DOES COMPARATIVE JUDGEMENT REDUCE STUDENTS' PERCEIVED COGNITIVE LOAD WHEN EVALUATING MATHEMATICS SOLUTIONS?

Jennifer Palisse[a], Deborah King[a], and Mark MacLean[b]

Presenting Author: Jennifer Palisse (jmpalisse@student.unimelb.edu.au)
[a]School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia
[b]Department of Mathematics, University of British Columbia, Vancouver British Columbia BC V6T 1Z4, Canada

KEYWORDS: comparative judgement, comparison, cognitive load, peer assessment

ABSTRACT

Comparative judgement can be used as a method of facilitating peer evaluation in educational settings. Early empirical findings indicate that evaluating peer work comparatively leads to improved learning outcomes when compared to evaluating peer work sequentially. This study explored the role of perceived cognitive load as a potential contributing factor in explaining why comparative judgement enhances learning outcomes. Undergraduate mathematics students were randomly assigned to one of three conditions: a sequential condition where students evaluated solutions one-at-at-time (N = 164), a pairs condition where students evaluated solutions in pairs that shared structural features (N = 186), or a second pairs condition where solution-pairs did not share commonalities (N = 182). To measure perceived cognitive load, students completed a questionnaire and provided written reflections on their experience of evaluating worked solutions. From these, we identified instances where students expressed difficulty during evaluation. Results indicate that presenting solutions to questions as pairs reduced the perceived cognitive load compared to presenting solutions sequentially, and the effect was more pronounced when the paired solutions shared common features. These differences were small, indicating that cognitive load may only play a small role in the effectiveness of comparative judgement for learning.

Proceedings of the Australian Conference on Science and Mathematics Education, The University of Tasmania, 30 August – 1 September 2023, pages 77-82, ISSN 2653-0481.

INTRODUCTION

In recent years, educators have adopted comparative judgement as a new tool to facilitate peer assessment among students (for a review of current research, see Bartholomew and Jones, 2022). Comparative judgement involves showing students two samples of work, side-by-side, and asking them to select which one they believe is of higher quality. Students complete several of these comparisons, producing a ranking of work from best to worst based on their selections.

Empirical evidence indicates that comparative judgement is an effective pedagogical tool in certain disciplines. Students studying a Design Practice subject who engaged in comparative judgement activities tended to achieve better learning outcomes than those who evaluated solutions sequentially, as judged by the quality of students' design projects (Bartholomew, Strimel, et al., 2019), portfolio work (Canty et al., 2017), and quality of Point of View statements (Bartholomew et al., 2022). In the context of English, Bouwer et al. (2018) report that students who engaged in comparative judgement produced essays of higher quality than students who assessed example essays one-at-a-time by applying marking criteria. However, this positive effect was not replicated in subsequent studies of similar design (van Daal et al., 2023).

One reason comparative judgement might improve learning outcomes is that it helps learners adopt the vocabulary of experts. It has been argued that the process of comparing peer work facilitates learners' abilities to express and communicate characteristics that make up high-quality work (Kimbell, 2020). This means learners may be more likely to adopt the vocabulary of experts and this, rather than a better understanding of the assessment criteria, is what is thought to move students from novice to expert (Bartholomew, Mentzer, et al., 2019).

While previous studies have shown the positive effects of comparative judgement on learning outcomes, the mechanisms underlying these effects are largely unexplored. This study explores one potential mechanism: cognitive load. According to Cognitive Load Theory (Sweller et al., 2011), instructional design will only be beneficial to learning when the required cognitive load stays within the limits of working memory. We were interested in exploring whether the presentation of worked solutions as pairs, as an instructional design approach, might reduce cognitive load. Specifically, we examined the potential influence of comparative judgement on perceived cognitive load and whether this relationship may contribute to the effectiveness of comparative judgement for learning reported in the literature. By investigating the role of cognitive load, this study aims to shed light on the underlying mechanisms that may explain the potential benefits of using comparative judgement in educational settings.

LEARNING THROUGH COMPARISONS AND COGNITIVE LOAD THEORY

Comparing and contrasting alternative strategies and solutions has been a key pedagogical approach of reform mathematics classrooms in primary and secondary schools (Chazan & Ball, 1999). Learning through comparison has been found to improve mathematics learning more so than learning from problems one-at-a-time (Alfieri et al., 2013). The mechanism of comparing aids in the identification of underlying structures which can then be stored as a mental representation (Gray & Holyoak, 2021). Once such structures are stored in memory, they can be applied to similar problems.

