2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #36907

Using Adaptive Comparative Judgment to Holistically Assess Creativity of Design Solutions: A Comparison of First-Year Students and Educators' Judgments

Dr. Clodagh Reid, Technological University of the Shannon: Midlands Midwest

PhD in spatial ability and problem solving in engineering education from Technological University of the Shannon: Midlands Midwest. Graduated in 2017 from the University of Limerick with a B. Tech (Ed.). Member of Technology Education Research Group (TERG).

Dr. Sheryl A. Sorby, University of Cincinnati

Dr. Sheryl Sorby is currently a Professor of STEM Education at the University of Cincinnati and was recently a Fulbright Scholar at the Dublin Institute of Technology in Dublin, Ireland. She is a professor emerita of Mechanical Engineering-Engineering Mec

Mr. Gibin Raju, University of Cincinnati

Gibin Raju is a second-year doctoral student pursuing Engineering Education in the Department of Engineering and Computing Education at the College of Engineering and Applied Sciences. His research interests are focused on spatial skills, cognitive stress, cognitive load, STEM accessibility issues, workforce development, engineering pathways, STEM education, ID/ODD, and education practices.

Dr. Jeffrey Buckley, Royal Institute of Technology (KTH)

Dr Jeffrey Buckley received his PhD from KTH Royal Institute of Technology, Sweden, in the area of spatial ability and learning in technology education. He is a qualified post-primary teacher of Design and Communication Graphics and Construction Studies.

Dr. Niall Seery, Technological University of the Shannon

Dr. Niall Seery is also the Director of the Technology Education Research Group (TERG) and is a Guest Professor in Technology Education at the Royal Institute of Technology (KTH) in Sweden.

Using Adaptive Comparative Judgement to Holistically Assess Creativity of Design Solutions: A Comparison of First-year Students and Educators Judgements

Abstract:

This Complete Research paper investigates the holistic assessment of creativity in design solutions in engineering education. Design is a key element in contemporary engineering education, given the emphasis on its development through the ABET criteria. As such, design projects play a central role in many first-year engineering courses. Creativity is a vital component of design capability which can influence design performance; however, it is difficult to measure through traditional assessment rubrics and holistic assessment approaches may be more suitable to assess creativity of design solutions. One such holistic assessment approach is Adaptive Comparative Judgement (ACJ). In this system, student designs are presented to judges in pairs, and they are asked to select the item of work that they deem to have demonstrated the greatest level of a specific criterion or set of criteria. Each judge is asked to make multiple judgements where the work they are presented with is adaptively paired in order to create a ranked order of all items in the sample. The use of this assessment approach in technology education has demonstrated high levels of reliability among judges (~0.9) irrespective of whether the judges are students or faculty. This research aimed to investigate the use of ACJ to holistically assess the creativity of first-year engineering students design solutions. The research also sought to explore the differences, if any, that would exist between the rank order produced by first-year engineering students and the faculty who regularly teach first-year students.

Forty-six first-year engineering students and 23 faculty participated in this research. A separate ACJ session was carried out with each of these groups; however, both groups were asked to assess the same items of work. Participants were instructed to assess the creativity of 101 solutions to a design task, a "Ping Pong problem," where undergraduate engineering students had been asked to design a ping pong ball launcher to meet specific criteria. In both ACJ sessions each item of work was included in at least 11 pairwise comparisons, with the maximum number of comparisons for a single item being 29 in the faculty ACJ session and 50 in the student ACJ session. The data from the ACJ sessions were analyzed to determine the reliability of using ACJ to assess creativity of design solutions in first-year engineering education, and to explore whether the rankings produced from the first-year engineering students ACJ session differed significantly from those of the faculty. The results indicate a reasonably high level of reliability in both sessions as measured by the Scale Separation Reliability (SSR) coefficient, $SSR_{faculty} = 0.65 \pm 0.02$, $SSR_{students} = 0.71 \pm 0.02$. Further a strong correlation was observed between the ACJ ranks produced by the students and faculty both when considered in terms of the relative differences between items of work, $r = .533$, $p < .001$, and their absolute rank position, $\sigma = .553$, $p < .001$. These findings indicate that ACJ is a promising tool for holistically assessing design solutions in engineering education. Additionally, given the strong correlation between ranks of students and faculty, ACJ could be used to include students in their own assessment to reduce the faculty grading burden or to develop a shared construct of capability which could increase the alignment of teaching and learning.

Key Words: Engineering Education, Design, Assessment, Adaptive Comparative Judgement.

Background

In the recent past, an increasing emphasis has been placed on the development of undergraduate engineering students' design capabilities, with a focus on enabling students to develop innovative and creative design solutions [1], [2]. This change has been driven by design being mandated by ABET as a core skill that graduates should be equipped with upon graduation. To foster the development of these skills, design projects have been incorporated into the first year of many engineering programs in addition to capstone senior design projects [1], [3].

