ENHANCED ASSESSMENT AND LEARNING THROUGH ADAPTIVE COMPARATIVE JUDGEMENT

Tony Topping, Matt Murphy, Samuel Saunders

School of Engineering, University of Liverpool
The Centre for Innovation in Education, University of Liverpool

ABSTRACT

The School of Engineering, University of Liverpool is trialling the use of Adaptive Comparative Judgement (ACJ) as tool to allow students to engage in a peer learning and assessment process. By examining three case studies, this paper describes how ACJ has been used in undergraduate engineering modules to enhance the assessment process for learning. The first case study in this paper describes how ACJ was used as an approach to provide feed-forward in a conceptual engineering design activity. It explores how ACJ has been used to facilitate a more structured approach to reviewing and contextualising examples of previously submitted work, in order to help students properly understand assignment expectations; and to provide some form of feed-forward. The second case study describes how ACJ is being used to facilitate peer assessment with the aim of increasing learner benefit. It was used to facilitate and inform peer discussion in an engineering ethics topic; and to help students reflect on their own attitudes and learning. The third explores how ACJ can be used to inform the final grade of a summative assessment and how the outcomes of completing an ACJ activity compare between students and teachers. The paper presents student perceptions and the authors experiences of using ACJ and discusses how such activities have been structured. Although the overall findings indicate that ACJ can be of benefit to teaching, more work is needed to optimise its deployment.

KEYWORDS

Formative Assessment, Peer Assessment, Peer Learning, Adaptive Comparative Judgement, Standard 11

INTRODUCTION

A key goal of a teacher is to foster an environment where students are able to engage in deep learning (Biggs & Tang 2011). In an effort to expand on current teaching strategies, The School of Engineering is exploring how to engage students in a peer learning and assessment process, the benefits of which are well documented (Topping 1998, 2009). Integrating a peer learning and assessment approach into an activity can support the development of skills such as collaboration and teamwork along with building confidence in the ability to communicate and externalise concepts to peers (Tanner et al. 2019).

Hansson et al. (2011) acknowledge that peer learning is often used to increase students' levels of deep learning, with their study concluding that it is an effective way to improve skills and that students participating in the activity feel more involved.

Although Hermon and McCartan (2011) note that there can be some issues with peer assessment, particularly with students inexperienced with the process, their study shows that helping students to develop self and peer assessment skills is worthwhile and that students report an increase in motivation just by participating in the process. A study by Jonsson et al. (2022) shows that students gain inspiration form viewing peer's work and find it useful to see how others solved the same problem, also noting that students found it useful to have input from someone other than just their supervisor. Peer assessment can also be a way to leverage the final piece of work and extend the learning of an activity after submission. As Bartholomew et al. (2022) points out, by not participating in the assessment of their work, students miss out on the benefits of applying higher order thinking skills such as critical evaluation and synthesis.

It is clear that a peer assessment and learning process can bring added benefit to an activity. But these strategies can be difficult to plan and implement with large cohorts and small teaching teams; strategies that employ deep learning often rely on small group teaching and intensive contact between students and teachers (Poot et al. 2020). This paper describes how The School of Engineering is implementing peer assessment and learning elements to a number of existing activities, using RM Compare as a tool to facilitate the process and to overcome some of the issues that come with managing larger cohorts.

PROPOSED INTERVENTION - RM COMPARE

RM Compare is an online digital tool that uses Adaptive Comparative Judgement (ACJ) to facilitate flexible approaches to both peer learning and peer assessment. ACJ is based on the work of Thurstone (1927) and later Pollitt (2004, 2012) who argue that making comparative holistic judgements, rather than using rubric based grading, can produce better results when evaluating the quality of work. To use the tool, work is uploaded to the platform before individuals are invited to participate in a judging session. Participants are presented with two pieces of work and asked to judge which they think best meets the assessment criteria; usually a single holistic statement. This pairwise comparison is repeated by different judges against different pieces of work over a number of rounds. As the judging rounds progress and a rank order of work begins to form, the tool will start to present participants with pieces of work that are closer in quality i.e., they have similar rank scores determined by judgements from previous rounds. This adaptive comparison can speed up the judging process and improve the overall reliability of the rank order (Pollitt 2012).

