to analyse feedback in more detail and in doing so we evaluated historical student feedback on the dissertation module which had two assessment points. The intention here was to see if there were patterns in the feedback in the early assessment which could support an increase in attainment at the final submission point. As the first researcher's investigation into praise had revealed that teachers believed this to be the purpose of praise and were giving feedback whilst referring to it as praise, it was established that there was a strong connection between the two projects and it was decided to merge them into a single project. It should be noted that the researchers do not seek to define "good" feedback as that which results in measurable improvement in student work; rather we seek to devise and trial models of feedback which achieve the stated aim of the teachers in both initial investigations which was to facilitate improvement in students' attainment.

Research Context
This research paper will discuss the preliminary findings from a pilot study on attainment driven feedback. This research does not intend to debate the forms of feedback within higher education as there is significant evidence of this within the academic literature (Merry and Orsmond, 2008; Lunt and Curran, 2010; Ball, 2010; Hennessy and Forrester, 2014), however, what this paper will do is present the findings which demonstrate that there is a strong link between the content of feedback and the impact on attainment. This is seen as important as many of our students are driven by higher grades (Lund University, 2013) and the feedback we give can support this increase and add value to our students' learning (Carless and Boud, 2018; Hawk and Lyons, 2008; Kauffman, 2015).

In particular we discuss our findings that feedback is least effective when commoditised, which has been found to frequently be the case (Dunworth and Sanchez, 2016), and much more effective when the student is involved in the feedback process, such as by a meeting to discuss work. This should not be surprising as it has been found before (for example Bloxham and Campbell, 2010; Carless et al, 2011, McArthur and Huxham, 2013) that an active dialogic process is favoured by students and enables them to be more effective independent learners who can make accurate judgements of the quality of their own output (Nicol and Macfarlane-Dick, 2006; Sadler, 1985). Despite this well-established knowledge, we found that the majority of feedback is still commoditised and passed to the student as an entity.

The two small departmental projects which informed our investigations on feedback for this paper found that that at every level, from primary school to postdoctoral, combining praise and feed-forward (Jones, 2005) as typified in "praise burger" (Darwent et al., 2018) style models is at best ineffective and more often destructive, demotivating and undermining. Outside of education this seems to be accepted readily, e.g. in the Harvard Business Review: "If you give a feedback sandwich, you risk alienating your direct reports. In addition, they are likely to discount your positive feedback, believing it is not genuine." (Shwartz, 2013). Learners, educators and dictionaries all mean different things when they use the term "praise". Bloxham *et al.*, suggests that there is an overreliance on the students or markers perceptions of feedback, rather than the interaction between them (Bloxham et al., 2011). When individuals have a different understanding or interpretation of feedback it defeats the object and becomes ineffective in supporting the student's progression.

Furthermore, the two smaller departmental projects found that standardisation of assessment information was beneficial to students' understanding of what was required of them and they liked receiving information in a familiar way. This meant that consistency across their modules was important to their understanding of the task(s) being set, as this had a direct impact on their understanding of the feedback they then got. Feedback needed to be personalised and specific to the student (Austen and Malone, 2018) but not critical, from their perspective, and not blandishments if the mark did not reflect a "good" mark from their perspective. The result of this being that students who have a negative perception of the feedback they receive are demotivated and lack determination (Hawk and Lyons, 2008; Kauffman, 2015) which impacted on their attainment in later assessments. However, positive feedback and feed-forward commentary on students' work yields an increase in attainment (Hawk and Lyons, 2008; Patchan *et al.*, 2016). Making clear links between assessment information and students understanding of feedback was a key finding from the departmental project and from within the literature (Bloxham *et al.*, 2011; Austen and Malone, 2018; Hawk and Lyons, 2008; Kauffman, 2015; Patchan *et al.*, 2016).

Anecdotally, many academics in Higher Education know that their students are disappointed and frustrated with the feedback they receive from their assessors, particularly when they do not receive the marks they had hoped for (Alfehaid et al., 2018). Many institutions see this in national statistics, such as the National Student Survey (NSS England) and local feedback around assessment and feedback. Many assessors know that feedback is fundamental to a student's learning and overall performance and attainment and yet it is an aspect which is consistently highlighted as problematic. Indeed, detailed examination of NSS results consistently indicate that nationally the lowest levels of satisfaction are given for assessment and feedback (Williams, Kane, Sagu and Smith, 2008), for example in 2015 only 68% of respondents found faculty feedback to be helpful (Bell and Brooks, 2017). The following year HEFCE reported that satisfaction levels relating to assessment and feedback were again the lowest, with full time students reporting an average of 73% satisfaction (HEFCE, 2016). More recent NSS results show that the situation continues to be of concern.

At this point it is important to highlight the difference between attainment – commonly referred to as 'marks' or 'grades' such as in Alfehaid et al. (2018) and performance. Attainment generally, and especially in education, refers to the highest award that the student has successfully achieved to date. For example, Education Scotland (2022) describes attainment as "Attainment is the measurable progress which children and young people make as they advance through and beyond school … " Performance, however, is the measure of how nearly one's attainment matches an expectation, such as a target set by a teacher, a national average, personal aspiration or the pass criteria for an assessment. Talib and Sansgiry (2012) define performance thus: "Academic performance is the extent to which a student, teacher, or institution has attained their short or long-term educational goals and is measured either by continuous assessment or cumulative grade point average." This difference is significant since our research started from the assumption that "*good feedback practice is broadly defined … as anything that might strengthen the students' capacity to self-regulate their own performance*" (Nicol and MacFarlane-Dick 2006: 205).

Our aim was to explore what types of feedback meet the students' desire for better attainment and support staff to best utilise their assessment time by providing students with useful, attainment driven feed-forward and feedback. This research explored feedback characteristics which supported student development and attainment, in turn, leading to greater opportunities for improvements in student attainment. Specifically, this means that if staff confront feedback by addressing the emotional and practical dimensions, student feedback literacy may be improved (Sutton, 2012) which, in turn, may improve attainment (Deenen and Brown, 2016).

Methodology

This project involved using historical feedback from 181 final year undergraduate dissertation projects. The dissertation involved two assessment points, and both sets of feedback were used in this analysis. Although the data was taken from one module, the students were on three different courses, two of which are jointly taught cross-departmentally within the institution. The first assessment point was six weeks after the start the module and consisted of a short piece, 1500 words, which summarised some of the literature students have found so far, what their chosen methodology will be, proposed central research question and any ethical issues they see arising from their work. The purpose of this assessment was twofold; firstly, the students getting ethical approval for their research and, secondly, for the student to demonstrate they had started to consider the key areas which would support their final assessment point, the dissertation project. The second assessment point was the submission of the dissertation at the end of the semester two. The aim of this piece of work is for the student to demonstrate their capabilities with independent study, to select an appropriate approach for the question and consider their chosen topic from a critical standpoint. All the dissertation projects followed a similar pattern for the structure but there were some variations depending on the type of methodology used.

The feedback for the dissertation project was done by the supervisor of the students work. There were 25 assessors across two departments working on this project. Assessors were expected to return feedback on the assessment within 15 workings days of submission. Although there was some standardisation for the feedback given, namely, maximum number of words used, electronic qualitative feedback and a feedback matrix set against broad learning outcomes for the module, for example, on the whole assessors applied their own style at both assessment points.

Ethical approval for this project was granted by the host institution. All the feedback analysed was electronic feedback, either typed or spoken. The first assessment point was in semester 1 (T1) of the institutions' academic calendar around the six-week mark. The second assessment (T2) came at the end of the module in the institutions' end of year assessment period. The methodology used was a mixed methods approach in that we conducted a content analysis of T1 and T2 feedback data in order to uncover themes, categories and names for categories. The aim here was to explore the feedback across the dissertation project module to look for the presence of certain words, themes, or concepts within the qualitative data set. In order to protect the reliability and validity of the data a research assistant was hired to conduct the analysis of the T1 and T2 data. They were blind to the purpose and did not have any preconceived ideas about what they might find. They labelled the feedback with no bias (Mackieson et al., 2018 and Smith and Noble, 2014). The study started with no preconceived categories, but categories were developed and labelled as patterns emerged.