While there have been significant endeavors to incorporate design into engineering programs, an issue remains with the assessment of design tasks and projects. These forms of activities have traditionally been assessed through portfolios, rubrics, and criterion grading tools [4]–[7], which have some limitations when assessing open-ended design projects. Some of these limitations include excessive time investment in grading, timeliness of feedback, validity issues, and low levels of reliability [8], [9]. In terms of reliability and validity, it can be quite difficult to capture and quantify an element of design performance such as creativity through a grading rubric, where "points" are typically given for certain features of the design. Given the nature of creativity and design, in that they can manifest and be represented in a number of ways [10], it is necessary that assessment tools for engagement in a holistic task are able to appropriately measure this phenomenon holistically. Adaptive Comparative Judgement (ACJ) is one such tool which affords the opportunity to holistically assess design problem solving. As such, this study aims to investigate the use of ACJ for assessing creativity of first-year engineering students design solutions.

Adaptive Comparative Judgement

ACJ is a holistic assessment approach which involves intentionally and adaptively pairing two items of work which are assessed by a number of individual judges to produce a rank order of performance within a group [3], [4], [11], [12]. The intentional and adaptive pairing of items of work is driven by an algorithm [13] which pairs work based on maximizing the information gained resulting from decisions made by panellists to accelerate the achievement of a reliable rank order of performance [3], [4], [7], [11]. Software, such as RM Compare [14], uses this algorithm to automate the presentation of specific items of work to the judges. Figure 1 demonstrates how pairwise comparisons are presented to the user. The criteria that the work is assessed on can be specified through a 'holistic statement' presented to the judges. Specified criteria may include judges own perception of professional constructs, for example quality of design, innovation, etc., or judges may be provided with a specific definition of a construct. Previous research has demonstrated validity in both approaches [15].

Figure 1: ACJ pairwise comparison process

Judges of ACJ sessions have previously included students, professionals and instructors/ faculty [3], [5]–[7], [16]. The number of judgements that must be made in an ACJ session is typically specified as a function of the number of items of work to be assessed [13], with this number typically being divided by the total number of judges to determine the number of judgements asked of each individual judge. However, current developments of the ACJ process are leading to these configurations becoming more idiosyncratic in response to specific research questions [13], [17].

While making judgements through an ACJ system, there is also the opportunity to enable feedback to be provided, allowing the system to be used as a formative or summative assessment tool for design tasks. For instance, if an ACJ session with feedback enabled was to be conducted in the middle of a design project with students engaging in the design project acting as judges, the students could receive feedback on their progress from multiple perspectives in a timely manner to support them in refining their design before submission of their "final" product [7]. This would be to the benefit of both faculty and students as it would reduce the burden on faculty to provide feedback to all students in a large group while also increasing the feedback and timeliness of the feedback that students receive. Arguably for an assessment mechanism the most important factors considered are the reliability and validity of the assessment tool. The validity of ACJ refers to the validity of the rank-order that is produced and is directly tied to the cohort of judges which is assembled [18]. The reliability of ACJ sessions is described by the Scale Separation Reliability (SSR) coefficient which, in the context of comparative judgement, has strong indications that it reflects an interrater reliability index [19]. It is computed as

$$SSR = \frac{\sigma_\alpha^2 - MSE}{\sigma_\alpha^2} \qquad 1$$

where σ_α^2 is the standard deviation of the estimated parameter values squared, and MSE is the mean squared standard error, or the mean of the standard error values after they have been squared.

Previous studies that have used ACJ to assess different elements of design have reported SSR values of ~0.9 [3], [5], [7], [11], [16], [20]. Additionally, as this assessment approach relies on judgements being made by various individuals to set out the standard of work, it can also address issues of bias or varying views held by judges within a cohort via misfit statistics.

Considering the promising opportunities outlined through previous research of using ACJ to holistically assess design performance, this research aimed to investigate the use of ACJ to assess the creativity of first-year engineering students' design solutions. The research also sought to explore the differences, if any, that would exist between the rank order produced by first-year engineering students and faculty. This work will contribute towards addressing a gap in research on the use of ACJ in assessment in engineering education specifically focusing on creativity.

Method

Setting and Participants

The research described in this paper, is part of a larger study examining the link between spatial skills and engineering design. To address the research aims for this aspect of the study, work was conducted in two stages set out in Figure 2. In the initial stage, n = 127 undergraduate engineering students in the first and final year of their studies at a large public R1 university in the College of Engineering and Applied Science were invited to solve a design task. The participants were recruited through recruitment flyers that were shared throughout the college. The specific design task the students were asked to complete is described subsequently in this paper. The solutions for these students were then collated to produce the two separate ACJ sessions which included the same items of work (participant design solutions).