Whilst this tool can be used by instructors to quickly and reliably assess students work or as a tool to aid moderation and standardisation of grades, of particular interest to the authors is the tool's ability to facilitate deep learning through critical evaluation. By asking students to use RM Compare and participate in judging sessions, it is proposed that the tool can bring added benefit to existing activities; especially as the tool can be used regardless of cohort size.

Assignments that have open ended solutions can often be challenging for students as there is no single correct answer and it is not always clear what is expected of them to succeed. It is common practice to expose students to exemplar work from previous cohorts to help them better understand the task, but students might not properly engage; it can be difficult for students to internalise the criteria for quality just by viewing alone. Several studies have shown that using ACJ to facilitate feedforward or as a learning by evaluation tool can be beneficial; it promotes student growth and learning and can increase student understanding of assessment criteria (Bartholomew et.al. 2022).

Kimbell (2018) notes that by evaluating peers work, students crystalise their own learning and improve their understanding of the difference between a good piece of work and the best piece of work. Bartholomew et al. (2019) add to this argument by reporting that students found it helpful to receive feedback from an ACJ process as it gave insights into where exactly improvement was needed. Seery and Canty (2017) describe how the exposure to peers' work can help a student position their own work relative to another's and that viewing the solutions formulated by peers to the same task gives insights to approaches and concepts that may not have been considered previously. Bartholomew et al. (2022) agrees that ACJ is an ideal way to provide an opportunity for students to be exposed to new ideas; a critical element in an engineering design activity and many CDIO themed activities.

Of particular interest to the authors, is a study by Canty (2012) that describes how during a design activity a student felt that they were "blinkered by one idea and missed out on a chance to be really creative" This is an attitude often encountered by the authors in their own engineering design activities but as Canty (2012) argues, the use of ACJ can be more critical than the task itself; the student can learn more by viewing their peer's concepts and approaches, even if they feel that their own design was lacking in comparison. Indeed, Bartholomew and Yoshikawa (2018) have suggested that ACJ is particularly suited to open ended problems and problem-based learning activities that are common in CDIO subjects, with Tanner et al. (2019) concluding that ACJ is compatible with the CDIO initiative and brings several benefits similar to the ones described above.

Much of the literature around ACJ, some of which is cited above and in particular a study by Kimbell (2022), report that it can be a reliable way to grade a piece of work. Given the time pressures that come with grading large volumes of work in large cohorts, the authors decided to use this pilot study to also explore how the rank order produced from students participating in an ACJ judging session could be used make their assessment more efficient.

APPROCHES & RESULTS

In order to investigate the possible benefits of using ACJ to enhance learning, the School of Engineering piloted its use in the academic year 2022-23 in two modules; '*ENGG111: Professional Engineering - A Skills Toolkit'*, a first-year engineering skills module (n=220) and; '*MECH212: Engineering Design'*, a second-year engineering design module (n=170). After the students had completed the activities in case studies 1 and 2, they were invited to complete a survey consisting of a mix of Likert scale and open ended questions. The survey invitation was sent to 390 students and had a response rate of 14% (54 students).

Case Study 1: Peer Review for Formative Learning

ENGG111 is a wide ranging first year engineering skills module where a new engineering ethics activity has been introduced (worth 1.5 credits). The first part of this activity introduces students to Equality, Diversity and Inclusion (EDI) topics; providing an awareness of the importance of EDI to the engineering profession and; as a foundation to activities in later years of study. ACJ was deployed here as a tool to allow students to participate in a peer assessment activity; the process not only requiring students to critically evaluate their peers work but also allowing them to be exposed to the views and experiences of the whole class, an important element to any EDI activity (Florian and Pratt 2015).Students worked in groups of six to create an infographic that demonstrated their understanding of EDI topics; first carrying out some research into EDI and then holding group discussions on infographic content.

After students submitted their EDI posters, the judging session was introduced to students during a lecture; an overview of how ACJ works was given and; a rationale provided for the activity. Students were given a document instructing them on how to join the session and how to use the RM Compare interface. Each student viewed approximately 5 pairs of work, judging them against the holistic statement *"Which of these 2 posters best improves your understanding of EDI?"* Participation in the task was mandatory; a 50% penalty would be applied to the final grade of anyone who did not participate.