While comparing examples appears beneficial for learning, the relationship between comparing and its associated cognitive load is unclear. According to Cognitive Load Theory (Sweller et al., 2011), working memory has a limited capacity to retain information at any given time. The amount of working memory required for different tasks can vary greatly which can impact our ability to process information, thereby affecting the ease with which one can learn new skills from a task. Learning becomes easier when extraneous cognitive load is decreased, that is, the load imposed by the way information is presented to the learner. The question at hand is whether presenting worked solutions as pairs imposes a higher or lower cognitive load than when compared to presenting worked solutions sequentially.

When examples are presented sequentially, learners are required to compare the current solution with one stored in working memory to make a comparison, which leads to an increase in cognitive load (Begolli & Richland, 2016). Based on this observation, it is reasonable to posit that presenting examples in pairs may help reduce cognitive load. Conversely, the act of comparing may *increase* cognitive load due to the multiple processes required when comparing examples. Such processes include understanding the two individual solutions both in isolation and in relation to each other, holding solution steps from both solutions in working memory, identifying suitable mappings between them, and filtering relevant information while ignoring irrelevant information (Richland et al., 2016). The cognitive effort required is further exacerbated for learners with low prior knowledge as their working memory resources must be split between interpreting unfamiliar solutions and making comparisons (Durkin et al., 2017).

In the context of comparative judgement, the impact that learning by comparing examples might have on cognitive load is less clear. The presentation of solutions side-by-side should arguably reduce cognitive load as it is the delivery of pairs that eliminates the need for students to compare a solution to one stored in memory. However, the act of comparing requires students to both understand and compare multiple solution approaches, which can increase cognitive load. As such, an argument could be made that comparative judgement both increases *and* decreases cognitive load.

METHODS

The results presented in this paper are part of a larger study. As such, the description of the research design includes those aspects which are relevant to the analysis of these results.

PARTICIPANTS

Participants were drawn from an undergraduate first-year calculus subject at an Australian university. All students in the cohort were invited to participate with the opportunity to opt-out, with a total of 532 providing consent to participate.

DESIGN

Ethics approval for this study was sought and received from the Human Research Ethics Committee at The University of Melbourne (reference number: 2022-23405-26158-4). All participants provided informed consent.

At intervention, students evaluated a set of six solutions. Students were randomly assigned to one of three conditions: sequential (N = 164), aligned pairs (N = 186), and misaligned pairs (N = 182). For the sequential condition, students were shown the six solutions one-at-a-time and asked to give a score between one and five, where five indicated a high-quality piece of work. For the aligned pairs condition, students were shown the solutions as pairs and asked to select which of the two examples they thought was 'better'. Solution-pairs had similar solution approaches (e.g., both solutions solved the problem algebraically or both solutions solved the problem graphically). The misaligned pairs condition presented students with the solutions as pairs that did not align closely. All three groups were shown the same set of six solutions and all solutions were seen once by students.

INSTRUMENTS

Questionnaire

A cognitive load questionnaire was used, adapted from Leppink et al. (2014), which consisted of nine items. Students responded to each item on a five-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*). Example items include "Reading other students' work was very complex and challenging" and "I needed to use a high amount of mental effort to understand these students' work". Students' responses to the nine items were aggregated to give a perceived cognitive load score out of 45.

Written reflections

On each of two consecutive weekly assignments, students were asked to find the solution set for an algebraic rational inequality. Students then compared their two solutions and were asked to comment on any changes they had made or aspects they had decided to keep the same. Students were also asked to reflect on any learnings gained by evaluating other students' work.

PROCEDURE

As part of a weekly mathematics assignment, students were asked to solve the following rational inequality: Find the set of real numbers, x, such that $\frac{5x-2}{x+5} > 6$. For the following week's assignment, students reviewed six worked solutions to this same problem either sequentially, or in pairs, depending on their assigned experimental condition. The six solutions included correct and incorrect answers, high- and low-quality solutions, a variety of solution approaches, and solutions that were messy or neat. Students were not provided with a marking scheme and were instructed that there was no right or wrong way to evaluate the worked solutions. Students reviewing solutions sequentially assigned each solution a score out of 5 whereas those reviewing solutions in pairs decided which of each pair was better. As part of the same assignment, students were asked to solve a similar rational inequality: Find the set of real numbers x, such that $\frac{x+3}{x-4} \leq 2$. Students were asked to submit their written reflection as part of this assignment.

Marks were allocated for both rational inequality problems, the evaluations, and written reflection. The rational inequality problems were each scored out of three marks, the evaluations and written reflections contributed an additional four marks. Each weekly assignment counted as 2% towards the final grade.

RESULTS

QUESTIONNAIRE RESULTS

The perceived cognitive load questionnaire had a high level of internal consistency, with a Cronbach's alpha of 0.83.