In the second stage, n = 25 engineering faculty from various universities attending the First-Year Engineering Experience (FYEE) conference completed the first ACJ session. The second ACJ session was completed by n = 46 first-year engineering students from the same large public R1 university as the initial research stage; however, it should be noted that due to the timing of the panel and the data collection activities, none of the panelists had personally completed the ping pong design task. The student ACJ panel was conducted as part of a classroom activity. For both ACJ panels all participants for each session took part at the same time in large open plan spaces.

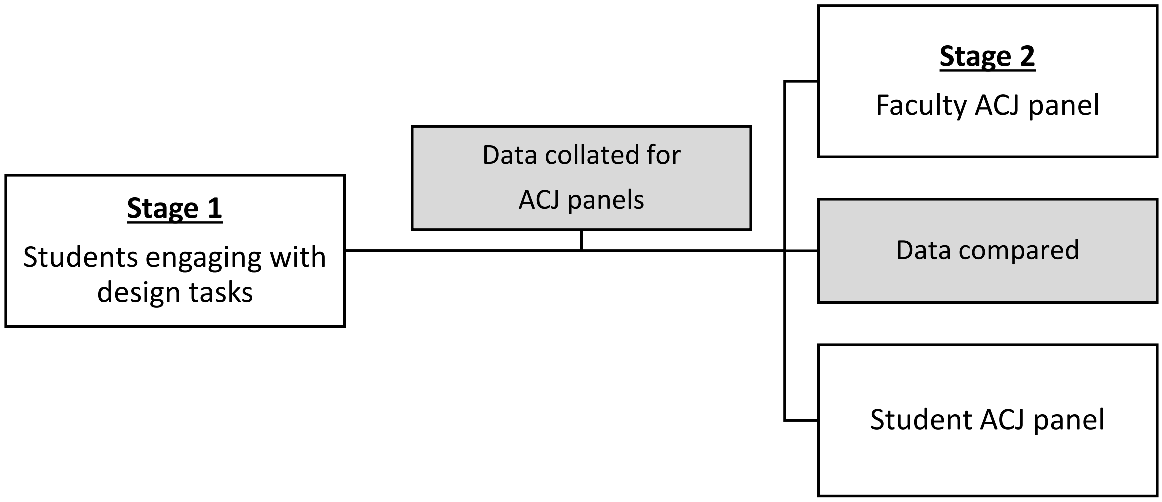

Figure 2: Overview of research approach

Data Collection

In the first stage of data collection the 127 undergraduate engineering students engaged with three design problems previously used in research investigating the design process [21]. This study focuses on the first problem that students engaged with where they were asked to design a ping-pong ball launcher. The instructions for participants included that the launcher they designed must be capable of providing an accurate landing with long flight time and that the ping pong ball could not be thrown at the target. The students could design any device that would meet these conditions. They were provided with specific dimensions of the ping-pong ball launcher and a distance from the launcher to the landing area. Participants were asked to detail all elements of their design including any relevant diagrams or calculations that they had completed. Additionally, they were asked to outline any assumptions that they had made and to attempt to keep their design simple yet effective. The participants were provided with as much time as they wished to solve the problems.

Following the first stage, 26 participants problem-solving solutions were removed from the study because of missing data relating to the specific problem. This resulted in 101 participants solutions being anonymized and loaded into two ACJ sessions, created through the RM Compare software, which had the same parameters. Judges were instructed, both by the facilitator (first-author) and a holistic statement above their judging window, to compare the solutions they were presented with and select the one the believed to be "better" based only on the criterion of "creativity". A definition of creativity was not provided to the participants to investigate the alignment of faculty and students own perceptions of the construct of creativity. An example of the representation of information to the ACJ judges is outlined in Figure 3.