Engagement with this activity was high, 82% of the class participated, likely because of the grade penalty for non completion. When respondents to the survey were asked if they had a better understanding of the EDI topics after judging other groups posters, 15.2% 'Strongly agreed', 45.7% 'Agreed', 26.1% were 'Neutral', 10.9% 'Disagreed' and 2.2% 'Strongly disagreed'. This indicates that there is some benefit to having students judge other groups work and the continued development and use of ACJ would be worthwhile. Whilst somewhat effective in helping learners improve their understanding of a topic more generally, it is clear that ACJ can help a student understand the quality of their own work; when asked if the activity had helped them to better understand the quality of their own work, 43.5% 'Strongly agreed', 45.7% 'Agreed', 8.7% were 'Neutral' and 2.2% 'Disagreed'. When respondents were asked if viewing and judging other groups posters with RM Compare would be better than just viewing posters in an exhibition, 35.8% 'Strongly agreed', 34.8% 'Agreed', 23.9% were 'Neutral', 4.3% 'Disagreed' and 2.2% 'Strongly disagreed'. This is an interesting result with implications on the design of other poster activities. The authors have experienced issues in the past with the logistics and cost of organising poster exhibitions, particularly with large cohorts. This result suggests that using RM Compare to facilitate an engaging poster exhibition may offer a solution to these issues.

Case Study 2: Feed-Forward in an Engineering Design Activity

MECH212 is a second year engineering design module where students work in groups of six to follow a formal design process from first brief; through problem definition, background research, and product design specification; to conceptual design; then fully embodied 3D CAD and manufacturing pack. ACJ was deployed here as a formative, feed-forward exercise to help students to properly understand the task; understanding what exactly makes a piece of work successful and; preventing them from taking the wrong approach to completing it. Mid-way through the module, students are tasked with producing a concept design poster that demonstrates their design thinking; showing details of how they arrived at a final concept via initial concepts, concept selection and concept development. At the start of the poster assignment process, students were invited to judge and compare all the posters from the previous year's assignment. The judging session was introduced to students during a lecture; an overview of how ACJ works was given and; a rationale provided for why they should participate i.e. *by participating you will be more likely to produce a better poster and score a higher grade for the task.* Students were given a document instructing them on how to join the session and how to use the RM Compare interface. Each student viewed approximately 10 pairs of work, judging them against the holistic statement *"Which of these 2 posters best describes the evolution of the Concept Design?"* Participation in the task was not linked to the final grade and no extra credit was awarded for completion.

Engagement with this activity was much lower than the activity in case study 1, only 26% of the class participated. This is likely due to there being no grade penalties for non completion in this activity.

When respondents from to the survey were asked if they had a better understanding of the poster assignment after judging last year's posters, 25% 'Strongly agreed', 50% 'Agreed' and 25% 'Disagreed'. This indicates that students do gain benefit from using ACJ to facilitate a feed-forward activity. Figure 1 shows how the final group grades differed depending on whether group members completed the feed-forward activity. These results show that by completing the feed-forward activity, a groups work will be of higher quality. However, it is noted that at least half of the group must have completed the activity to gain this benefit. This is likely because, in the authors experience, student approaches to group work tends to be siloed; they complete their own sections of work before compiling it at the last minute. Another interpretation of these results is that the type of student who is likely to produce a good piece of work is more likely to complete this type of activity, hence the correlation between ACJ completion and high grade. When asked if the activity had helped them to better understand the quality of their own work, 62.5% 'Strongly agreed', 12.5% 'Agreed', 12.5% 'Disagreed' and 12.5% 'Strongly disagreed'. This follows on from the findings in case study 1 that ACJ is beneficial to student understanding of the overall quality of their work.