The feedback for the module contained two key elements, a tick box matrix which was built into the virtual learning environment and a space for qualitative comments. The qualitative comments had no content guidance attached other than a maximum word count, therefore, assessors were free to comment upon any aspect of the work based on what they had read. When the content analysis was conducted, we were, therefore, not bound by any specific guidance the assessors had been given. The categories which were developed in the T1 and T2 content analysis were found to broadly match the Stanford WISE (Yeager et al., 2013) feedback categories of, firstly, feedback description - lecturer explicitly describes the nature of feedback being offered; secondly, high standards - lecturer emphasises and explains high standards used to evaluate the student work and generate the instructional feedback; and thirdly, assurance of student ability - lecturer states explicitly that the student has the skills necessary to successfully meet those standards (Yeager et al., 2013). It should be noted here that the no individual assessor followed the Stanford WISE model in full, but when looking at the module feedback as a whole, it was evident that these characteristics exist.

The second element of analysis was quantitative in nature and involved looking at the relationship between the categorisation of feedback and attainment across T1 and T2 data. The aim was to assess if, or how much, value was added. To do this a focus was placed on factors which may have affected the grade point average (GPA) between T1 and T2. Conducting various statistical analysis also supported the identification of factors which may have influenced attainment between T1 and T2. This included specific feedback characteristics but also specific demographics; for example, gender, ethnicity and course studied, to identify whether there were any variations across these categories too. The demographics of the participants was typically female dominated (79% female) with ethnicity being identified as white [any] (95%).

The feedback matrix was not part of this data analysis due to its tick box approach and standard nature.

Findings
Feedback for both assessments on the dissertation project module consisted of written or oral feedback and a final year undergraduate standard (for the subject area) matrix. 99.5% (n = 180) of the feedback was given in electronic written form using the institution's web-interface and a completed matrix. 0.5% (n = 1) received only matrix feedback. Despite the optionality, no one received oral feedback. 25 assessors from across two departments at the institution were providing feedback on this module. Throughout the analysis of written feedback, it was clear to see that assessors had provided consistent feedback within their own marking but that this wildly differed between assessors and across departments. Written feedback varied between less than 100 words in some cases to over 400 words in others. However, what was apparent was that assessors gave well-meaning feedback which reflected on what they had read and gave praise to achievements that the individual work had demonstrated.

In analysing the data we discovered and labelled six characteristics of feedback from the 25 assessors. These were: structural feedback, follow-up meetings, unfocussed critiques, focussed critiques, unfocussed affirmation, focussed affirmation. Structural feedback was defined by some routine characteristics of academic work, for example, the way the assessor structures the work, a greeting, grade included. The follow up meeting was offered on very few occasions, but this was done by way of offering a face-to-face meeting. The unfocussed critiques were where assessors made broad statements about the inadequate features of the work without being specific, for example, *your work has been rushed*. Focussed critiques are where the lecturer highlights what the student did wrong and why; and how they can improve to get a better result, for example, "*as it stands this work doesn't meet the pass criteria because []... however if you do []... then it will meet the criteria*". Affirmation unfocussed is concerned with what the student has done well without identifying specific examples, for example, *this is great work*. Finally, focused affirmation states what the student has done well and explains why it is good. It might even suggest further refinement: for example, "*this is excellent work because you did []… and you could even improve it further by []…*"

Although the structured feedback varied hugely in style and content across the various assessors, it was clear that they were following, either intentionally or unintentionally, a "praise burger" (Darwent *et al.*, 2018) model. The praise burger is structured by three key elements which are order specific, affirmation-criticism-affirmation. It is a very formulaic approach where assessors state something positive, then negative and then finishing on another positive. In this project, we found this structure was typically linked to unfocussed critiques as more general comments, for example, "*it is great that you have... [ ] however, I have notice that your... [ ] needs work, but I am pleased that... [ ]*". The affirmation element, which assessors wrote more words on, derived from the positive aspects of the work they had read but fell into two categories. Some cited areas which had been executed well, for example, "*the methodology is good*" (Hattie and Timperley, 2007); whilst others made more general

comments about the work, for example, *"overall this is a good dissertation"* (Hesketh and Laidlaw, 2002). Feedback should be specific and directive, otherwise students are unlikely to engage with the feedback and unable to understand how to improve (Thompson, 1997).

When assessors are focusing on the middle of the "praise burger" the majority of the criticism detailed what had gone wrong without any direction on how they could have improved. This means that assessors either focussed on a very specific error or made unfocussed generalisations about the work. It can thus be seen that assessors – whether knowingly or otherwise – are closely following the "praise burger" (Darwent *et al, 2018*) principles of providing more positive affirmation (the two bread sections) and less critical or negative comment (the filling in the middle), but that their effort goes into managing this imbalance rather than ensuring that both elements are focussed and will facilitate students' development.

The Stanford WISE model (Yeager *et al.*, 2013) specifies three elements to feedback: noting why the strengths are strengths; noting where improvements can be made; and, noting why the assessor is confident that the learner is capable of making the improvements suggested. Whilst uncovering characteristics, themes and labels from the data it transpired that 18.8% (n = 34) of the pieces of feedback included either one or two of the three characteristics of the Stanford WISE model of feedback. Assessors were using the strengths and improvement elements of Standford WISE but no one used the "capable of making these improvements" element. No assessors feedback had all three Stanford WISE characteristics; we therefore assume that the use of some Stanford WISE elements was unintentional. This left 81.2% (n = 147) of assessments receiving none of the Stanford WISE elements. Even if elements of Standford WISE were used unintentionally, this research found that when it was used it had a more positive effect on the students' attainment than when it was not used at all.

One of the key findings of this pilot study was that there was an increasing positive effect on grade improvement according to how many of the Stanford characteristics were present in the feedback between T1 (first assessment point) and T2 (second assessment point). Of the 18.8% (n = 34) who received one or more of the Stanford WISE feedback elements at T1, 50% (n = 17) moved into the next grade band at T2. Although this was not statistically significant, literature tells us that statistical significance should not be used as the sole test for real life significance. Henkel (1976) states that statistical significance is "...of little or no value in basic social science research, where basic research is identified as that which is directed toward the development and validation of theory" and McLean and Ernest (1998) note that statistical significance does not equate to practical significance. Any increase in grades apparently arising from unconscious use of a particular feedback model (in this case Standford WISE) is of practical significance and merits further research (Musgrove, *et al.*, in draft).

When Stanford WISE elements are used in the context of other findings from the data, this pilot study found key elements of feedback to be a potentially valuable to improving a student's attainment between the first and second assessment point. There were other positive aspects which aided an increase in grades, such as meetings being offered, structured feedback and focussed commentary. From the 18.8% (n=34), where two Stanford WISE characteristics of feedback were present, positive effect was greatest where additionally a tutorial, or discussion, had been offered and the feedback was focussed throughout. Whilst records were not kept of whether the tutorial offer was taken up or not, there was a clear amplification of increased performance with the inclusion of a tutorial/meeting offer. This finding links to earlier research supporting the idea of including a dialogue when providing feedback as it leads to better student attainment (Alfehaid, 2018; Beaumont, 2011; Blair, 2014; Boud and Molloy, 2013; Carless *et al*., 2011; Nicol and Macfarlane-Dick, 2006). It is also connected to improving students' self-regulative skills, especially when in a dialogue with their peers (Alfehaid, 2018; Blair, 2014; Carless *et al.*, 2011), meaning students are more likely to achieve a higher grade in the future (Carless *et al.*, 2011).

By contrast, where the feedback was poorly structured and included unfocussed comments, such as "Great work!", "well done" or "you can do better than this" there was either no, or even a negative, change in performance between T1 and T2. A study carried out at Liverpool John Moores University's architecture department asked students about the types of feedback which they found most helpful and constructive (Smith, 2019). Smith's findings cite student comments such as "… can be inconsistent … vague …" and that [written] feedback would be more useful if it "… clearly showed what went well … so then in future projects you know what to repeat." (Smith, 2019). In addition, Smith's study shows that individual tutorials were more than three times more helpful than their next nearest rival; students offered very clear statements such as "you can ask questions, unlike written feedback" and "it facilitates more directed discussion and questioning…" (Smith, 2019). Smith found that two of the most common barriers to understanding feedback were lack of opportunity for discussion and lack of direction on how to improve, both of which accord with our findings and the Stanford WISE model (Yeager et al., 2013).

