Comparative judgement and affect: A case study

Jennifer Palisse
The University of Melbourne
<jennifer.palisse@student.unimelb.edu.au>

Deborah King
The University of Melbourne
<dmking@unimelb.edu.au>

Mark MacLean
The University of British Columbia
<maclean@math.ubc.ca>

Comparative judgement is a relatively new way of facilitating peer-assessment where students are shown pairs of other students' work and judge which of the two is *better*. Literature on example-based learning suggests that students should be able to learn from comparative judgement. We present the case of one student, Josie, whose understanding of rational inequalities did not improve while assessing other students work comparatively. We argue that her self-explanation attempts were limited because comparative judgement created an environment that threatened her goal of understanding. Knowing the correct answer may have helped alleviate some of these issues.

Comparative judgement is increasingly being used as a digital peer-assessment tool where students evaluate, provide feedback, and rank pairs of other students' work. It involves showing pieces of work as pairs, side-by-side, where students select which of the two they feel is *better*. This relies on the fact that humans are more reliable at comparing two objects against a given quality, for example, which object is heaviest, than they are at making an absolute judgement, such as stating how heavy an object is in kilograms (Thurstone, 1927).

In the context of mathematics education, comparative judgement has been used successfully to assess problem solving (Jones et al., 2014), students' understanding of fractions (Jones & Wheadon, 2015), and proof comprehensions (Davies et al., 2020), as well as understanding of calculus (Jones et al., 2019, 2013). Each of these studies focused on establishing reliability, that is, to demonstrate that the rankings produced by comparative judgement agree with those that would have been produced using traditional marking methods. High inter-rater reliability has consistently been demonstrated.

Other studies have explored what type of construct comparative judgement might be measuring. For example, Jones et al. (2013) advocate for the potential of comparative judgement to measure what they term as conceptual understanding. In their study, teachers were asked to give an estimate of their middle school students' mathematical ability. Students were then given a task where they were asked to order a set of fractions from smallest to largest and explain their solution. Teachers then used comparative judgement to rank students' responses. Jones et al. found that teachers' estimates of mathematical ability were a better predictor of final rankings than task accuracy, and argue that comparative judgement measures something other than procedural understanding.

Given the large number of studies already available exploring issues of reliability and validity, this paper instead focuses on recent calls to explore the potential use of comparative judgement as a learning tool rather than an assessment tool (e.g., Strimel et al., 2020). As of yet, there has been minimal research looking at the role comparative judgement might play pedagogically, particularly in the context of mathematics. From the few studies that do exist, it seems that students both in the secondary school setting (Jones & Alcock, 2013) and university setting (Potter et al., 2017) find pairwise peer-assessing valuable and worthwhile.

2021. In Y. H. Leong, B. Kaur, B. H. Choy, J. B. W. Yeo, & S. L. Chin (Eds.), *Excellence in Mathematics Education: Foundations and Pathways (Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia)*, pp. 313-320. Singapore: MERGA.

In the context of design, Bartholomew et al. (2019) have shown that comparative judgement can improve student outcomes, where students who participated in pairwise peer-assessment outperformed students who assessed their peers' work one at a time. This suggests that comparative judgement influences learning beyond the process of just giving and receiving feedback.

In this paper, we add to the limited knowledge of comparative judgement as a learning tool by presenting the case of one student, Josie, who did not find comparative judgement helpful. We present an analysis of Josie's experience to better understand what aspects of the comparative judgement process were helpful or otherwise. We suggest that the way in which we organised our comparative judgement activity may have limited Josie's willingness to engage in deep reflection, reducing the potential for learning to occur.

Before reviewing the case of Josie, we draw on the framework of learning from worked examples to provide some justification for why comparative judgement might be useful for students' learning.

Learning from Worked Examples

Learning from worked examples involves providing students with the following: the problem itself, the steps taken to reach a solution, and the final solution (Renkl, 2014). In a seminal study, Sweller and Cooper (1985) compared learning through worked examples and learning through problem solving in mathematics and found that learning through worked examples required less time to process, problems were solved faster, and students made less mathematical errors. Since then, several studies have replicated these findings (for an overview, see Renkl, 2014).