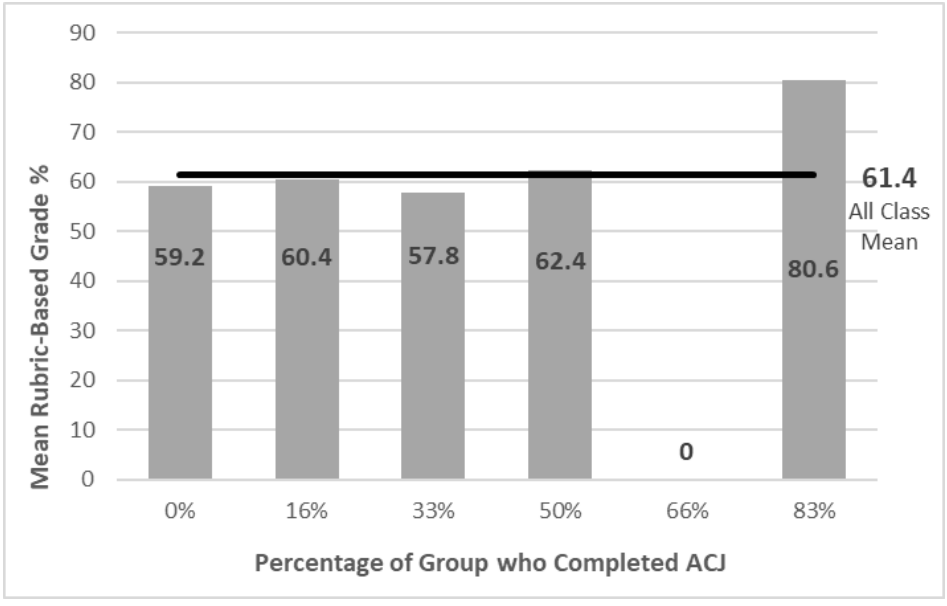

Figure 1. The relationship between the percentage of group members that completed the ACJ activity and that group's mean grade.

Case Study 3: Summative Peer Assessment

As a first step in exploring the relationship between rubric-based grades and grades based on ACJ ranking, a direct comparison was made between the instructor rubric-based grades given to concept design posters from last years cohort and the ACJ rank orders for the same posters produced by this years cohort during the feed-forward activity. The teaching team (n=5) that graded last years posters using a rubric were also asked to grade the same posters again using ACJ. In order to make a direct comparison between ACJ rank order and the rubric-based grade, a rank order of the rubric-based grades was produced. The grade for each poster was then placed in descending numerical order. This rank order of grades was then transposed directly to the ACJ rank order; essentially, awarding the top ACJ ranked poster the same grade as the top ranked rubric-based graded poster and so on.

Figure 2 shows the rubric-based grade of last years posters in comparison to the ACJ ranks produced by the same teachers who graded the posters and the rank produced by students who completed the feed-forward exercise. There is broad correlation between ACJ rank orders and rubric-based grades. The correlation between teacher rubric-based grade and teacher ACJ ranking shows that ACJ may be a reliable alternative method for teachers to grade work. Depending on the type of assignment and the number of pieces of work being graded, ACJ has the potential to reduce the amount of time taken to grade an assignment. For example, grading all the posters took around 3 hours using a rubric but took each judge around 30 minutes to complete the ACJ exercise. The correlation between student ACJ ranking and teacher ACJ ranking offers evidence that student rankings can be reliable; given that some students seemed distrustful of student rankings, this is an important finding and will be used to build trust in future activities. When respondents to the survey were asked if they would be comfortable with RM Compare being used to help inform the final grades in future assignments, 20.4% 'Strongly agreed', 31.5% 'Agreed', 31.5% were 'Neutral', 7.4% 'Disagreed' and 9.3% 'Strongly disagreed'. When asked if they trusted their classmates to use RM Compare properly, 13% 'Strongly agreed', 25.9% 'Agreed', 29.6% were 'Neutral', 22.2% 'Disagreed' and 9.3% 'Strongly disagreed'. Some of the larger differences between teacher grade and ACJ rank order were examined; it appears that when using a rubric-based grade system, students were graded more harshly i.e. whilst some posters were good overall, students lost points because they had misunderstood the exact criteria requirements for a single section of the task. Given that the purpose of the assignment is to allow students to demonstrate their *whole* design story, the holistic approach to grading used in ACJ may be more appropriate.