Normality of skewness and kurtosis values were examined for the cognitive load scale and were found to fall within an acceptable range. Next, an analysis of variance (ANOVA) was conducted to examine the effect of the three experimental conditions on perceived cognitive load. The Levene's test of determining homogeneity of variance was not violated ($p > 0.05$). Results indicated significant differences between groups, with a small effect size, $F(2, 531) = 3.08$, $p = 0.047$, $\eta^2 = 0.01$.

Table 1
Student perceived cognitive load by experimental condition

	N	Mean	SD
Sequential	164	26.9	6.1
Misaligned pairs	182	26.1	6.4
Aligned pairs	186	25.2	6.0

Means and standard deviations of results can be found in Table 1. The mean ratings for all three groups were between 25 and 27 on a 45-point scale, suggesting that students did not perceive the evaluative activity as highly demanding. However, Tukey post-hoc tests indicated significant differences between groups. Students who viewed solutions as pairs with high alignment reported lower cognitive load compared to those who viewed solutions sequentially ($p = 0.036$). There were no significant differences in perceived cognitive load between the misaligned pairs group and either the sequential ($p = 0.467$) or aligned pairs group ($p = 0.380$). While the mean ratings might suggest that students generally did not find that the activity required a high cognitive load, the observed differences between groups are still meaningful suggesting that presenting solutions as aligned pairs has the potential to further reduce cognitive load.

STUDENT REFLECTIONS

Thematic analysis techniques (Braun & Clarke, 2006) were used to analyse students' written reflections. The analysis involved a thorough reading of the reflections, from which we identified codes representing common patterns and ideas. These codes were then systematically compared and grouped into categories and overarching themes. This was an iterative process which inductively generated categories from the data itself. The analysis was exploratory in the sense that we did not search for anticipated themes.

One emergent theme centred around instances where students expressed difficulty with some aspects of the evaluation process, as indicated by comments such as "It can be confusing to read another person's working." A total of 31 students (5.8%) implied finding some aspect of evaluating other students' work challenging. We categorised these comments further into two groups: quality of explanation and neatness.

With regards to quality of explanations, 22 of the 31 students noted that solutions that lacked sufficient or high-quality explanations were difficult to follow:

> *The peer review activity gave me new knowledge about how important it is to have clear working as non-clear working makes it very difficult to look over.*

Almost all comments (12 of 13) made by students in the sequential group were related to the lack of explanation suggesting this was the most notable difficulty noticed by students in this group.

Five of the 31 students noted that including too much working or unnecessary steps were also hard to evaluate:

> *In the peer review activity, I learnt that sometimes more writing/more lines are not always the way to approach writing answers. Some students' work had many unnecessary steps which made it harder to mark.*

With regards to neatness, five of the 31 students noted that untidy or messy handwriting made solutions difficult to read, and therefore, difficult to understand:

> *After the peer review activity, I realised how much harder it can be to understand a person's solution when they are messy and poorly laid out, even if the answer is correct.*

We used a chi-squared test to examine whether a relationship existed between the experimental condition and the likelihood of students reporting some type of challenge with the evaluation process. Results indicate no significant differences across all three groups, $\chi^2(2, N = 525) = 4.85$, $p = 0.088$. However, we found significant differences when we compared only the sequential group with the aligned pairs group, $\chi^2(1, N = 343) = 4.90$, $p = 0.027$. Results suggest that students in the sequential group were more likely to comment finding some aspects of evaluating other students' work challenging than students in the aligned pairs group.

While the analysis here focuses on identifying instances where students expressed difficulty, it is important to note that they were not asked explicitly if they found the evaluative task challenging. Therefore, the 31 students who discussed finding part of the task challenging are unlikely to represent the full extent of students' experiences, and further exploration is warranted to understand the broader spectrum of student perceptions. Additionally, the instructions students were provided with directed them to reflect on their learnings, which might have guided their focus away from mentioning difficulties. Consequently, students may have underreported their challenges or prioritised discussing their learning experiences instead. While the overall number of students mentioning difficulties is small, results indicate a difference between the aligned pairs group and sequential group, which is consistent with findings in the previous section.

CONCLUSION

The results of this study indicate that presenting students with worked solutions one-at-a-time resulted in higher perceived cognitive load, whereas pairing solutions with structural similarities resulted in a lower perceived cognitive load. Students' written comments further support this finding: students who evaluated worked solutions one-at-a-time were more likely to comment on finding some parts of the evaluation process difficult compared to those who evaluated solutions as pairs with structural similarities.