The reasoning behind why example-based learning is effective draws on *Cognitive Load Theory* (Sweller et al., 1998). Learning by practicing problems without the guidance of a worked example places heavy demands on working memory. To solve a given problem, learners must work on finding a set of steps that can lead them to the desired goal. Because such an approach imposes a heavy load on working memory it generally does not lead to learning. Studying worked examples reduces cognitive load. Learners no longer need to find a set of required steps themselves as these are included within the worked examples. Learners can then avoid searching their own prior knowledge for solution methods. This thereby reduces working memory leaving cognitive resources available for self-explanation (Sweller et al., 1998). This is important as only when self-explanation is encouraged, that is, when students explain the reasoning behind a solution to themselves, do worked examples appear to be effective (Chi et al., 1989; Renkl, 2014).

We hypothesise that comparative judgement, as with worked examples, may also facilitate learning, where the solutions students compare are analogous with worked examples used during example-based learning. One key difference is the emphasis on *comparing*, however, this too has supportive evidence for learning in the literature on example-based learning. Empirical studies have demonstrated that presenting worked examples side-by-side, rather than one at a time, consistently results in greater learning gains (Star & Rittle-Johnson, 2009). By having two worked examples available simultaneously, the learner is no longer required to hold a representation of the previous worked example active in their working memory in order to compare with the next example. The effect is to reduce cognitive load making learning more likely. To put it simply, in the words of Rittle-Johnson and Star (2009, p. 529) "Experts agree; comparison is good."

Lastly, comparative judgement includes both correct and incorrect examples. The framework of example-based learning suggests that providing students with both correct and incorrect examples is more beneficial than providing correct solutions only since incorrect solutions help students recognise incorrect strategy choices by drawing attention to the feature of the problem that makes the strategy inappropriate (Booth et al., 2013).

Research Design

The results reported here form part of a larger pilot-study. As such, we limit our explanation of our research design and detail only aspects needed to understand this single case study.

Participants included first-year undergraduate students studying an entry level calculus course. Eight students participated in an individual problem-based interview. Students completed a pre-task that asked them to solve the rational inequality $\frac{x+1}{x-7} > 3$. No feedback was provided, and they were not told whether they had answered the problem correctly or incorrectly. The students were then shown six pairs of other students' work on the same problem. For each pair, students selected which of the two they thought was the better solution. What *better* meant was up to students to decide. Students decided based on a variety of factors including the choice of method, whether any mistakes had been made, and even the neatness of the handwriting. Students were not provided with a rubric, marking scheme, or correct answer to help inform their decisions. While assessing each solution, students were asked to think-aloud. The think-aloud method involves verbalising one's thoughts when first noticed and is seen as a valid way to capture individuals' working memory (Ericsson & Simon, 1993). Once students completed their pairwise judgements, they then completed a similar problem and solved $\frac{5x-2}{x+5} > 6$. This allowed for comparison between students' pre- and post-tasks to analyse any changes students might have made between tasks. A short semi-structured interview followed where we asked students to explain any changes they made to their solution technique between their pre- and post-tasks.

Analysis

From the literature, we anticipated that comparative judgement should be useful in improving students' understanding of rational inequalities. To measure this, we intended to compare students' pre- and post-tasks to see if students who had made errors in their pre-task had rectified their mistakes in their post-task. Surprisingly, for students who held misunderstandings during the pre-task, we found little evidence for improved understanding in the post-task.

Since we were unsure why there had not been any improvement our next steps were exploratory. We found students' think-aloud comments during comparative judgement enlightening and analysed students' comments using thematic analysis (Braun & Clarke, 2019). This involved a progressive process of systematically comparing and grouping segments of think-aloud data firstly into meaningful smaller groups, and later into broader themes. These themes were developed inductively from the data. We present two themes here: the frustrations caused by not knowing the correct answer, and the lack of willingness to interpret other students' solutions if knowledge from external, rather than internal, sources was expected.

Below is our analysis of one student, Josie, who had not found comparative judgement helpful. Josie represented an atypical case since she was able to complete both the pre- and post-tasks but was still unable to understand the different approaches used in the solutions presented to her. As such, Josie was the only student for whom there was no evidence to suggest any improvement in understanding had resulted from the comparative judgement task. As Josie was forthcoming with her opinion, her case was useful in illustrating why comparative judgement may not be helpful for all students.

Results and Discussion: The Case of Josie

Background

Josie completed the pre-task by rearranging $\frac{x+1}{x-7} > 3$ as $\frac{x-11}{x-7} < 0$ and solving for when the numerator was positive and denominator negative, and vice versa. Unfortunately, she made a minor numerical error, leading to an incorrect final answer. This proved to be significant as her answer differed to most of the answers in solutions provided for the comparative judgement task. Josie noted while completing the pre-task that she had never understood how to solve rational inequalities and did not know why the same techniques used for solving equations could not be used to solve inequations. For example, Josie did not understand why you could not multiply both sides by $x - 7$ and why there would be "two answers". We refer to this type of approach as a two-cases approach.