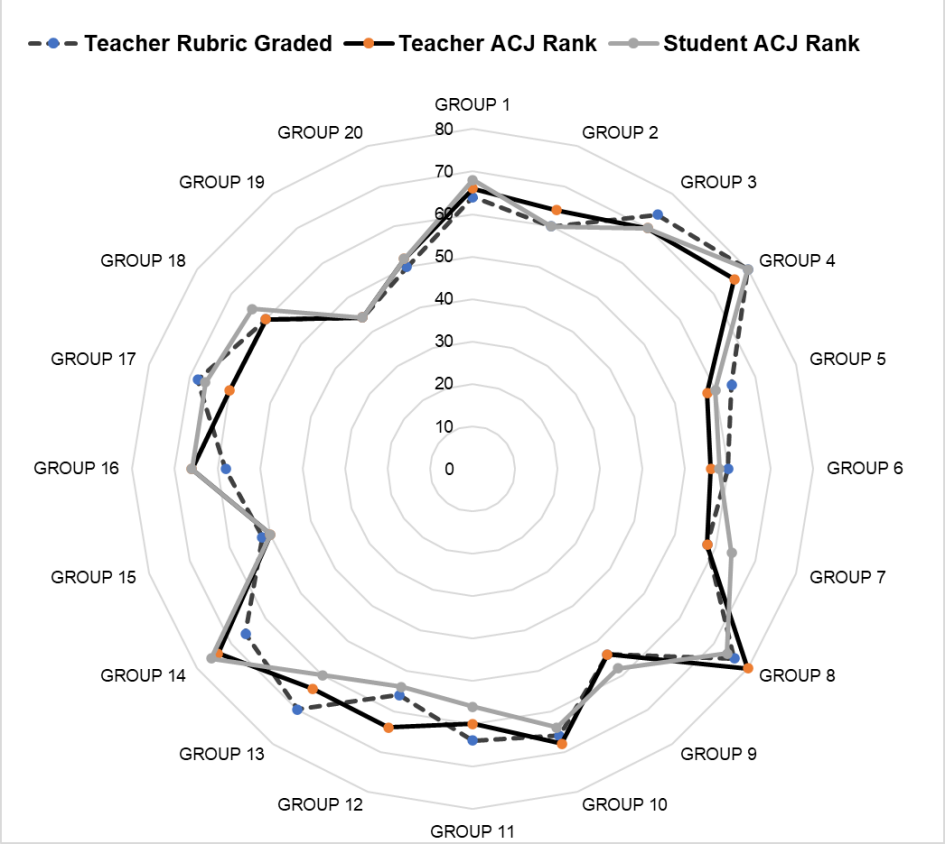

Figure 2. The relationship between final teacher rubric-based grade and the comparable grades created with the ACJ rank order produced by teachers and students.

Further Investigation: Using ACJ To Enhance Instructor Rubric-Based Assessment

After noting the correlation between grading shown above, the authors wanted to further explore how ACJ could be used in the grading process, asking the question "*Would it be possible to reduce the amount of time instructors spend grading student work by only asking instructors to grade a sample of work, and then using the output of a student ACJ session to generate grades for the whole class,?"* To investigate this, the instructor rubric-based grades given to concept design posters produced by last years cohort was taken as the starting point. This data is plotted in figure 3 as 'Rubric Grade'. Next, an ACJ informed grade was created, plotted in figure 3 as 'ACJ Grade'. To create the ACJ informed grade a plot was first created that described the relative position of each piece of work. This plot was created using the "Parameter Value", an output of the ACJ calculations assigned to each piece of work and used to determine the rank order. The top, middle and bottom rubric-based grades where then used to calibrate the upper, middle and lower limits of the new grade before extrapolating the grades between these limits using the Parameter Value curve. Figure 3 compares the spread between these two different grades showing good correlation between the two. It could have been possible to apply the grades in a linear fashion, as was done in the Summative Peer Assessment study above, but this would not have allowed for the relative differences in quality between adjacently ranked work to be demonstrated. The Parameter Value produced by ACJ does account for this, so would seem sensible to make use of it to produce a more realistic grade distribution.