Conclusion

This was a pilot project looking at the types of feedback given on a final year dissertation project module which had two points of assessment. T1 was an early assessment whereby the student had to write a report on what they intended to study, early findings and where they were heading with their dissertation and T2 was the actual dissertation project of 10,000 words. As this was historical data the project analysed the data that was present having had no influence on it, although one author was also one of the 25 assessors.

Assessors spend a significant amount of time on reading students' work and providing feedback so that a student can see a critique or commentary on the work they have produced. However, there is a misalignment between this being provided and it being used as a tool for learning and improvement. There are large differences in how students and staff view feedback (Adams, 2020), therefore, tutors must make students aware of when feedback is being provided (Adams, 2020) and students should increase their understanding of how feedback works in order to improve their feedback literacy for future success (Carless and Boud, 2018). Feedback only becomes useful when students use it to make improvements to their work or their learning strategies (Carless and Boud, 2018). Boud and Molloy (2013) suggest "that without understanding how feedback has been used, teachers are blind to the consequences of their actions and cannot act to improve learning". If students are unable to understand their feedback, they are unable to engage, and if feedback is not reflected on, or read, staff should not waste their limited time with providing written comments that will remain unused or misused by students (Price et al., 2011).

The model we propose consists of the positive aspects of feedback we identified and crucially the key addition of a face-to-face tutorial or meeting. Ajjawi and Boud's (2015) observation that there is an increasing need for feedback to be more of a social interaction between the tutor and the student, meaning that feedback becomes meaningful, understood and acted upon. Harvey (2019) discusses the benefits of tutorial discussions for feedback and benefits of marking and providing feedback face-to-face with students. Discussion between student and staff is essential as it allows for expectations to be regulated (Ajjawi and Boud, 2015; Orsmond and Merry, 2011), as a lack of dialogue leads to higher student dissatisfaction (Ali, 2018; Nicol, 2010) due to expectations not being met. Orsmond and Merry (2011) suggest a "lack of feedback dialogue means that students never become fully aware of the potential contribution of feedback to their learning and tutors never fully appreciate how their feedback is being used" meaning that students are not fulfilling their learning capabilities and staff feedback is not being used to its fullest extent. For those reasons, it is essential that the model includes a dialogue.

Our pilot model consists of the three main elements of the Stanford WISE model (Yeager *et al.*, 2013), specifically: noting why the strengths are strengths, noting where improvements can be made, noting why the assessor is confident that the learner is capable of making the improvements. We then advocate the additional pro-active offer of a face to face meeting/discussion, for example a tutorial or similar, to discuss the feedback. Our model actively eliminates unstructured, unfocussed critiques as these are unhelpful at best and can often have negative impact on attainment (Hattie and Timperley, 2007).

The next phase in the development of these preliminary findings is to test our new hypothesis on current students. The hypothesis will focus on Stanford WISE characteristics plus a personalised offer of face to face contact at T1 and test to see whether this has a significant impact on T2 attainment. Mackay *et al.*, (2019) "*concluded that assessment can act as a barrier between staff and students, especially where students are not given effective feedback*" (2019: 315). We assert that, by identifying a way of improving attainment as a direct result of our feedback, assessors will be better equipped to discuss models of feedback and resulting potential. This discussion should sit within enhanced professional development and should also be shared with their students, in order to increase transparency and fairness of assessment.

References
Adams, A. (2020) 'Student engagement with feedback and attainment :the role of academic self-efficacy', .*Assessment and Evaluation in Higher Education*, 45(2), pp. 317–329.
Ajjawi, R. and Boud, D. (2015) 'Researching feedback dialogue: an interactional analysis approach', *Assessment & Evaluation in Higher Education*, 42(2), pp. 252–265. https://doi.org/10.1080/02602938.2015.1102863
Alfehaid, L., Qotineh, A., Alsuhebany, N., Alharbi, S. and Almodaimegh, H. (2018) 'The Perceptions and Attitudes of Undergraduate Healthcare Sciences Students of Feedback: A Qualitative Study', *Health Professions Education*, 4(3), pp. 186–197. https://doi.org/10.1016/j.hpe.2018.03.002
Ali, N. (2018) 'Identifying predictors of students' perception of and engagement with assessment feedback', *Active Learning in Higher Education*, 19(3), pp. 239–251.
Austen, L. (2016) 'Balancing academic gatekeeping and interpersonal positioning: a qualitative analysis of written feedback to undergraduate students', *Practitioner Research in Higher Education,* 10(2), pp. 67-81.
Austen, L. and Malone, C. (2018) 'What students' want in written feedback: praise, clarity and precise individual commentary', *Practitioner Research in Higher Education*, 11(1), pp. 47-58.
Blair, A. (2014) 'Can dialogue help to improve feedback on examinations?', *Studies in Higher Education*, 39(6), pp. 1039–1054.
Bell, A.R. and Brooks, C. (2017) 'What makes students satisfied? A discussion and analysis of the UK's national student survey', *Journal of Further and Higher Education*, 42(8), pp. 1118-1142. 10.1080/0309877X.2017.1349886.
Ball, E. C. (2010) 'Annotation an effective device for student feedback: A critical review of the literature', *Nurse Education in Practice*, 10, pp. 138–143.
Bellon, J.J., Bellon, E.C. and Blank, M.A. (1991) *Teaching from a Research Knowledge Base: a Development and Renewal Process*. Facsimile edition. Prentice Hall, New Jersey, USA.
Bloxham, S. and Boyd, P. (2007) *Developing Effective Assessment in Higher Education: a practical guide.* Open University Press, England.
Bloxham, S. and Campbell, L. (2010) 'Generating dialogue in assessment feedback: exploring the use of interactive cover sheets', *Assessment & Evaluation in Higher Education*, 35(3), pp. 291-300, DOI: 10.1080/02602931003650045
Bloxham, S., Boyd, P. and Orr, S. (2011) 'Mark my words: the role of assessment criteria in UK higher education grading practices', *Studies in Higher Education*, 36(6). pp. 655-670. http://insight.cumbria.ac.uk/1179

Boud, D. and Molloy, E. (2013) *Feedback in higher and professional education: Understanding it and doing it well.* London: Routledge.

Beaumont, C., O'Doherty, E.M. and Shannon, L. (2011) 'Reconceptualising Assessment Feedback: A Key to Improving Student Learning?', *Studies in Higher Education*, 36(6), pp. 671-687.

Brookhart, S.M. (2008) 'How to give Effective Feedback to you Students*', Association of Supervision and Curriculum Development*, Alexandria, VA, USA.

Cambridge Online Dictionary https://dictionary.cambridge.org/dictionary/english/praise (Accessed: 08 February 2022).

Cambridge Dictionary (2019) https://dictionary.cambridge.org/dictionary/english/praise [accessed Cambridge Dictionary (2019) https://dictionary.cambridge.org/dictionary/english/praise (Accessed: 08 February 2022).

Carless, D. and Boud, D. (2018). 'The development of student feedback literacy: enabling uptake of feedback', *Assessment & Evaluation in Higher Education*, 43(8), pp. 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., Salter, D., Yang, M. and Lam. J. (2011) 'Developing sustainable feedback practices', *Studies in Higher Education*, 36(4), pp. 395–407.

Darwent, D. (2016) 'To praise or not to praise, is that the question?', Available at: https://blogs.shu.ac.uk/sioe/2016/06/21/to-praise-or-not-to-praise-is-that-the-question/?doing_wp_cron=1521187779.9995939731597900390625 (Accessed June 2019)

Darwent, D., Irons, C. and Taylor, L. (2018) The perilous praise burger. Available at: https://blogs.shu.ac.uk/sioe/2018/06/04/1516/ (Accessed: June 2019).

Deneen, C.C. and Brown, T.L.G. (2016) 'The impact of conceptions of assessment on assessment literacy in a teacher education program', Cogent Education, 3(1), Article 1225380. DOI: 10.1080/2331186X.2016.1225380

Dunworth, K. and Sanchez, H.S (2016) 'Perceptions of quality in staff-student written feedback in higher education: a case study', *Teaching in Higher Education*, 21(5), pp. 576-589. https://doi.org/10.1080/13562517.2016.1160219

Education Scotland (2022) Scottish Attainment Challenge: Available at: https://education.gov.scot/improvement/learning-resources/scottish-attainment-challenge/ (Accessed: 08 February 2022).