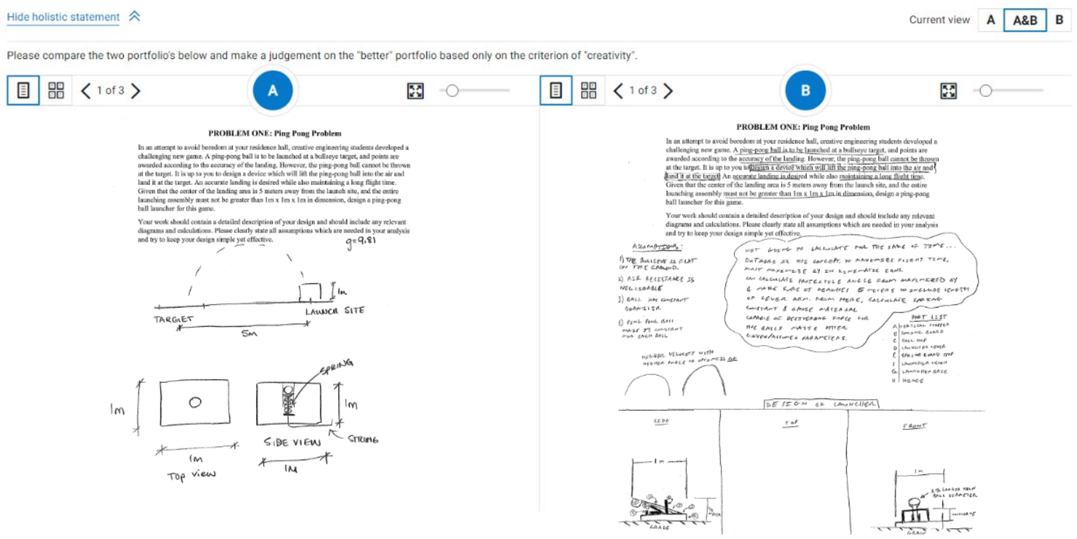

Figure 3: Presentation of information to judges through the ACJ tool.

The second stage involved n = 25 engineering faculty from various universities being invited to participate in an ACJ session where each individual was asked to complete between 40 and 41 ([101 pieces of work × 10]/25 judges) pairwise comparisons. The participants were instructed that there could only be one winner of each comparison - it was not possible to deem that both solutions were equally creative. Of the 25-faculty invited, 23 individuals participated in this session. As the rounds progressed, the evolution of the ranking and reliability could be observed in real-time.

In a second ACJ session 46 freshmen were invited to compare the same items of work that the engineering faculty had previously compared in stage two. The same instructions were provided to the participants in both ACJ sessions by the same facilitator. Each of the student judges were asked to make 21 to 22 ([101 pieces of work × 10]/46 judges) pairwise comparisons.

Results

The reliability of both sessions was moderately high ($SSR_{faculty} = 0.65 \pm 0.02$, $SSR_{students} = 0.71 \pm 0.02$); however, this value was notably lower than prior studies [3], [5], [7], [11], [16], [20]. Two factors are important to consider in this regard. First and more critical to this study is that these prior studies did not ask judges only to consider creativity, a construct for which no explicit definition was offered to the judges, and which has many possible components [22]. Instead, these prior studies were conducted in educational activities where judges were asked to choose the best piece of work or the piece of work most reflective of learning – an assessment question theoretically more likely to have a more consensual response. Second, neither session was "finished" as not all judges completed all of their assigned judgements. The faculty ACJ session was 83% complete with 17 judges making 41 judgements, and the remaining six judges made 5, 6, 14, 18, 31, and 37 judgements respectively. The student ACJ session was 98% complete with 39 of the judges making 22 judgements, two making 21 judgements each, and the remaining five judges made 2, 7, 11, 19, and 20 judgements respectively.

To exemplify the effect of this, it should be noted that an ACJ session is organised in terms of "rounds" where one round involves all items of work being involved in one comparison. As in this study both sessions involved 101 items of work, a round would consist of 51 pairwise comparisons. The SSR coefficient is then computed at the end of each round. As shown in Figure 4 the reliability of the ACJ process increases the more information (judgements) are included in the model which produces the rank. It is unlikely that these values would have been much higher if both sessions were 100% completed, but interpreting these results with respect to any future related work it should be noted that both reliability estimates could have been increased.

