Using Comparative Judgment to Assess the Quality of Explanations in Mathematics

Juan Pablo Mejía Ramos
Rutgers University

Tanya Evans
University of Auckland

Colin Jakob Rittberg
Vrije Universiteit Brussel

Matthew Inglis
Loughborough University

Keywords: Comparative Judgment, Explanation, Proof.

We describe evidence of the usefulness of the Comparative Judgment (CJ) approach to assess the quality of explanations in mathematics. In philosophy, some study how certain types of mathematics offer *explanations* of other mathematical phenomena. In the field of education, offering explanations is a central part of teaching mathematics, and understanding those explanations is a vital activity for learners. So, what makes a good explanation in mathematics?

In the philosophy of mathematics, explanation quality has been traditionally studied by investigating the properties of what are deemed to be exemplars of mathematical explanations (e.g., Steiner, 1978), often mathematical proofs. In education, the approach has been to use *general* frameworks (i.e., non-math specific) describing the features that high-quality instructional explanations may have (e.g., Wittwer & Renkl, 2008).

We propose a different approach: using the CJ method to explore the notion of explanation quality *as it exists among mathematicians and undergraduate students*. CJ approaches to understanding human judgment exploit the finding that people are better at comparing two objects against each other than at evaluating one object against specific criteria (Thurstone, 1927). Modern uses of comparative judgment in assessment rely upon the Bradley-Terry model (Bradley & Terry, 1952), which assumes that each stimulus has a numerical parameter which captures its quality on the dimension of interest. By presenting judges with repeated pairs of stimuli and asking them to assess which they would rate higher on the given dimension, empirical estimates of these parameters can be obtained.

Recent empirical studies (Mejía Ramos et al., 2021; Evans et al., 2022) have used the CJ approach to study the extent to which mathematicians and students can reliably judge the quality of explanations in mathematics. In the first study, 38 mathematicians made a total of 760 comparisons between two proofs of the same proposition. For each judgment, two proofs were randomly selected and positioned side-by-side for the participant to think about and select which argument best explained why the proposition holds (without focusing on how the proof might be received by a particular audience), an instruction aimed at investigating the more philosophical sense of explanation described above. Using a similar procedure, in the second study we asked 32 undergraduate students who had taken a Linear Algebra course and 16 mathematicians to judge explanations meant for a hypothetical mathematics undergraduate student who did not understand the concept of an abstract vector space, an instruction aimed at investigating the more pedagogical sense of explanation described above. Using split-half inter-reliability coefficients, we report very high levels of reliability in participants' judgments of explanation quality for both mathematicians and students. Furthermore, in the second study we found a very high correlation between the parameter estimates for mathematicians and those for students. Overall, these studies provide evidence of high level of agreement among participants regarding what makes a good explanation in mathematics. This method opens fascinating avenues for future research on the notion of explanation in mathematics.

References

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika, 39*, 324–345.

Evans, T., Mejía Ramos, J. P., & Inglis, M. (2022) Do mathematicians and undergraduates agree about explanation quality. *Educational Studies in Mathematics, 111*, 445-467.

Mejía Ramos, J. P., Evans, T., Rittberg, C., & Inglis, M. (2021). Mathematicians' assessments of the explanatory value of proofs. *Axiomathes, 31*, 575-599.

Steiner, M. (1978). Mathematical explanation. *Philosophical Studies, 34*(2), 135–151.

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

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49–64.