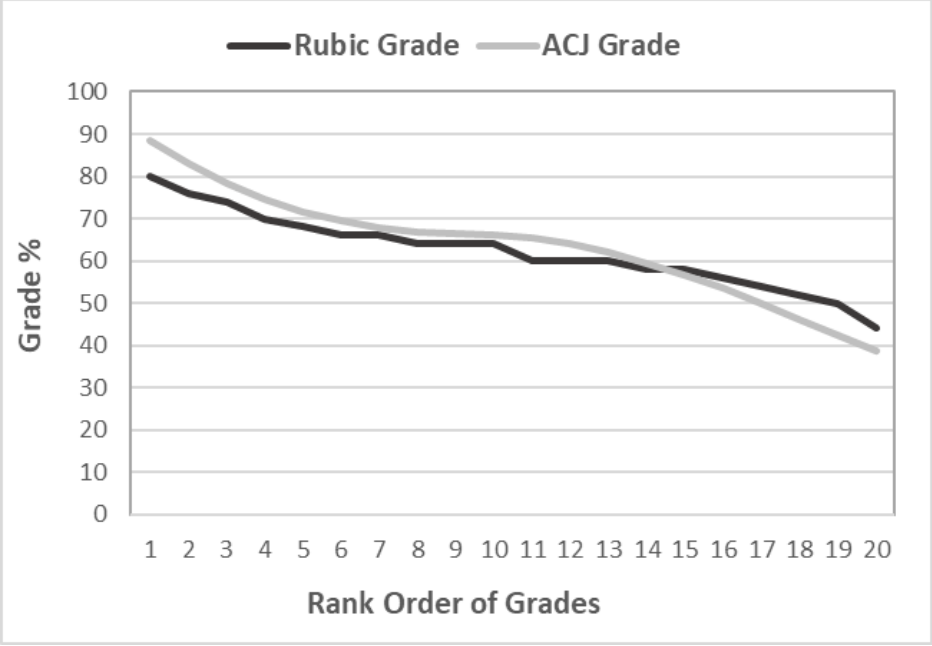

Figure 3. A comparison of the spread of grades using rubric-based grading vs using ACJ informed grading.

Figure 4 then compares the actual rubric-based grades to the grade a group would have received if the ACJ informed grades where used. Whilst it is interesting for this study to directly compare ACJ informed grades to rubric-based grades, it should be noted that the two methods for producing grades are different and therefore a difference in grades is to be expected. However, the fact that there is broad correlation between the grades produced by dissimilar grading methodologies is interesting and indicates a promising area of further work.

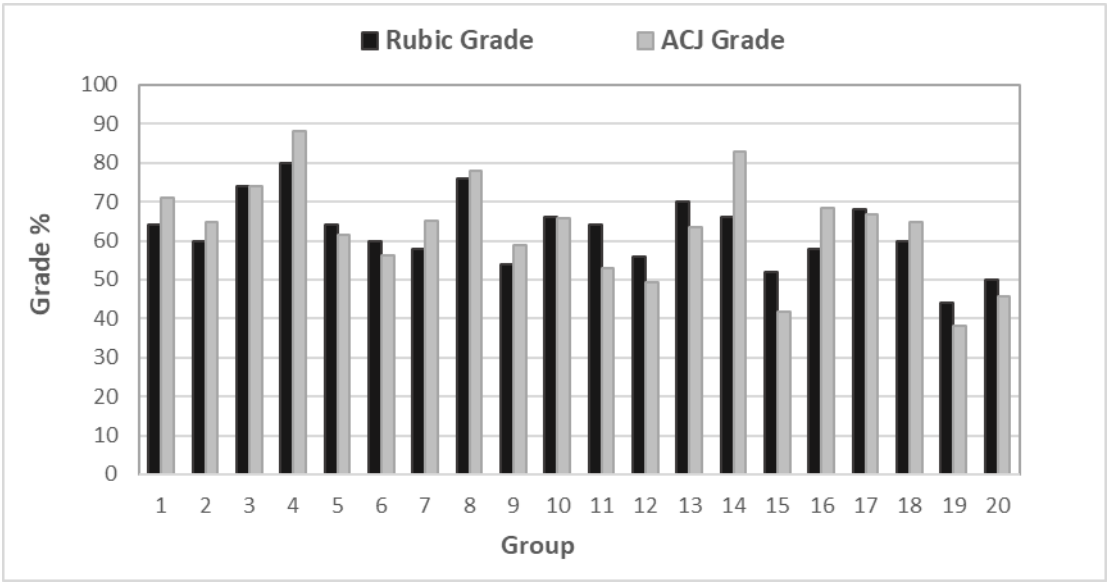

Figure 4. A comparison of actual rubric-based grade vs the hypothetical ACJ informed grade.