Josie's goal of understanding and not knowing the correct answer

Josie was shown sample solutions from other students, presented in pairs, from which she was asked to choose which of the two solutions she felt was *better*. For Josie, the better solution tended to be the one with the correct answer.

> The thing I mostly go off is if the answer is right. If one was right, one was wrong, most of the time I picked the one that was right.

Not knowing for sure the correct final answer proved to be a major source of frustration for Josie as her strategy for choosing between the two solutions relied heavily on which solution was correct. This meant she was at risk of not making what she felt to be the correct choice between solutions. As a result, Josie did not trust her pairwise judgements.

> But if I actually knew the right answer, then I'd be like well that one's better because it's the right answer.

Having been denied access to the correct answer after multiple requests to the interviewer, Josie felt frustrated and as such, perceived it to be the interviewer's fault, not her own, that she was unable to make a correct pairwise choice. The above statement was said with a tone of irritation suggesting she felt a sense of pointlessness to the whole activity. From Josie's point of view, there was no point in engaging with the task if her pairwise choices were going to be wrong anyway. Without a means of resolving her uncertainties, the task had little perceived value.

This sense of pointlessness was again felt as Josie began working on the post-task. She believed her answer to the new problem to be incorrect and attributed this again to not having been provided with the correct answer to begin with:

> Now see if you'd told me how to solve this [pre-task] then I could probably solve this. So, I'm just going to be doing the wrong thing. [Huff] Is that right?

Josie continued working with a 'huff'. These statements carried a strong sense of exasperation and her final comment "Is that right?" was said sarcastically. By this late point in the activity, Josie was frustrated.

Josie's inability to resolve her uncertainties appeared to have affected her efforts in understanding the relevant mathematical concepts. In one example, Josie had tried to understand why the student had solved for two-cases but ended her reflection shortly:

> I'm just like sort of trying to tell myself what they're doing and why they're doing it. But I don't know if that's the right answer… [Moves onto next solution]

Here Josie made a sincere attempt to understand what the student had written but quickly gave up. Her comment about not knowing the right answer was said with a slight antagonistic tone directed towards the interviewer. From Josie's perspective, there was no point spending time interpreting the student's work if there was no means of checking the correctness of her interpretation.

In a second example, Josie tried to understand why a particular student's solution included two cases, noting two columns, one labelled $x - 7 > 0$ and $x - 7 < 0$. She quickly gave up, this time with self-deprecating comments:

> I'm just trying to work this out for my own benefit here and think, why are they doing the
>
> $x - 7 > 0$ [Points at two column headings]? [Silence] Honestly, I'm probably just too dense to get this. [Quickly starts evaluating next solution]

Josie had been silent for some time, suggesting she had been deep in thought and genuinely tried to work out why the student was solving for a second case. Additionally, her self-criticism was not uttered with her usual air of exasperation, but instead said with a more sombre tone, suggesting a sense of hopelessness or even sadness. We interpret this to mean that Josie really had tried to understand the need for a second case but felt unable to do so. Furthermore, the negative emotional response suggests Josie may have been experiencing fear as she perceived her goal of understanding to be at risk. As such, Josie may have felt her own reasoning was not a safe strategy (Sumpter, 2013) and rather than risk investing more time into resolving her confusion, she avoided this situation by quickly moving on to evaluating the next solution.

Together these instances paint a picture of a student who found comparative judgement frustrating primarily because she had no internal means available to resolve her uncertainties. This subsequently created an environment that repeatedly placed Josie's goals of wanting to understand the problem under threat. As a coping strategy, Josie often appealed to not having the correct answer as the cause of not being able to understand. By placing blame on external factors, that is, not having the answer, Josie was able to increase her own feelings of safety rather than risk relying on her own reasoning. As a result, Josie seemed to have viewed comparative judgement as somewhat pointless for learning. From Josie's perspective, what was the point of investing time and effort into understanding someone else's solution if her reasoning of what the student had written could not be verified? This provides some explanation into why comparative judgement was not helpful in this case: if there was no point spending time interpreting each worked solution, then self-explanations were unlikely to be generated. With no self-explanations, learning from worked examples was unlikely to be successful (Chi et al., 1989). As such, if comparative judgement is to be used with the educational goal of improving mathematical understanding, we recommend providing students with, at minimum, the correct answer before completing any pairwise judgements. We suspect had Josie known the correct answer with certainty, she might have been willing to spend more time reflecting and interpreting what other students had written, making learning more likely.