Harvey, T. (2019) **'***What went wrong? Students' and lecturer reflection on why face- to-face feedback was ineffective*. Carlisle: University of Cumbria.

Hattie, J.A.C. and Timperley, H (2007) 'The power of feedback', *Review of Educational Research*, 77(1), pp. 81-112.

Hawk, T. and Lyons, P. (2008) 'Learner's perceptions of care and respect offered by instructors', *Industrial and Commercial Training*, 40(4), pp. 197–205. https://doi.org/10.1108/00197850810876244

Henkel, R.E. (1976) *Tests of significance*. Newbury Park, CA: Sage.

Hennessy, C. and Forrester, G. (2014) 'Developing a framework for effective audio feedback: a case study', *Assessment & Evaluation in Higher Education*, 39(7), pp. 777-789.

Hesketh, E.A and Laidlaw, J. (2002) 'Developing the teaching instinct - 1: Feedback', *Medical teacher*, 24, pp. 245-8. 10.1080/014215902201409911.

Kauffman, H. (2015) 'A review of predictive factors of student success in and satisfaction with online learning', *Research in Learning Technology*, 23.

Jessop, T., El Hakim, Y. and Gibbs, G. (2014). 'The whole is greater than the sum of its parts: a largescale study of students' learning in response to different programme assessment patterns', *Assessment and Evaluation in Higher Education*, 39, pp. 73-88

Jones, C.A. (2005) '*Feedback for Learning*', Assessment for Learning: Vocational Learning for 16-19s. Institution for Learning and Skills Development.

Lund University (2013) 'School system favors pupils driven by worry and conscientiousness, Swedish study shows', ScienceDaily, 24 January 2013. Available at: www.sciencedaily.com/releases/2013/01/130124091534.htm (Accessed 07 February 2022).

Lunt, T. and Curran, J. (2010) '"Are you listening please?" The advantages of electronic audio feedback compared to written feedback', *Assessment & Evaluation in Higher Education*, 35(7), pp.759-769.

McArthur, J. and Huxham, M. (2013) Feedback unbound: From master to usher, in Merry, S., Price, M., Carless, D. and Taras M. (eds.) *Reconceptualising Feedback in Higher Education*. London. Routledge.

MacKay, J.R.D., Hughes, K., Marzetti, H. Lent, N. and Rhind, S.M. (2019) 'Using National Student Survey (NSS) qualitative data and social identity theory to explore students' experiences of assessment and feedback', *Higher Education Pedagogies*, 4(1), pp. 315–330. https://doi.org/10.1080/23752696.2019.1601500

McLean, J.E. and Ernest, J.M. (1998) 'The Role of Statistical Significance Testing in Educational Research', *Research in the Schools*, 5(2), pp. 15-22.

Mackieson, P., Shlonsky, A. and Connolly, M. (2018) 'Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis', Qualitative Social Work. Available at:, https://doi.org/10.1177/1473325018786996 (Accessed: 08 February 2022).

Merry, S. and Orsmond, P. (2008) 'Students' attitudes to and usage of academic feedback provided via audio files', *Bioscience Education*, 11(1), pp.1-11.

Musgrove, A., Darwent, D. and Baker-Green, K. (*in draft*) How can our feedback support improvement in students' grade: A case study.

Nicol, D. and 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), pp. 199–218. https://doi.org/10.1080/03075070600572090

Nicol, D. (2010) 'From monologue to dialogue: improving written feedback processes in mass higher education', Assessment and Evaluation in Higher Education, 35(5), pp. 501–517.

Orsmond, P. and Merry, S. (2011) 'Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback', *Assessment & Evaluation in Higher Education,* 36(2), pp. 125-136.

Patchan, M., Schunn, C. and Correnti, R. (2016) 'The Nature of Feedback: How Peer Feedback Features Affect Students' Implementation Rate and Quality of Revisions', *Journal of Educational Psychology*, 108(8), pp. 1098–1120 https://doi.org/10.1037/edu0000103

Price, M., Handley, K. and Millar, J. (2011) 'Feedback: focusing attention on engagement', *Studies in Higher Education*, 36(8), pp. 879-896.

Sadler, R. D. (1985) 'The origins and functions of evaluative criteria', *Educational Theory*, 35(3), pp. 285-297.

Schwartz, R. (2013) The "Sandwich Approach" Undermines Your Feedback. Available at: https://hbr.org/2013/04/the-sandwich-approach-undermin (Accessed: 08 February 2022).

Smith J. and Noble H. (2014) 'Bias in research', Evidence-Based Nursing 2014, 17, pp. 100-101 http://dx.doi.org/10.1136/eb-2014-101946

Smith, C. (2019) 'What do they do with my feedback?' Liverpool: Liverpool: John Moores University.

Sutton, P. (2012) 'Conceptualizing Feedback Literacy: Knowing, Being, and Acting', *Innovations in Education and Teaching International,* 49(1), pp. 31–40. DOI 10.1080/14703297.2012.647781

Talib N. and Sansgiry S.S. (2012) 'Determinants of academic performance of University students' Pakistan', *J Psychol Res,* 27(2), pp. 265–78.

Yeager, D. S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Brzustoski, P., Master, A., Hessert, W. T., Williams, M. E. and Cohen, G. L. (2013). 'Breaking the Cycle of Mistrust: Wise Interventions to Provide Critical Feedback Across the Racial Divide', *Journal of Experimental Psychology*, Advance online publication, Doi: 10.1037/a0033906.

Williams, J., Kane, D., Sagu, S. and Smith, E. (2008). Exploring the National Student Survey Assessment and feedback issues. Available at: http://www.heacademy.ac.uk/assets/York/documents/ourwork/research/NSS_Assessment_and_Feedback_ExecSummary_31.04.08.pdf (Accessed: 08 February 2022).

Informal Peer Learning of Diverse Undergraduate Students: Some Learners Make Meaning through Collaborative Activity

Russell Woodward*and Nicola Pattinson**
*The TEC Partnership, **University of Lincoln

Abstract

This paper explores student reported informal peer-learning in a business degree cohort at a UK community college. Literature on conceptualisation and debate regarding peer-learning is examined along with established typologies with respect to equality, mutuality and structure. Following this a critical review of evidence-based studies is undertaken. Crucially we propose a framework of conceptual, functional and experiential themes of peer learning, linked to aspects of student diversity in such courses/institutions. This framework is applied in our questionnaire study investigating the perceptions of second- and third-year students about peer learning. The main findings are that the thematic directionally distinct peer-learning seems not to be prevalent but that across themes and settings some students are more generally engaged in peer-learning than others. The content-based theme of two-way peer-learning that is strongest is guidance in making sense of subject literature; the weakest is guidance on environment aspects of business, e.g. economics. The settings of strongest peer-learning are reported to be those of in-class activities plus group assessment work. The main teaching recommendation is therefore greater deployment of in-class activities that require collaborative learning and require peer learning. The main recommendations for further research include more open explorations of student peer engagement where students can specify significant in-class events and experiences.

Keywords

Introduction

As noted by Boud (2001) a leading writer on peer-learning, if we as tutors set ourselves as the mediation point for all that students need to know, we are making a serious mistake. He goes on to assert that peer-learning in the educational setting is key to skill development as regards people effectively learning from each other in life and work, and as regards people learning practical discernment across information received through all sorts of channels.

There are many different forms of peer education including peer tutoring as covered by Duran (2017). While peer-learning itself is variable, as Boud (2001) asserts is essentially involves the sharing of knowledge and understanding between participants in a similar situation (i.e. cohort) who do not have a role as teacher or expert practitioner. Boud also asserts the reciprocal two-way nature of peer-learning.

In line with reasoning developed later, we propose in this study that both the value and the propensity for peer learning may be strong when there is balanced diversity in terms of socioeconomics, demographics and prior learning. We connect this to complementary differences in terms of relative preferences, strengths and weaknesses. For reasons more closely explored in the literature-based analysis, it is argued here that a degree cohort at a community college – and in particular a cohort studying a pre-management business degree is likely to be very diverse in these key respects. We also

Citation

propose themes of peer learning in this context which we term functional, conceptual and experiential.

With these proposals in mind, this study seeks to conduct an initial indicative piece of research on the extent and themes of informal peer-learning across a business studies degree cohort at a community college in the UK.