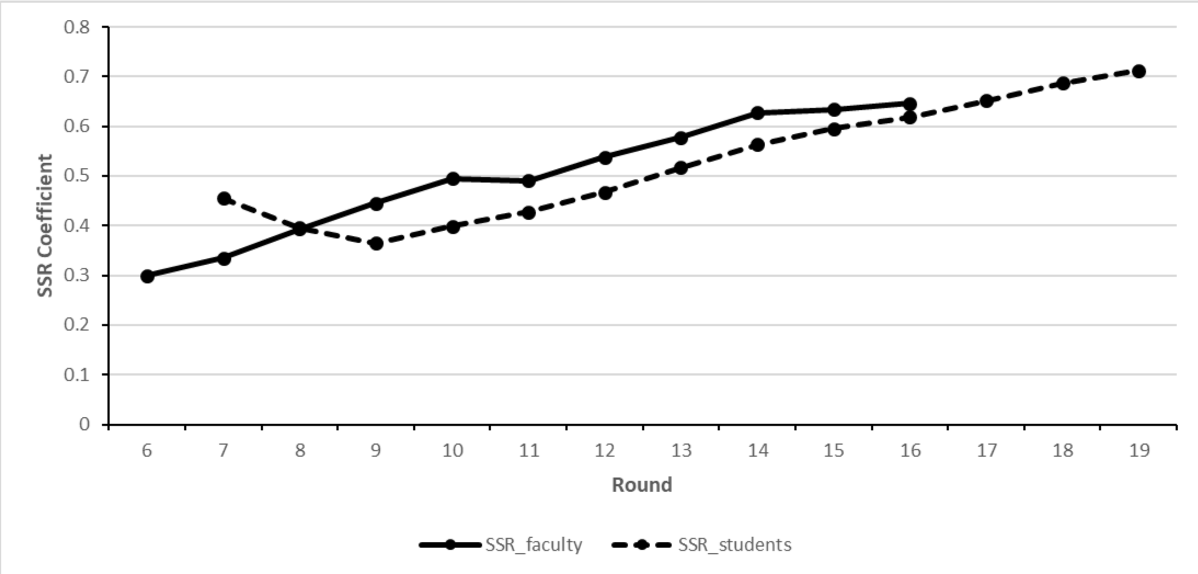

Figure 4: SSR coefficient computed at the end of each round of both ACJ sessions. *

***Note, SSR is only computed in the RM Compare system after the 6th or 7th round as during the initial rounds pairing of items not driven by the adaptive algorithm.**

Next, it was of interest to determine the degree of association between the ranks produced by the engineering faculty and students. It should be noted that the output of an ACJ session is a rank with two modes of interpretation. First, there is the absolute rank, i.e. relative 1st place, relative 2nd place, … relative last place. As an absolute rank no information pertaining to how much better or worse the judge cohort perceived one item of work to be in comparison to another is offered. Second, the rank contains portfolio (individual pieces of work) parameter values, which are also referred to as ability scores [23]. These describe relative differences between the items of work (with standard error) and give an indication of how much better or worse the judges pieces items of work to be in comparison to each other. The parameter values are transformed to have a mean score of 0, and therefore portfolios close to 0 can be interpreted as reflective of the perceived theoretically average pieces of work. Critically, the ranks do not describe the quality of the work. It is not clear from the ranks whether the judges considered the work to be of high or low quality, the only information contained in the rank is the perceived ordering of the pieces of work relative to each other.

The faculty produced rank is presented in Figure 5 and the student produced rank is presented in Figure 6. To determine the degree of association between the ranks as an indicator of how much the faculty and students agreed on the demonstrated levels of creativity within the pieces of work, both a Pearson's correlation and Spearman's correlation were conducted on the parameter values of each rank. The Pearson's correlation revealed a strong and statistically significant correlation, $r = .533$, $p < .001$. The Spearman's correlation conducted on the parameter values, being a rank transformed Pearson's correlation, was identical, if a Pearson's correlation had been conducted on the absolute rank positions of the pieces of work. It two revealed a strong and statistically significant correlation, $\sigma = .553$, $p < .001$.

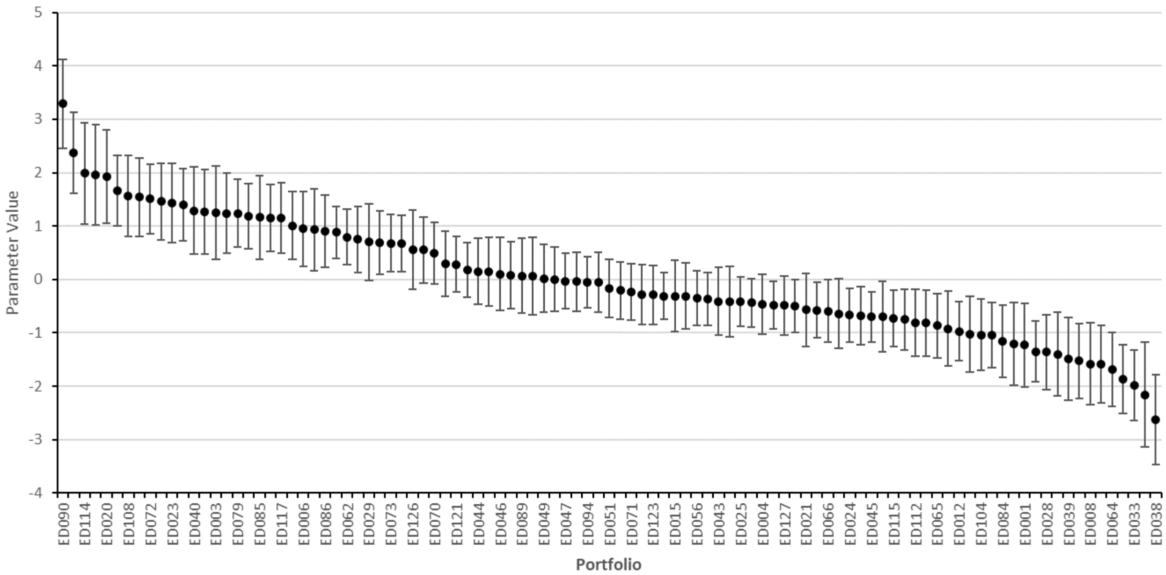

Figure 5: ACJ rank of the items of work produced by the engineering faculty. Note that every second portfolio ID is omitted due automated sizing of the axis.