CONCLUSIONS

All respondents to the survey either 'Strongly agreed' (65.2%) or 'Agreed' (34.8%) that the RM Compare platform was easy to use and when asked if they would like to see RM Compare used in other activities, 25.9% 'Strongly agreed', 38.9% 'Agreed', 31.5% were 'Neutral', 1.9% 'Disagreed' and 1.9% 'Strongly disagreed'. These usability results, when taken with the results showing that ACJ can bring additional learner benefits, indicate that there would be value in continuing to develop and deploy ACJ activities. The results around using ACJ to grade work are also promising; as class sizes get bigger and staff time becomes more precious, anything that has the potential to save time would be welcomed. Importantly, ACJ is scalable; it can be used in a class of 50 or 500. This scalability is not only useful when thinking about grading for large cohorts but also when designing activities for large cohorts. There is evidence to show that students have benefited somewhat from the deep learning that comes with critical thought, all without major facilitation from instructors. Add to this the finding that students are receptive to the idea that ACJ could replicate a poster exhibition, all be it asynchronously and at a fraction of the cost, suggest a number of promising directions for further work. The authors will continue to use ACJ; the results from this initial study informing how they will develop strategies to make further use of ACJ and; forming the basis of a new study into the possibility of using ACJ to assist with grading.

Student feedback and the authors experience of deploying ACJ have highlighted a number of areas of further work that would improve future activities. The different levels of student engagement between the activities indicate that how the activities are deployed is an important factor in how effective they are. It had been assumed that students in second year of study would be mature enough to be left to complete the activity in their own time and that being told the activity would likely increase their grade would be incentive enough to do so. This assumption was wrong. The authors strongly suggest allocating class time to complete these activities as a strategy to improve engagement, beyond linking completion to final grade; from past experience this is the best way to get students to complete not-for-credit activities such as module evaluation surveys.

The strategy by which the authors introduced these activities to students should also be noted; students were informed that the use of ACJ and the RM Compare software was experimental and this perhaps may have discouraged some students from engaging. Some students completing the activity in case study 1 reported that some posters were taking too long to read and so weren't fully engaging with the judgements. The authors ran a number of ACJ pilot sessions with other staff and found that if the holistic statement (judging criteria) or the pieces of work being judged were too complex, judgments would default to being made based on the aesthetics of a piece of work. These findings indicates that thorough consideration should be given to both the work being judged and the criteria for judging, as well as the number of judgments students are asked to make. Some students in case study 1 reported that some of their classmates weren't engaging with the session properly because they had realised that they could quickly complete the activity by randomly selecting a piece of work without properly engaging in a judgement. It is possible to have the RM Compare software require students to leave rationale for their judgments which may overcome this issue.

FINANCIAL SUPPORT ACKNOWLEDGEMENTS

The authors would like to thank Jack Scott, an MEng student, contributing to the design of the assignments and working to gain initial feedback from peers on the new activities. The authors would also like to thank Dr Vincent Page who helped with data analysis.

FINANCIAL SUPPORT ACKNOWLEDGEMENTS

The authors received no financial support for this work and have no affiliation with RM Compare or their employees.

REFERENCES

Bartholomew, S.R., Mentzer, N., Jones, M. et al. (2022) Learning by evaluating (LbE) through adaptive comparative judgment, *Int J Technol Des Educ 32, 1191–1205*.

Bartholomew, S. R., Yoshikawa, E. (2018). A systematic review of research around Adaptive Comparative Judgment (ACJ) in K-16 education. *2018 CTETE Monograph Series*.

Biggs, J., & C. Tang. (2011). Teaching for Quality Learning at University, Fourth Edition, *The Society for Research in Higher Education. Berkshire: Open University Press*

Boud, D., Cohen, R., & Sampson, J (2001). Peer Learning in Higher Education: Learning from and with Each Other (1st ed.). *Routledge*

Canty, D. (2012). The impact of holistic assessment using adaptive comparative judgment of student learning, *PhD Thesis, University of Limerick, Ireland.*

Florian, L., & Pratt, L. (2015). Embedding equality and diversity in the curriculum: an education practitioner's guide*. York: Higher Education Academy.*