The literature-based analysis of this paper looks at the different depictions of peer-learning in context of the wider notion of peer education, as well as the theoretical underpinnings, findings and methods of existing studies in peer-learning. There is also assessment of education literature based reasoning for suspecting diversity and strong peer-learning potential/benefit in community college pre-management business degree cohorts.

Thereafter the study goes forward with primary research in the form of an embedded case study at a community college on a business studies degree cohort using an anonymised questionnaire tool – structured to discern and classify different themes of peer-learning both provided to and received from cohort classmates. Following analysis of these findings – there is discussion of possible strategies to more greatly facilitate and enable the informal peer-learning such as it is.

Literature Based Analysis

As recognised by Boud (2001) peer-learning can be very informal and incidental in its essential sense of individuals learning with and from each other. The key defining aspects, namely that the process involves status equals in similar situations where none/neither have a formal tutor or instructor position are agreed also by Topping (2005). It should be noted however that most treatments of peer-learning depict it as a learning strategy deployed or an actively implemented process, (Hammond et al, 2010). There are two main reasons for this. One, with there being anecdotal and experiential belief that some peer-learning occurs informally and aids student attainment, there is also conscious organisation to increase and enhance this. This aspect constitutes initial justification for distinct schemes such as PASS, Peer Assisted Study Schemes in Australia (Dancer et al., 2015), PAL and PALS, Peer Assisted Learning and Peer led Academic Learning Schemes respectively in the UK (Hammond et al, 2010; Keenan/HEA, 2014) and the numerous schemes originating and centred in the United States under the terminology SI or Supplemental Instruction (Capstick et al., 2004; Martin, 2008). The further motive for active strategies and formal schemes to be dominant in the literature is the reasonable desire to move from the anecdotal and experiential reflection to some kind of testable scenarios where the effects of peer-learning can be assessed. In particular Hammond et al. (2010) are critical of some of these positivistic efforts due to difficulties/issues in deriving separable intervention and control cohorts, and the possibility of confounding factors when comparing groups and their academic performance.

Within the broad coverage of peer education there are formats which deviate sizeably from the Boud (2001) status equal/no tutor/reciprocity depiction. Topping et al. (2017) classify different forms of peer education within the dimensions of Equality, Mutuality and Structuring with there being significant variations on all three. As noted by Falchikov (2001) a good deal of peer education is cross level peer tutoring whereby a student leader from a higher level helps in classroom-based activities. This approach is critiqued in practice with the mutual respect aspect being seen as stronger between authentic same level peers (Falchikov, 2001). The formally specified tutor aspect has also been critiqued in that the presence of a more advanced learner may still encourage a passive approach to learning (Ladyshewsky, 2000). There are of course some practical feature advantages where peer education involves a formally set tutor role. Looking across the different approaches surveyed by Duran (2017) including peer tutoring and peer-learning by teaching, an established peer tutor position can enable students so appointed to undertake planning both for quality of explanation and for formative questioning of the peer tutoring recipient to check for depth of learning. Having noted these

aspects – and their proximity to those of traditional teaching – this study focuses on the basic Boud articulated approach. This is partly because of the parity of respect and active engagement points made above, but also because this study seeks to ascertain – at least thematically – the kind of peer-learning that is going on informally. Therefore, in the classification of Topping et al. (2017), the peer-learning explored here, in so far as it is prevalent, involves: strong equality, between cohort classmates with none given a tutor role; sizeable mutuality as students may receive and provide peer guidance albeit in an ad hoc, non-synchronised and not strictly pair-wise way; and a low degree of structuring because of that informal ad hoc nature and no particular planning.

In terms of argued benefits of peer learning, a number of relevant and general points apply from the literature. In conceptual terms, Boud (2001) sees the benefit of reciprocal peer (same class) learning being that students are both learning from and contributing to each others' learning. Boud links this to Habermas' (1984) notion of an 'ideal speech act' wherein issues of power and dominance are not prominent. The notion of social interdependence in the work of Vygotsky (e.g. 1986) – linked to peer education by Dancer et al (2015) surely has its strongest applicability in the reciprocal peer-learning setting. The connection to constructivist and social constructivist learning is also important as this kind of peer-learning sees education being enhanced by social interaction and works on the premise that everyone has some existing relevant knowledge to contribute. Crucial in specific terms here is Vygotsky's notion of zones of proximal development in which student capability, given help (including peer help) can feasibly be extended. The reciprocal peer-learning setting is almost certainly also optimal for any community of learning benefits to come through and help new students cope with the challenges of higher education, as again related to peer education by Dancer et al (2015). It should be noted here, however, that the reciprocity of peer-learning is unlikely to be simultaneous in real time since some students are likely to be stronger than others depending on the topic being covered. This is an important practical point to recognise especially in view of the thematic nature of peer-learning investigation to be put forward in this study. In context of widening participation in higher education, Allan and Clarke (2007) cite Biggs (2003) as asserting a role for peer-learning in study skills accumulation linked to meta-cognition and more widely learning to learn. This is significant in terms of being generic across all curricular but student type specific relating to those whose prior study skills may have been limited. Another beneficial aspect reasoned in the literature is Topping's (2005) self-disclosure point whereby student willingness to admit ignorance or confusion and ask for guidance may be greatest in interaction with peers who hold no position of authority. It is noted that the argument presented here relates specifically to that context of status equals in informal settings.

In relation to longstanding core pedagogy, perhaps the most straightforward piece that indicates a reciprocal benefit for peer-learning is the Cone of Experience/Pyramid of Learning based on the work of Dale (1954) and the National Training Laboratories Institute (NTL, 1954). In these model depictions, while learning retention rates for listening to a lecture are portrayed as likely to be very low, explaining or teaching to others what has been covered is up there with practical use of the material as providing a very high retention rate. In the later work of Dale (1969) methods such as explaining to others are depicted as giving a far higher cognitive capability in the material (in the taxonomy of Bloom et al (1956)) than watching/listening to lectures.

In terms of empirical studies, evidence of reported effectiveness of informal peer-learning was gathered from students by Hossain and Sormunen (2019). In their research, numerous students from a Library and Information Science cohort reported that a great deal of their information and communications technology learning came informally from peers. Their quantitative and qualitative research – examining process within informal peer-learning – showed that the learning tended to take place when students asked peers for guidance – upon reaching practical obstacles in their computer use, when working on course subject assessments. The informal and status-equals aspect is important to note here. It is also recognised that while focussed on a particular subject cohort, the content

conveyed in the peer learning is of a generic cross-curricular nature. A relevant study using peer evaluation of business degree learning, though taken forward specifically to apply to group work and mentoring, is that by Brutus and Donia (2010). The category themes used in their research tool very much link to ways in which students can rate and detail how peers are contributing to their learning – namely: cooperation, practical contribution, conceptual contribution and work ethic. Peer learning flows along these themes could be reflective of diversity related differences in strengths, weaknesses and preferences. As such this categorisation aspect is very close to the one deployed in this study. The difference about this Brutus and Donia (2010) research is that it focusses on sessions explicitly dedicated to group working and mentoring. There is merit in studying the extent and nature of such processes during regular teaching and study activity, and that is undertaken in this paper.

There are a number of reasons for thinking business cohorts may be diverse such that complementary strengths, weaknesses and preferences exist. Broadly speaking it may be that very different people choose the pre-management business degree because business study itself is diverse on a number of meaningful criteria. Based on the work of McFarlane (1997) it is fair to say that business study has a sizeable academic content element alongside a crucially vocational format in its learning outcomes. Furthermore, reflections on business studies curriculum and teaching methods, such as that by Mulligan (1987) in dual cultures context, indicate the field to have a sizeable humanities content focus alongside deployment of scientific techniques for subject analysis. In view of these balance, dichotomous or mixture aspects a business management studies degree cohort is highly likely to comprise, on the one hand students interested in the conceptual and contextual, as well as a data analytical applied treatment of material, and on the other students interested in the functional practicalities as well as the people and creative side of business and management activity. This intuitive point also corresponds to empirical work finding very strong learning styles differences across business studies cohorts (Biberman and Buchanan, 1986; Loo, 2002). In so far as salient differences in interests, preferences and motives reflect different relative strengths and weaknesses, pre-management business studies cohorts may have a diverse and perhaps complementary range of aptitudes and deficiencies.