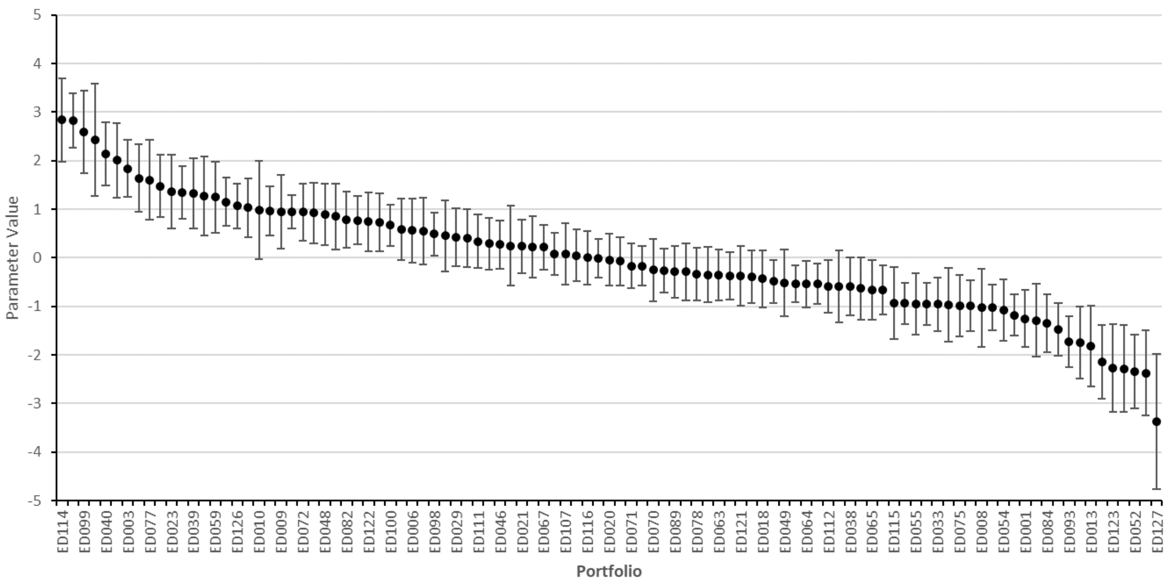

Figure 6: ACJ rank of the items of work produced by the engineering students. Note that every second portfolio ID is omitted due automated sizing of the axis.

While both correlations are interpretable as "strong" [24]–[26], they indicate approximately 28% - 30% shared variance, meaning there is still considerable difference. As items of work at the extremes of the ranks are more clearly separable than those in the middle in terms of within-rank perceive quality variation, the ten highest and lowest ranked items of work are presented in Table 1. What is clear is that within the lists of top and bottom ranked items of work, there are very few that are contained within both lists. This indicates that there is a statistically significant association which is interpreted here as being a strong degree of consensus in terms of how creativity is perceived. There is still considerable practical difference which generates the question of what qualities or factors did the judges of both cohorts consider when making decisions pertaining to creativity?

Table 1: Ten highest and lowest portfolios from each rank.

Ten highest ranked portfolios			**Ten lowest ranked portfolios**		
Faculty Rank	***Portfolio***	***Students Rank***	***Faculty Rank***	***Portfolio***	***Students Rank***
1	**ED090**	2	92	ED116	49
2	ED078	59	93	ED039	13
3	**ED114**	1	94	ED067	45
4	ED010	19	95	ED008	86
5	ED020	51	96	**ED076**	100
6	ED063	60	97	ED064	69
7	ED108	12	98	ED075	83
8	ED007	24	99	ED033	81
9	ED072	23	100	**ED052**	99
10	**ED105**	10	101	ED038	71

Note. Items of work in bold font are those which appear in the lists for both faculty and students.

