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

Practitioner Research
In Higher Education

Issue 15(1) pages 24-35

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

Citation
Ellison, C. (2023) ‘Effects of adaptive comparative judgement on student engagement with peer formative feedback’, *PRHE Journal,* 15(1), pp. 24-35.

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.

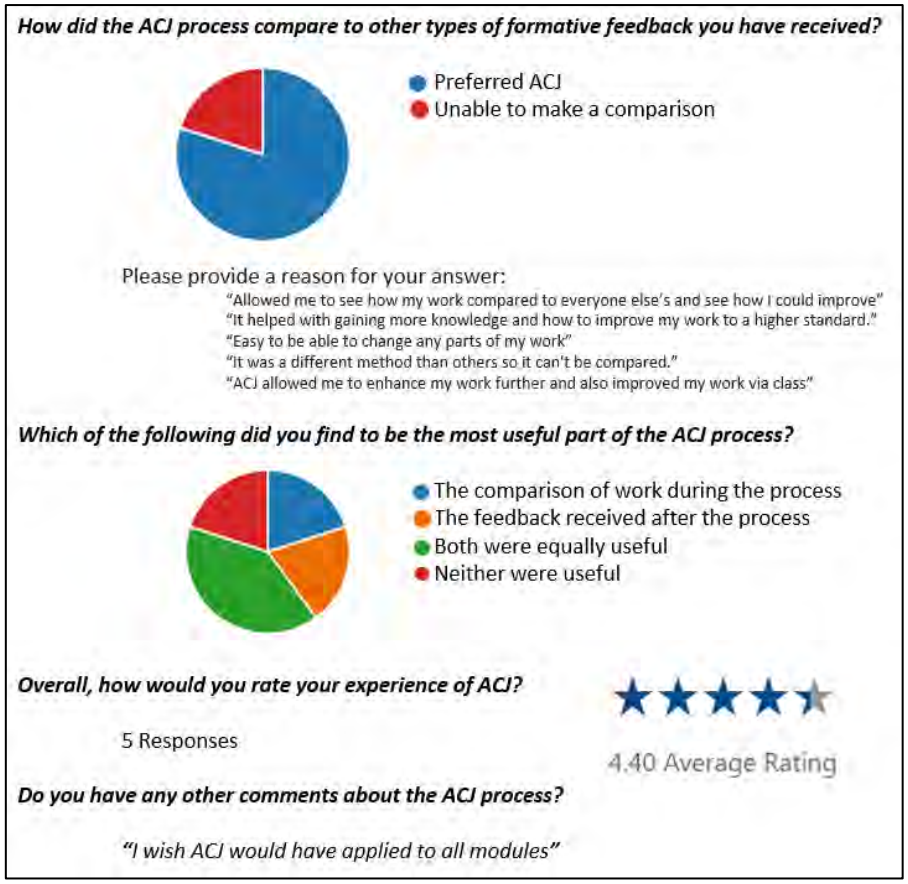

Figure 1. Key questions and answers relating to student engagement with adaptive comparative judgement.

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.

Acknowledgements
The funding of the Salford Community for Excellence in Learning and Teaching (SCELT) Scholarship is gratefully acknowledged, along with the mentorship of John Dermo. Use of the RM Compare software as part of the Early Adopter scheme is also gratefully acknowledged.

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