The higher education setting of degree delivery at UK community colleges brings further diversity to cohorts – business studies and otherwise. As noted in works by the Education and Training Foundation and the UK Government Department for Business, Innovation and Skills, community (often called further education) colleges are at the relative forefront of the UK's Widening Participation and Lifelong Learning Agendas, including with regard to higher education and especially comparative to universities (ETF, 2017; BIS, 2012). In specific terms this means that community college degree cohorts contain a distinctly higher proportion of students without traditional academic entry qualifications, students from lower socioeconomic groups, and students from families without relatives having participated in higher education (ETF, 2017; BIS, 2012). Further, the proportions of students over 25 are distinctly higher at UK community college higher education, and the intake of students with relevant vocational entry qualifications is also a salient feature of degree cohorts at these colleges (ETF, 2017; BIS, 2012). The combination of these distinct proportion aspects of degree cohort composition mean that the following can reasonably be said regarding community colleges. Mature students with working life experience but limited academic qualifications, as well as students, young and older, coming through from vocationally focussed post-compulsory business diploma courses e.g. BTEC (Business and Technician Education Council), are likely to be just as numerous in community college business degree studies classrooms as students coming straight from a successful sixth form e.g. A (Advanced) levels background having studied a range of academic subjects including humanities and social sciences in post compulsory education. As such the diversity of relevant prior exposure, in terms of real-life experience of organisational workplaces, business intrinsic functions study and engagement with conceptual learning is likely to be substantial and quite possibly systematic across a UK community college business studies degree cohort.

The diverse and possibly complementary nature of preferences, aptitudes and prior experiences across a typical pre-management business studies degree cohort at a UK community college means that reciprocal peer-learning between the classmates may be critical, but also feasible and potentially very beneficial. In so far as this diversity is prevalent, there may be scope for thematic enquiry as to the nature and extent of peer-learning that informally takes place.

In terms of a stylised categorisation: The peer-learning guidance chiefly coming from matures could be depicted as Experiential, while that from BTEC vocational diploma completers could be termed Functional, and the guidance from sixth form A level completers could be classified chiefly as Conceptual.

There is no guarantee that the thematic peer-learning depicted above will be taking place; the diversities may themselves also form obstacles to the process. Whether looked at as learning style variances or aptitude variances, e.g. between the abstract conceptual and the experientially practical, the potential for zones of proximal development to be too far removed from where peers are operating, for non-teacher students to guide each other, is a possible problem. The life experience differences may also obstruct social integration, including on self-disclosure, between sub-groups, thereby inhibiting sizeable peer-learning. In particular there is research (Chapman, 2013) finding non-traditional matures entering universities to be prone to 'imposter syndrome' in the depiction of Clance and Imes (1978) and an enduring self identification as being different from traditional HE students and belonging to a distinct community of practice in the model of Wenger (1998). It is hoped that in the community college small class setting with an even balance of student types as set out above, the facility and potential for authentic integration will be greater for the matures giving sizeable scope for peer-learning. Further, the non-residential nature of community college means that some of the university campus lifestyle oriented alienations (Chapman, 2013) e.g. the contrasting dispositions to student union drinking cultures – should not apply to the detriment of the classroom.

The course and institution factors for diversity linked to complementary strengths, weaknesses and preferences indicate that peer learning flowing along conceptual, functional and experiential themes may be strong in business studies at community colleges. Further, the obstacles to certain groups having a peer learning role in HE generally are reasoned to be minimised for an FE college setting. While across studies on peer learning, there has been attention to informal and status-equals contexts, widening participation cohorts, subject specific courses and thematic categorisations, no study has combined these aspects for focus on a community college business degree in relation to regular teaching and learning activity. This paper goes forward with indicative initial research along such lines.

Methodology
Without specific interventions and distinct control groups, ability to assess student outcomes gain from peer-learning is greatly limited in the informal setting. However, students' own views about the nature and direction of peer learning provided and received can be ascertained, be it asynchronous in practice or otherwise. As such, in terms of research philosophy, this study can reasonably be depicted as an interpretive, rather than a positivistic investigation.

Within investigation of the thematic scope of informal peer-learning, there needs to be some consideration of setting. Lectures, in so far as they are not overly formalised, may frequently involve peer questioning and peer explanation as topics and terms are covered. Again, turning to institutional comparison, BIS (2012) note that the main mode of delivery in community college higher education, is much more small group oriented and interactive than that in most universities.

Also important are activities including in class, some of which may be explicitly group organised while others may involve in class tasks in which discussion with classmates is allowed/encouraged during completion, ahead of feeding answers back to the tutor. There is also the setting of group assessment to which, as noted by Capstick et al (2004), peer-learning can be linked, explicitly or implicitly.

In Appendix 1 there is a questionnaire tool containing scale question statements relating to provision/contribution and receipt/benefit from and to classmates/peers of guidance/learning on the following themes:

- Conceptual and theory
- Real life experiences
- Functional areas
- Contextual aspects of business
- Making sense of subject literature

These questions are present in pairs on the questionnaire in the form:

I provide guidance to classmates on *theme*
Not at all   Little   Quite a bit   A great deal
I receive guidance from classmates on *theme*
Not at all   Little   Quite a bit   A great deal

This enables cross tabulation of responses for each individual student.

The questionnaire also has questions regarding settings in which peer learning may be taking place. In view of the fact that community college degree classes may not fit the discrete lecture, seminar, supervision format of established universities, these settings are referred to as phases in the questionnaire.

Again, these questions are present in pairs on the questionnaire in the form:
I provide guidance to my classmates during *setting/phase*.
Not at all   Little   Quite a bit   A great deal
I receive guidance from my classmates during *setting/phase*
Not at all   Little   Quite a bit   A great deal

In terms of research participants and sampling aspects, the student respondents came from years 2 and 3 of a university validated business undergraduate degree comprising three strands: Business with Marketing; Business with Accounting; Business with Organisation Behaviour, delivered at a UK community college. Invitation to participate and questionnaire distribution took place for both year groups when they were attending shared modules (taken by all strands). With regard to number and relevant diversity these cohorts taken together constituted 23 students, with institution data showing that:
- 9 students fitted the classification UK non-traditional mature (on age and level 3 post compulsory qualifications criteria)
- 6 students were of traditional undergraduate age with UK level 3, post compulsory qualification in A levels (academic subjects)
- 5 students were of traditional undergraduate age with UK level 3 post compulsory qualification in BTEC (vocational business diploma).
- 3 students were classified international

This profile does quite closely correspond to the even balance of student types as asserted above for community colleges degree education.

In terms of ethics, it was felt that the regular assurances of anonymity and confidentiality were particularly important in this study since questions especially about peer/classmate help received may be a sensitive esteem/status matter for some students. This primary research process including questionnaire tool, was put through and received ethical approval centrally at the institution concerned. The questionnaire itself does not include student classifying questions e.g. on age and prior education. In view of the small cohort size such questions would probably diminish confidence of anonymity despite assurances. Findings are viewed as being indicative on extent, themes and direction of peer-learning across groups that are heterogenous in the way described.

Limitations
Without age and prior education/experience classifying questions, this primary research can only yield indicative findings rather than effective verification on the issues.

Also, this research does not deploy intricate longitudinal sociometry as applied to the feedback aspect of informal peer-learning between teacher education students by Headington (2018). It is fair to say that this study is less comprehensive. Such a sociometric approach may ultimately be warranted with regard to community college business degree cohorts. At this stage an exploration of particular themes of peer-learning, through a relatively simple tool is viewed as useful for the pursuit of readily anonymised easy to view indicative findings. While the data strictly represents a snapshot, the questions do ask students to answer based on the experience throughout their degree cohort studies to date.

In terms of student clarity regarding the questionnaire: Although a period of explanation regarding themes and settings took place ahead of questionnaire distribution, this was not lengthy due to concerns of priming students to answer in a particular way. Despite the students being at years 2 and 3 on the degree and so having some experience of undergraduate study, we cannot rule out ambiguity or patchy understanding on the subtler themes and settings questions.

Results: Presentation and Comment
As noted above, twenty-three students were invited to take part and all twenty-three completed the questionnaire. The smallness of sample size along with the number of answer options mean that formal statistical analysis, e.g. Chi Squared, is not effective, but table presentation conveys key indicative themes.