Discussion and Conclusions

The aims of this research were to investigate the use of ACJ for holistically assessing creativity of engineering students design solution and to examine the differences that may exist between the rank order produced by engineering students and faculty. To address the first element of the research aim, the reliability of the positioning of items of work in the rank was analyzed. It was found that the reliability of both ACJ sessions was moderately high. This is promising in terms of measuring constructs like creativity and design in engineering education, which manifest and are defined in various ways [10], [22], in a holistic manner. Although the reliability was found to be moderately high, it is lower than previous studies [3], [5], [7], [11], [16], [20]; however, these studies were assessing work based on less subjective criteria. The reliability of the rank in the current study could have been increased if all ACJ participants had completed their sessions. This was due to a limitation in completing the sessions where some participants attempted to complete their judgements on mobile phones which resulted in issues displaying the items of work. This can be mitigated in future studies by only allowing for completion of the sessions on a laptop or tablet device. The reliability of the sessions may also have been increased by providing a definition of creativity, however, future research is necessary to determine whether this would be the case.

To investigate the second element of the research aim, examining differences in student and faculty produced ACJ rank, the reliability and correlation between the two ranks were compared. The reliability determined in the student rank was marginally higher than that of the engineering faculty; however, it is important to note that the student rank reached a higher level of completion. With respect to correlation analysis, strong and statistically significant correlations were found between each of the ranks. These findings are significant when considered in the context of assessment, and more specifically, issues currently faced with the assessment of design solutions and projects in engineering education. Issues with assessing open-ended design projects have included low levels of reliability, excessive time investment and timeliness of feedback [8], [9]. Through this study, ACJ has demonstrated a moderately high level of reliability when used by both students and faculty to assess an element of design and strong correlation between both ranks. This strong correlation suggests that in practice ACJ could be used by students to reliably assess the work of their peers on design tasks which would significantly reduce grading and time demands on faculty. Through the ACJ system that was employed in this research, RM Compare, there is the capacity to include a feedback feature. Using this feature in student graded ACJ sessions would enable students to receive timely feedback from multiple perspectives to support them in refining their design solutions, as has previously been the case in Technology education contexts [7]. Involving students in the assessment process through ACJ not only affords them the opportunity to receive a greater level of feedback but also provides them with exposure to the work of their peers which may allow them to recognize whether their work is meeting a good standard. From a faculty perspective, as this research and previous research has found strong correlations between student and faculty grading through ACJ, enabling students to act as graders in the place of/in tandem to faculty would significantly reduce the time spent grading and providing feedback on design projects to large class groups.

Acknowledgement

This work was made possible by a grant from the National Science Foundation (NSF #2020785). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] A. A. Anwar and D. J. Richards, ‘Comparison of EC and ABET Accreditation Criteria’, *J. Prof. Issues Eng. Educ. Pract.*, vol. 144, no. 3, p. 06018001, Jul. 2018, doi: 10.1061/(ASCE)EI.1943-5541.0000364.

[2] E. Starkey, C. A. Toh, and S. R. Miller, ‘Abandoning creativity: The evolution of creative ideas in engineering design course projects’, *Des. Stud.*, vol. 47, pp. 47–72, Nov. 2016, doi: 10.1016/j.destud.2016.08.003.

[3] G. J. Strimel, S. R. Bartholomew, S. Purzer, L. Zhang, and E. Y. Ruesch, ‘Informing engineering design through adaptive comparative judgment’, *Eur. J. Eng. Educ.*, vol. 46, no. 2, pp. 227–246, Mar. 2021, doi: 10.1080/03043797.2020.1718614.

[4] A. Pollitt, ‘The method of Adaptive Comparative Judgement’, *Assess. Educ. Princ. Policy Pract.*, vol. 19, no. 3, pp. 281–300, Aug. 2012, doi: 10.1080/0969594X.2012.665354.