Josie's views on the role of the expert and usefulness of other students' solutions

In this section we discuss who Josie felt was responsible for explaining mathematical ideas so that she could understand them, and whether she felt comparative judgement was effective in improving her understanding.

Ultimately, Josie felt the responsibility for her understanding of the mathematics content lay with her lecturers – an external source of expertise:

> I guess I don't have as good an understanding of inequalities as I should. I guess you can't multiply across like that [multiply by $x - 7$ without considering both $x - 7 > 0$ and $x - 7 < 0$]. I don't know why you can't. That never really got explained to me.

Furthermore, she felt frustrated with her mathematics lecturers, who she felt were not providing adequate explanations:

> Cause there's a lot of stuff I don't understand, and the lecturers aren't really thorough with explaining it.

From Josie's point of view, the responsibility for generating understanding came from her lecturers. It was their role to provide a thorough enough explanation to generate understanding and her role to receive such knowledge. We suggest that by placing expectations of understanding on an external source, it increased Josie's feelings of psychological safety by removing her ownership of her own understanding. By claiming her lecturers had failed in providing her with suitable explanations, it may have resolved her experience of doubt by surrendering control to an external authority (Bendixen, 2002). From Josie's point of view, it was not her fault if she could not understand when the responsibility for understanding lay with her lecturers.

These sentiments were echoed as Josie explained why she felt comparative judgement had not been helpful for her understanding. Firstly, Josie acknowledged that comparative judgement had been useful in pointing out that her initial assignment had been flawed. However, it had not been useful for understanding why they were flawed:

> But yeah it definitely helped me figure out oh yeah, I need another answer. Stuff like that. But it didn't help me understand why.

Here Josie was referring to when she had completed a comparative judgement activity as part of a class assignment earlier in the semester. For her assignment, she had multiplied both sides of the inequality by the denominator but failed to consider when the denominator was either positive or negative, yielding the partial solution $x < 11$. By looking at the solutions of her peers, she realised that there was "another answer".

Josie continued explaining why comparative judgement might not have helped her understanding:

> Well there's either really smart kids that just go bang, bang, bang. That's the answer and they don't really explain it. I don't understand what you're doing [the smart kids]. Or they're just not getting the right answer in general.

Given it was likely Josie believed the role of her lecturers was to provide her with an explanation, it may be that as she read these solutions, she similarly felt that it was the responsibility of the student (as an external source of knowledge) to provide all necessary explanations. That is, it was not her role to spend time interpreting the solutions but rather the solutions should have been presented in a way that did the thinking for her. Because of the value Josie placed on figures of authority such as her lecturers, it was surprising to see that Josie did not value the work of the "really smart kids". This speaks to the interplay between Josie's views on the role of the expert and her goal of understanding. While Josie viewed her role as the receiver of knowledge from those more knowledgeable than her, she simultaneously expected the knowledge giver to provide such knowledge in a way that was easy to understand. This meant that for Josie, comparative judgement failed to provide an environment conducive to learning as the solutions were not written in a clear and easy to understand manner.

Lastly, linking back with our previous theme, Josie implied that incorrect solutions were not useful in helping her mathematical understanding. Because of the emphasis she placed on knowing the correct final answer, it was unlikely that Josie thought she could learn from incorrect examples. However, even the correct worked solutions seemed equally unvalued as they did not include enough detail or explanation to help resolve Josie's knowledge gaps ("they don't really explain it"). This is true - none of the worked solutions in this study included much detail or written explanation as to exactly why two cases needed to be considered. We found this surprising given that research in the area of learning from worked examples suggests that learning from incorrect examples is often more beneficial to learning than correct examples only (Booth et al., 2013). Results here suggest that this may not carry into the context of comparative judgement. This might be because the comparative judgement solutions we used were not labelled as either correct or incorrect, whereas studies exploring incorrect worked examples typically label such solutions as incorrect. One avenue of future research could be to explore whether marking comparative judgement solutions as either correct or incorrect has any influence on how students interact with the written solutions.