With regard to results data presentation, responses to question pairs i.e. receive/provide guidance on each specific theme or setting/phase were cross tabulated into single table boxes. This gives an indication as to comparative balance in the following sense: On a theme/setting/phase where students are divided between strong net recipients and strong net provides of guidance, the most populated cells will be top right and bottom left corners. On themes/settings/phases where individuals are balanced between being providers and recipients of guidance, cell population will follow the top left to bottom right diagonal; top left dominant for minimal provider and recipient involvement, and bottom right dominant for sizeable provider and recipient involvement.

Table 1. Conceptual Guidance.

I receive conceptual and theory guidance from my classmates	I provide conceptual and theory guidance to my classmates			
		Not at all	Little	Quite a bit	A great deal
	Not at all	0	3	0	0
	Little	1	7	2	0
	Quite a bit	0	0	4	1
	A great deal	0	1	2	2

The notion of sizeable net recipients/providers in line with diverse complementary strengths, weaknesses and preferences, is not at all evident here. There is however an even balance between those reporting significant and not significant peer learning activity on conceptual and theory grounds, where significant peer learning involvement comprises provision and receipt of guidance.

Table 2. Real life experience guidance.

I receive real life experience guidance from my classmates	I provide real life experience guidance to my classmates			
		Not at all	Little	Quite a bit	A great deal
	Not at all	2	0	1	0
	Little	1	6	0	1
	Quite a bit	1	1	8	1
	A great deal	0	0	0	1

Again thematic net recipients and providers are not reported, though a good number report being recipients and providers. One aspect to note here, is that real life experience relevant to business can itself be diverse, e.g. as employee, product consumer, household budget manager. As such, the scope for individuals being both recipients and providers is reasonable.

Table 3. Functional Areas Guidance.

I receive functional areas guidance from my classmates	I provide functional areas guidance to my classmates			
		Not at all	Little	Quite a bit	A great deal
	Not at all	1	0	0	0
	Little	0	7	2	1
	Quite a bit	0	1	7	1
	A great deal	0	0	1	3

On this third theme there is also a lack of reported net recipient/provider status. There is slight balance in favour significant peer learning involvement, both providing and receiving, over not so.

Table 5. Guidance on the contextual aspects of business.

I receive guidance from classmates on the contextual aspects of business	I provide guidance to classmates on the contextual aspects of business			
		Not at all	Little	Quite a bit	A great deal
	Not at all	3	1	0	0
	Little	0	13	1	0
	Quite a bit	0	0	2	0
	A great deal	0	0	0	3

Here the majority are reporting as neither a significant recipient nor provider of peer learning on this theme. On the face of it this is quite surprising, suggesting that any prior learning on humanities and social science is not translating into insight bearing for business that is being shared. There is a possibility that the student respondents are not firmly clear on the meaning of business contextual aspects, in terms of linking to modules/topics like business environment.

Table 6. Guidance on making sense of the subject literature.

		I provide guidance to classmates on making sense of the subject literature			
I receive guidance from classmates on making sense of the subject literature		Not at all	Little	Quite a bit	A great deal
	Not at all	0	0	0	0
	Little	0	5	4	0
	Quite a bit	0	1	8	0
	A great deal	0	0	2	3

Here, most respondents are reporting significant involvement, with the majority as both recipients and providers.

Table 7. Guidance during activities phases of class.

		I provide guidance to my classmates during activities phases			
I receive guidance from my classmates during activities phases		Not at all	Little	Quite a bit	A great deal
	Not at all	0	0	0	0
	Little	0	5	0	1
	Quite a bit	0	1	12	1
	A great deal	0	0	2	1

Here the vast majority report being involved in peer learning, with involvement as both a recipient and provider. With this being a setting rather than a theme aspect, the net recipient/provider distinction is not expected.

Table 8. Guidance during lecture phases of class.

		I provide guidance to my classmates during lecture phases			
I receive guidance from my classmates during lecture phases		Not at all	Little	Quite a bit	A great deal
	Not at all	0	0	0	0
	Little	0	6	3	1
	Quite a bit	0	3	6	1
	A great deal	0	0	1	2

We note here that there is still a sizeable number actively involved in peer learning during lecture phases. This is significant in that while university lectures are distinct from seminars/classes with high numbers and chiefly teacher centred delivery, in FHE colleges, lecture phases occur in the same small class interactive setting as the seminars.

Table 9. Guidance during group assessment work.

	I provide guidance to classmates during group assessment work			
I receive		Not at all	little	Quite a bit	A great deal
guidance from	Not at all	0	0	2	0
classmates	Little	0	1	1	2
during	Quite a bit	0	1	10	1
assessment work	A great deal	0	0	1	4

Overall reported peer learning activity levels here are much higher, but the setting concerned is itself very much oriented to peer learning, so high numbers here accord with the mentoring and group work findings of Brutus and Donia (2010).

Discussion and Conclusion

A significant amount of peer-learning guidance is being reported as taking place.

The thematically distinct direction flows of peer-learning guidance, as depicted as likely from the literature and context-based analysis, do not seem to be happening, e.g. people just providing conceptual guidance while receiving experiential and functional guidance, and so forth; the lower left and upper right areas of the thematic question response tables contain virtually no students.

Within each theme there seem to be some people significantly involved in peer-learning exchange, providing and receiving, and some people less so. The upper left to lower right diagonal of the response tables contain the bulk of the students in most cases.

In thematic terms it is a bit stronger for functional and experiential peer guidance, than conceptual. Interesting also it is very low with regard to peer guidance on the contextual or environmental aspects of business. These like economics, politics and law represent subjects that are not active functions of business, but still have a relevant bearing. As such it may be that the delivery here is almost entirely driven by the tutor. The peer-learning is reported as more sizeable, though still chiefly two-way, with regard to guidance on making sense of the subject literature.

The reported peer-learning is especially prevalent in the activities' sessions within classes and during group assessment work; in the latter, there is only one student reporting no significant involvement – provision or receipt – in peer-learning guidance. The peer-learning is less prevalent during lecture phases of classes but even in those, there is a non-trivial amount being reported.

An indicative conclusion to draw is that the thematic cohort diversity of the business degree subject and the community college setting is in itself not driving the direction of the peer-learning guidance. One aspect of the community college setting which may be enabling peer-learning in generic terms, is the smallness of class size and as such the facility for activities to take place within classroom delivery rather than the teacher centred lecture to a large cohort. The general prevalence of peer-learning reported, and in particular that regarding making sense of the subject literature, does tentatively further justify the core assertion from Boud (2001) that learning beyond the tutor's actions and stipulations is significant. It should be once again noted that this prevalence is from the students' own point of view and has not involved specific formal peer-learning settings, nor a control group for which that peer-learning formal setting is not present. The sizeable though partial involvement in peer-learning across the board is encouraging in line with the cone/pyramid depictions which say that explaining to others is a strong factor in strengthening information retention and understanding, especially in that the vast majority are actively involved in peer-learning during group assessment

work. The appreciable minority reporting no significant involvement in peer learning, provision or receipt, in any other phase, is a concern.

In terms of educational recommendations, continuance and where possible expansion of in-class activities, where helping each other is specified, on community college business degrees is greatly encouraged. Linking to the prevalence of peer learning on making sense of subject literature, some of these activities could include business article/report comprehensions. Linking to the lack of peer-learning regarding business environment context subjects, it is also encouraged that in class-group activities, connecting the topics of economics and law to business functions/problems, are made prevalent.

Turning to recommendations for further research, noting that the student participants had received part of their learning within the period of the pandemic, it would perhaps be interesting to investigate how much, during lockdown phases, peer-learning has taken place online between classmates when working on assessments, this may be non-synchronous e.g. questions and guidance on social media or specific forums.

In view of the potential for ambiguity around student appreciation of meaning, particularly of conceptual/contextual aspects of business study, there may be merit in having some two-way dialogue within or ahead of the primary research aspect itself, such that students can highlight any confusion and have it addressed, ensuring common understanding between researchers and participants. The value of this in terms of authentic findings may outweigh any tutor-student anonymity assurance in the data collection.