[5] D. N. Seery, J. Buckley, A. Doyle, and D. D. Canty, ‘The Validity and Reliability of Adaptive Comparative Judgements in the Assessment of Graphical Capability’, p. 7, 2016.
[6] S. R. Bartholomew, G. J. Strimel, and A. Jackson, ‘A Comparison of Traditional and Adaptive Comparative Judgment Assessment Techniques for Freshmen Engineering Design Projects’, p. 15, 2018.
[7] S. R. Bartholomew, G. J. Strimel, and E. Yoshikawa, ‘Using adaptive comparative judgment for student formative feedback and learning during a middle school design project’, *Int. J. Technol. Des. Educ.*, vol. 29, no. 2, pp. 363–385, Mar. 2019, doi: 10.1007/s10798-018-9442-7.
[8] D. R. Sadler, ‘Formative assessment and the design of instructional systems’, *Instr. Sci.*, vol. 18, no. 2, pp. 119–144, 1989.
[9] J. Buckley, D. Canty, and N. Seery, ‘An exploration into the criteria used in assessing design activities with adaptive comparative judgment in technology education’, *Ir. Educ. Stud.*, Sep. 2020, doi: 10.1080/03323315.2020.1814838.
[10] R. Kimbell, ‘Wrong ... but right enough’, *Des. Technol. Educ. Int. J.*, vol. 16, no. 2, pp. 6–7, 2011.
[11] S. R. Bartholomew and M. D. Jones, ‘A systematized review of research with adaptive comparative judgment (ACJ) in higher education’, *Int. J. Technol. Des. Educ.*, vol. 32, no. 2, pp. 1159–1190, Apr. 2022, doi: 10.1007/s10798-020-09642-6.
[12] E. Hartell and J. Buckley, ‘Comparative judgement: An overview’, in *Handbook for Online Learning Contexts: Digital, Mobile and Open*, A. Marcus Quinn and T. Hourigan, Eds. Switzerland: Springer International Publishing, 2021, pp. 289–307. [Online]. Available: https://doi.org/10.1007/978-3-030-67349-9_20
[13] S. Verhavert, A. Furlong, and R. Bouwer, ‘The accuracy and efficiency of a reference-based adaptive selection algorithm for comparative judgment’, *Front. Educ.*, vol. 6, 2022, Accessed: Mar. 28, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2021.785919
[14] RM Compare, ‘RM Compare’, *RM Compare*, 2023. https://compare.rm.com/ (accessed Feb. 10, 2023).
[15] N. Seery, D. Canty, and P. Phelan, ‘The validity and value of peer assessment using adaptive comparative judgement in design driven practical education’, *Int. J. Technol. Des. Educ.*, vol. 22, no. 2, pp. 205–226, May 2012, doi: 10.1007/s10798-011-9194-0.
[16] N. Seery, J. Buckley, T. Delahunty, and D. Canty, ‘Integrating learners into the assessment process using adaptive comparative judgement with an ipsative approach to identifying competence based gains relative to student ability levels’, *Int. J. Technol. Des. Educ.*, vol. 29, no. 4, pp. 701–715, Sep. 2019, doi: 10.1007/s10798-018-9468-x.
[17] N. Seery, R. Kimbell, and J. Buckley, ‘Using Teachers’ Judgments of Quality to Establish Performance Standards in Technology Education Across Schools, Communities, and Nations’, *Front. Educ.*, vol. 7, 2022, Accessed: Mar. 20, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2022.806894
[18] J. Buckley, N. Seery, and R. Kimbell, ‘A review of the valid methodological use of adaptive comparative judgment in technology education research’, *Front. Educ.*, vol. 7, no. 787926, pp. 1–6, 2022, doi: 10.3389/feduc.2022.787926.
[19] S. Verhavert, S. De Maeyer, V. Donche, and L. Coertjens, ‘Scale Separation Reliability: What Does It Mean in the Context of Comparative Judgment?’, *Appl. Psychol. Meas.*, vol. 42, no. 6, pp. 428–445, Sep. 2018, doi: 10.1177/0146621617748321.
[20] D. Canty, N. Seery, E. Hartell, and A. Doyle, ‘Integrating Peer Assessment in Technology Education through Adaptive Comparative Judgment’, presented at the Pupils

Attitude Towards Technology (PATT 34), 2017. [Online]. Available: https://www.researchgate.net/publication/320299812
[21] C. J. Atman, ‘Design timelines: Concrete and sticky representations of design process expertise’, *Des. Stud.*, vol. 65, pp. 125–151, Nov. 2019, doi: 10.1016/j.destud.2019.10.004.
[22] J. C. Kaufman and V. P. Glăveanu, ‘A review of creativity theories: What questions are we trying to answer?’, in *The Cambridge handbook of creativity, 2nd ed*, New York, NY, US: Cambridge University Press, 2019, pp. 27–43. doi: 10.1017/9781316979839.004.
[23] J. Buckley and D. Canty, ‘Assessing performance: Addressing the technical challenge of comparing novel portfolios to the “ACJ-Steady State”’, in *PATT39: PATT on the Edge - Technology, Innovation and Education*, St. John’s, Newfoundland and Labrador, Canada, 2022, pp. 523–537.
[24] T. Schäfer and M. A. Schwarz, ‘The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases’, *Front. Psychol.*, vol. 10, p. 813, 2019, doi: 10.3389/fpsyg.2019.00813.
[25] J. Cohen, *Statistical power analysis for the behavioral sciences*. New Jersey: Lawrence Erlbaum Associates, 1988.
[26] D. C. Funder and D. J. Ozer, ‘Evaluating Effect Size in Psychological Research: Sense and Nonsense’, *Adv. Methods Pract. Psychol. Sci.*, vol. 2, no. 2, pp. 156–168, Jun. 2019, doi: 10.1177/2515245919847202.