Hansson, J., Kolmskog, P., Odemark, Y. & Högfeldt, A. (2011). Peer Learning With Small Means: A Case Study Of Implementing Peer Learning In A Laboratory Exercise. *Proceedings of the 7th International CDIO Conference, DTU, Denmark*

Hermon, P. & McCartan, C. (2011). Peer Rating for Feedback in Group Projects, *Proceedings of the 7th International CDIO Conference, DTU, Denmark*

Jonsson, J., Täck, E., Hulthén, E. & Eriksson, A. (2022). PEER REVIEW IN A PRODUCT DEVELOPMENT COURSE – IMPLEMENTATION AND RECEPTION, *Proceedings of the 18th International CDIO Conference, Reykjavik University, Iceland*

Kimbell, R. (2018). Constructs of quality and the power of holism. Paper presented at the PATT35, Athlone Institute of Technology, Athlone, Ireland.

Kimbell, R. (2022) Examining the reliability of Adaptive Comparative Judgement (ACJ) as an assessment tool in educational settings. *Int J Technol Des Educ 32, 1515–1529*.

Pollitt, A. (2004). Let's stop marking exams. Available at: http://www.cambridgeassessment.org.uk/images/109719-let-s-stop-marking-exams.pdf. (Accessed 10th January 2023)

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19(3), 281–300.*

Poot, R., Gielen, K., Brincker, L., Filius, R. & Wiegant, F. (2020) Deep Learning through Peer Feedback in Higher Education; a Manual for Teachers, *Utrecht University,* Available at: https://cat-database.sites.uu.nl/wp-content/uploads/sites/778/2020/10/18012-B2-Deep-Learning-through-Peer-feedback-final-sept2020.pdf (Accessed 10th January 2023)

Seery, N., Canty, D. (2017). Assessment and Learning: The Proximal and Distal Effects of Comparative Judgment. In: *de Vries, M. (eds) Handbook of Technology Education. Springer International Handbooks of Education. Springer, Cham*

Tanner, D., Canty, D. & Power, J. (2019) Combined Strategies to Promote Active Learning and Retention, *Proceedings of the 15th International CDIO Conference, Aarhus University, Denmark*

Thurstone, L. L. (1927). A law of comparative judgment*. Psychological Review, 34, 273–286.*

Topping K. (1998) Peer assessment between students in colleges and universities, Review of Educational Research, 68 (3), pp. 249 – 276, DOI 10.3102/00346543068003249

Topping K.,J. (2009) Peer assessment, Theory into Practice, 48 (1), pp. 20 – 27, DOI 10.1080/00405840802577569

BIOGRAPHICAL INFORMATION

Tony Topping is a Learning Technologist in the School of Engineering and a Master of Arts candidate in Academic Practice, both at the University of Liverpool. He has a technical background, having worked as Teaching and Research Technician for 15 years, where he developed and delivered authentic learning experiences. His current work focuses on blending pedagogic and technical knowledge; and the use of technology to enhance the teaching activities of the school.

Matt Murphy is a Senior Lecturer in Engineering Design and Director of Education at School of Engineering, University of Liverpool, UK. He is a metallurgist by background with a PhD and 10-year technical research career in the additive manufacture (3D printing) of metals. For the last 14 years Matt has worked primarily on learning & teaching and has held several leadership positions in the School of Engineering, with special responsibility for curricular and pedagogic reform. Matt teaches a range of courses in materials science & manufacturing, but most enjoys leading the group design project modules. He establishes and supervises a broad range of student projects in fields such as renewable energy, urban farming, sustainable transport; veterinary healthcare; circular economy; and the local recycling of thermoplastics. In recent years Matt has focused on developing authentic learning and assessment experiences that seek to replicate industrial practice within taught programmes.

Samuel Saunders is an Educational Developer in the Centre for Innovation in Education (CIE) at the University of Liverpool. He has worked in higher education since 2015, and has a background in academic, pedagogic and curriculum development. Sam is particularly interested in assessment and feedback, programme and learning design, research-led/research-informed teaching, and decolonising pedagogy/assessment. Sam also holds a PhD in nineteenth-century literature and culture, and has published widely in this area.

Corresponding author

Tony Topping
School of Engineering
University of Liverpool
Liverpool, L69 3GH
United Kingdom
topping@liverpool.ac.uk