Reflecting again on study limitations, it is noted that the research undertaken was quite restricted, in terms of specifying particular themes and settings for peer-learning responses within studies, and not giving students the ability to highlight phases or events that they thought relevant. In view of this it may be beneficial to deploy an adapted version of Brookfield's (2017) Critical Incident Questionnaire, whereby students would be enabled themselves to pick out phases or aspects within their studies where they have felt most peer engaged/peer distanced and where they have felt most affirmed/helped or puzzled/confused from peer interaction. Analysis of evidence gained that way would of course be more qualitative.

References:
Biberman, G., and Buchanan, J. (1986) 'Learning Style and Study Skills Differences across Business and other Academic Majors', *Journal of Education for Business*, 61, pp. 303-307.
Biggs, J. (2003) Aligning Teaching for Constructing Learning. York: Higher Education Academy. pp.1-4.
BIS (2012) Understanding Higher Education in Further Education Colleges. BIS Research Paper No 69. London: Department for Business, Innovation and Skills.
Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D. R. (1956) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain.* New York: David McKay.
Brookfield, S.D., (2017) *Becoming a Critically Reflective Teacher*. New Jersey: John Wiley & Sons.
Boud, D. (2001) 'Making the Move to Peer-Learning', in Boud, D., Cohen, R. and Sampson, J. (eds.) *Peer-Learning in Higher Education: Learning from and with each other*. London: Kogan Page (now Routledge). pp. 1-20.
Brutus, S., and Donia, M. B. L. (2010) 'Improving the Effectiveness of Students in Groups with a Centralized Peer Evaluation System', *Academy of Management Learning and Education*, 9(4), pp: 652-662.

Capstick S, Fleming H and Hurne J (2004) Implementing Peer Assisted Learning in Higher Education: The Experience of a New University and a Model for the Achievement of a Mainstream Programme, in *Peer Assisted Learning Conference Proceedings*, 2004, Bournemouth University, UK.

Chapman, A. (2013) A (Re) 'Negotiation of Identity: from 'Mature Student' to 'Novice Academic', *Widening Participation and Lifelong Learning*, 14(3), pp. 44-61.

Dale, E. (1954) *Audio-Visual Methods in Teaching*. 2nd Edition. New York. Dryden Press.

Dale E. (1969) *Audio-Visual Methods in Teaching*. 3rd Edition. New York: Holt, Rinehart & Winston.

Dancer, D., Morrison, K., and Tarr, G. (2015) 'Measuring the Effects of Peer-Learning on Students' Academic Achievement in First Year Business Statistics', *Studies in Higher Education*,40(10), pp: 1808-1828.

Duran, D. (2017) 'Learning by Teaching. Evidence and Implications as a Pedagogical Mechanism', *Innovations in Education and Teaching International*, 54(5), pp: 476-484.

ETF (2017) College Based Higher Education. Preston: Education and Training Foundation.

Falchikov, N. (2001) *Learning Together: Peer Tutoring in Higher Education.* London: Routledge Falmer.

Habermas, J. (1984) *The Theory of Communicative Action*, Vol. 1 (trans. T. McCarthy). London: Heinemann).

Hammond, J. A., Bithell, C. P., Jones, L., and Bidgood, P. (2010) 'A First Year Experience of Student Directed Peer Assisted Learning', *Active Learning in Higher Education*, 11(3), pp. 201-212.

Headington, R. (2018) 'Students' Informal Peer Feedback Networks', *Practitioner Research in Higher Education: Special Assessment Issue*, 11(1), pp 4-14.

Hossain, M., and Sormunen, E. (2019) 'Does Informal Peer-Learning Stimulate Acquisition and Development of ICT Skills?', *Education for Information*, 35(2), pp, 129-148.

Keenan, C., (2014) Mapping Student-Led Peer-Learning in the UK. York: Higher Education Academy.

Ladyshewsky, R. (2000) 'Peer Assisted Learning in clinical education: A review of terms and learning principles', *Journal of Physical Therapy Education,* 14(2), pp. 15–22.

Loo, R. (2002) 'The Distribution of Learning Styles and Types for Hard and Soft Business Majors', *Educational Psychology*, 22, pp. 349-360.

Macfarlane, B. (1997) 'In Search of an Identity: Lecturer Perceptions of the Business Studies First Degree', *Journal of Vocational Education and Training*, 49(2), 5-20.

Martin, D. (2008) 'Foreword', *Australian Journal of Peer-Learning,* 1(1), pp. 1–3.

Mulligan, T. M. (1987) 'The Two Cultures in Business Education', *Academy of Management Review,* 12(4), pp.593-599.

NTL (1954) *The Learning Pyramid*. Bethel. Maine: National Training Laboratories Institute for Applied Behavioural Science.

Topping, K. J. (2005) 'Trends in Peer-Learning. *Educational Psychology',* 25(6), pp: 631-645.

Topping, K. J., Buchs, C., Duran, D., and van Keer, H. (2017) *Effective Peer-Learning: from Principles to Practical Implementation*. Abingdon: Routledge.

Vygotsky, L. S. (1986) *Thought and Language*. Cambridge MA: MIT Press.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Wenger, E. (2000) 'Communities of Practice and Social Learning Systems', *Organisation,* 7(2), pp. 225-246.

Appendix 1: Peer Learning Questionnaire

This questionnaire is part of a Research Project on capturing and exploring informal peer learning in diverse groups, focussing on a business degree cohort at a UK community college. The data obtained from this questionnaire will be used to help achieve that research aim.

Please answer the questions honestly. The information you provide is anonymous and will be treated as strictly confidential. By completing this questionnaire you are giving your consent for your information to be used for the above research purposes. The information you provide will be held under the terms of the latest Data Protection legislation. It should further be noted that following submission of the paper for publication all such data will be destroyed. If you wish to pass over some questions or only partially complete the questionnaire that is fine and it is still valid for you to submit the questionnaire.

Thank you.

Russ Woodward
Nicola Pattinson

These questions relate to your entire studies experience on the business degree and is in no way focussed just on the covid online phases.

In all questions, please circle the response that most closely reflects your honest view on the applicability of the statement concerned.

Question 1.
A) I receive guidance from my classmates on theory and conceptual aspects of business studies
Not at all   Little   Quite a bit   A great deal

B) I provide guidance to my classmates on theory and conceptual aspects of business studies
Not at all   Little   Quite a bit   A great deal

Question 2.
A) I receive guidance from my classmates on real life aspects of business studies
Not at all   Little   Quite a bit   A great deal

B) I provide guidance to my classmates on real life aspects of business studies
Not at all   Little   Quite a bit   A great deal

Question 3.
A) I receive guidance from my classmates on the functional module areas of business studies, e.g. marketing, accounts, people management
Not at all   Little   Quite a bit   A great deal

B) I provide guidance to my classmates on the functional module areas of business studies, e.g. marketing, accounts, people management
Not at all   Little   Quite a bit   A great deal

Question 4.
A) I receive guidance from my classmates on the contextual module areas of business studies, e.g. economics, politics/government and law.

Not at all   Little   Quite a bit   A great deal
B) I provide guidance to my classmates on the contextual module areas of business studies, e.g. economics, politics/government and law.
Not at all   Little   Quite a bit   A great deal

Question 5
A) I provide guidance to my classmates during lecture phases of business classes
Not at all   Little   Quite a bit   A great deal

B) I receive guidance from my classmates during lecture phases of business classes
Not at all   Little   Quite a bit   A great deal

Question 6
A) I provide guidance to my classmates during activities phases of business classes
Not at all   Little   Quite a bit   A great deal

B) I receive guidance from my classmates during activities phases of business classes
Not at all   Little   Quite a bit   A great deal

Question 7.
A) I receive guidance from my classmates during group assessment work
Not at all   Little   Quite a bit   A great deal

B) I provide guidance to my classmates during group assessment work
Not at all   Little   Quite a bit   A great deal

Question 8.
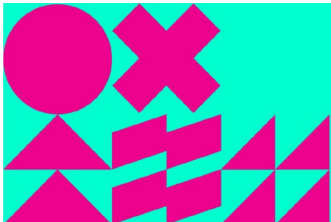   A)   I receive guidance from my classmates on how to make sense of the course literature
Not at all   Little   Quite a bit   A great deal
    B)   I provide guidance to my classmates on how to make sense of the course literature
Not at all   Little   Quite a bit   A great deal

PRACTITIONER RESEARCH
IN HIGHER EDUCATION

University of CUMBRIA