Final Comments

For Josie, comparative judgement was useful in helping her notice that she had been incorrect, but not useful in improving her understanding of solving rational inequalities. We noted two reasons for this. Firstly, it seemed that comparative judgement threatened Josie's goals of understanding – Josie did not feel her own reasoning was a safe strategy and did not feel she had a way of resolving her uncertainties. Josie's coping strategy was to blame the interviewer for not providing her with the answer. As Hannula (2006, p. 169) states "students may decide not to pursue learning goals when they feel that one or more of their psychological needs are thwarted."

Secondly, learning from other students' solutions did not appear to align with Josie's expectations of where knowledge comes from. From Josie's perspective, it was the role of the worked solution to provide her with a clear and easy to understand explanation and her role to receive such knowledge. In short, she expected to understand each solution without needing to think and the solutions we included in this study lacked the type of explanation Josie expected.

What ties these themes together is the interplay between Josie's desire to understand but not wanting to take ownership of generating this understanding. This seems to have impacted her ability to act strategically, limiting the amount of time she spent reflecting on the underlying mathematics, making learning from the worked solutions less likely. As such, the way in which we set up our comparative judgement activity may not be appropriate for the purposes of improving understanding for students like Josie.

References

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

Bendixen, L. D. (2002). A Process Model of Epistemic Belief Change. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing* (pp. 191–208). Erlbaum.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34. https://doi.org/10.1016/j.learninstruc.2012.11.002

Braun, V., & Clarke, V. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (Vol. 16, Issue 3, pp. 57–71). Springer. https://doi.org/10.1037/13620-004

Chi, M., Bassok, M., & Lewis, M. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *182*, 145–182.

Davies, B., Alcock, L., & Jones, I. (2020). Comparative judgement, proof summaries and proof comprehension. *Educational Studies in Mathematics*, *105*(2), 181–197. https://doi.org/10.1007/s10649-020-09984-x

Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis: Verbal reports as data (revised edition). In *Overview of Methodology of Protocol Analaysis*. Bradford Books/MIT Press.

Hannula, M. S. (2006). Motivation in Mathematics: Goals Reflected in Emotions. *Educational Studies in Mathematics*, *63*(2), 165–178.

Jones, I., & Alcock, L. (2013). Peer assessment without assessment criteria. *Studies in Higher Education*, *39*(10), 1774–1787. https://doi.org/10.1080/03075079.2013.821974

Jones, I., Bisson, M.-J., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal*, *45*(3), 662–680. https://doi.org/10.1002/berj.3519

Jones, I., Inglis, M., Gilmore, C., & Hodgen, J. (2013). Measuring conceptual understanding: the case of fractions. In A. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education (PME 37)* (Vol. 3, pp. 113–120).

Jones, I., Swan, M., & Pollitt, A. (2014). Assessing Mathematical Problem Solving Using Comparative Judgement. *International Journal of Science and Mathematics Education*, *13*(1), 151–177. https://doi.org/10.1007/s10763-013-9497-6

Jones, I., & Wheadon, C. (2015). Peer assessment using comparative and absolute judgement. *Studies in Educational Evaluation*, *47*, 93–101. https://doi.org/10.1016/j.stueduc.2015.09.004

Potter, T., Englund, L., Charbonneau, J., MacLean, M. T., Newell, J., & Roll, I. (2017). ComPAIR: A New Online Tool Using Adaptive Comparative Judgement to Support Learning with Peer Feedback. *Teaching & Learning Inquiry*, *5*(2), 89–113.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Rittle-Johnson, B., & Star, J. R. (2009). Compared With What? The Effects of Different Comparisons on Conceptual Knowledge and Procedural Flexibility for Equation Solving. *Journal of Educational Psychology*, *101*(3), 529–544. https://doi.org/10.1037/a0014224

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, *102*(4), 408–426. https://doi.org/10.1016/j.jecp.2008.11.004

Strimel, G. J., Bartholomew, S. R., Purzer, S., Zhang, L., & Ruesch, E. Y. (2020). Informing engineering design through adaptive comparative judgment. *European Journal of Engineering Education*, 1–20. https://doi.org/10.1080/03043797.2020.1718614

Sumpter, L. (2013). Themes and Interplay of Beliefs in Mathematical Reasoning. *International Journal of Science and Mathematics Education*, *11*(5), 1115–1135.

Sweller, J., & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, *2*(1), 59–89.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251-296. https://doi.org/10.1023/A:1022193728205

Thurstone, L. (1927). A Law of Comparative Judgement. *Psychological Review*, *34*, 274–286.