While students in all three groups perceived the evaluative activity as moderately demanding in terms of cognitive load, the differences between groups remain meaningful. By presenting solutions as pairs, the extraneous cognitive load appears to have been reduced which may offer a partial explanation for the positive learning outcomes observed in previous studies (Bartholomew, Mentzer, et al., 2019; Kimbell, 2020). By presenting worked solutions in pairs, students no longer need to generate a solution pathway from a solution stored in memory, thereby reducing the demands on working memory (Sweller et al., 2011). This reduction in extraneous cognitive load may allow students who compare solutions to engage in more efficient cognitive processing during the evaluation process by allowing for better allocation of working memory resources. As such, the reduction in perceived cognitive load observed in this study may indicate that cognitive load may contribute to the effectiveness of comparative judgement for learning. However, it is important to acknowledge that the effect size observed was small, suggesting that while cognitive load may play a role, it is unlikely to be the sole or main contributing factor.

Lastly, we note that students reported lower cognitive load when paired solutions shared similarities. This aligns with the principles of variation theory, which suggests that when providing multiple examples, learning can be hindered when too many aspects are varied at once, and key aspects should be varied one-by-one at first, then varied simultaneously following mastery (Marton, 2015). The higher cognitive load reported by students in the misaligned pairs group may be a result of the greater variation between paired solutions. This provides some insight into the use of comparative judgement for learning in the context of peer assessment. Typically, adaptive algorithms are used in comparative judgement that select pairs based on their current ranking, rather than their level of variation. This means students see pairs that are of similar standard in ranking as opposed to pairs that have been selected due to high or low variation. Unless pairs can be carefully selected to minimise initial variation, comparative judgement may be more effective once students have established some familiarity with the content and can better cope with higher variation.

REFERENCES

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Bartholomew, S. R., & Jones, M. D. (2022). A systematized review of research with adaptive comparative judgment (ACJ) in higher education. *International Journal of Technology and Design Education*, *32*(2), 1159–1190. https://doi.org/10.1007/s10798-020-09642-6

Bartholomew, S. R., Mentzer, N., Jones, M., Sherman, D., & Baniya, S. (2022). Learning by Evaluating (LbE) through adaptive comparative judgment. *International Journal of Technology and Design Education*, *32*(2), 1191–1205. https://doi.org/10.1007/s10798-020-09639-1

Bartholomew, S. R., Mentzer, N., & Jones, M. D. (2019). Learning by Evaluating. *106th Mississippi Valley Technology Teacher Education Conference*.

Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2019). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education*, *29*, 363–385. https://doi.org/10.1007/s10798-018-9442-7

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*(2), 194–213. https://doi.org/10.1037/edu0000056

Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying criteria to examples or learning by comparison: Effects on students' evaluative judgment and performance in writing. *Frontiers in Education*, 3, 86. https://doi.org/10.3389/feduc.2018.00086

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Canty, D., Seery, N., & Hartell, E. (2017). Integrating peer assessment in technology education through Adaptive Comparative Judgment. *PATT34 Technology & Engineering Education–Fostering the Creativity of Youth Around the Globe*, *July*, 10–14. https://www.researchgate.net/publication/320299812

Chazan, D., & Ball, D. (1999). Beyond Being Told not to Tell. *For the Learning of Mathematics*, *19*(2), 2–10.

Durkin, K., Star, J. R., & Rittle-Johnson, B. (2017). Using comparison of multiple strategies in the mathematics classroom: Lessons learned and next steps. *ZDM - Mathematics Education*, *49*(4), 585–597. https://doi.org/10.1007/s11858-017-0853-9

Gray, M. E., & Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain, and Education*, *15*(3), 250–263. https://doi.org/10.1111/mbe.12288

Kimbell, R. (2020). Capability, Quality and Judgement: Learners' Experiences of Assessment. In P. Williams & D. Barlex (Eds.), *Pedagogy for Technology Education in Secondary Schools. Contemporary Issues in Technology Education* (pp. 201–217). Springer. https://doi.org/10.1007/978-3-030-41548-8_11

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Marton, F. (2015). *Necessary conditions of learning*. Routledge.

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2016). Supporting mathematical discussions: The roles of comparison and cognitive load. *Educational Psychology Review*, *29*(1), 41–53. https://doi.org/10.1007/s10648-016-9382-2

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory. In J. Spector & S. Ljoie (Eds.), *Explorations in the Learning Sciences, Instructional Systems and Performance Technologies, 1*. Springer.

van Daal, T., Snajder, M., Nijs, K., & Van Dyck, H. (2023). Peer Assessment Using Criteria or Comparative Judgement? A Replication Study on the Learning Effect of Two Peer Assessment Methods. In O. Noroozi & B. De Wever (Eds.), *The Power of Peer Learning* (pp. 73–101). Springer International Publishing. https://doi.org/10.1007/978-3-031-29411-2_4

Proceedings of the Australian Conference on Science and Mathematics Education, The University of Tasmania, 30 August – 1 September 2023, pages 77-82, ISSN 2653-0481.