Educational Studies in Mathematics (2022) 111:445–467
https://doi.org/10.1007/s10649-022-10164-2

Do mathematicians and undergraduates agree about explanation quality?

Tanya Evans[1] · **Juan Pablo Mejía-Ramos**[2] · **Matthew Inglis**[3]

Accepted: 30 May 2022 / Published online: 29 July 2022

Abstract
Offering explanations is a central part of teaching mathematics, and understanding those explanations is a vital activity for learners. Given this, it is natural to ask what makes a good mathematical explanation. This question has received surprisingly little attention in the mathematics education literature, perhaps because the field has no agreed method by which explanation quality can be reliably assessed. In this paper, we explore this issue by asking whether mathematicians and undergraduates agree with each other about explanation quality. A corpus of 10 explanations produced by 10 mathematicians was used. Using a comparative judgement method, we analysed 320 paired comparisons from 16 mathematicians and 320 from 32 undergraduate students. We found that both mathematicians and undergraduates were able to reliably assess the quality of a set of mathematical explanations. Furthermore, the assessments were largely consistent across the two groups. Implications for theories of mathematical explanation are discussed. We conclude by arguing that comparative judgement is a promising technique for exploring explanation quality.

Keywords Mathematical explanation · Explanation quality · Mathematical practices · Undergraduate mathematics · Comparative judgement

✉ Tanya Evans
t.evans@auckland.ac.nz

Juan Pablo Mejía-Ramos
jpmejia@math.rutgers.edu

Matthew Inglis
M.J.Inglis@lboro.ac.uk

[1] Department of Mathematics, University of Auckland, 38 Princes Street, 1010 Auckland, New Zealand

[2] Graduate School of Education, Department of Mathematics, Rutgers University, 10 Seminary Place, 08901 New Brunswick, New Jersey, USA

[3] Department of Mathematics Education, Loughborough University, Epinal Way, LE11 3TU Loughborough, UK

1 Introduction

Explanations are central to both teaching and learning. Teachers and lecturers routinely offer instructional explanations as part of classroom practice, either during instructor-led exposition or in response to students' questions (e.g., Lew et al., 2016; Treagust & Harrison, 1999). However, explanations vary in quality, to the point where instructional explanations may sometimes have negative effects. Specifically, instructional explanations do not always lead to successful learning (Chi et al., 2001; Lew et al., 2016; VanLehn et al., 2003; Wittwer & Renkl, 2008), and generating low-quality explanations can in some circumstances limit understanding. For example, Lachner and Nückles (2016) conducted a classroom experiment to compare the effects of different explanations of an optimisation problem (involving finding extremum of a function) on senior high school students. Two types of explanations were compared: one produced by research mathematicians (emphasising principles and conceptual rationale), and the other by mathematics teachers (emphasing procedural steps of how to solve a problem). The students were randomly assigned to three groups, with two groups receiving explanations prepared by either mathematicians or teachers, and the third (control) group receiving no explanation. The researchers found that the students who received the explanations elucidating principles and conceptual rationale behind the steps (offered by mathematicians) outperformed the other students on a subsequent application test (Lachner & Nückles, 2016).

These findings emphasise the importance of taking into consideration explanation quality: in order to maximise student learning, instruction needs to be based on high-quality explanations and omit low-quality explanations. But, despite the importance of explanation quality, Wittwer and Renkl (2008) noted that the issue of what constitutes an effective explanation has been widely neglected by education researchers. Thus, when teachers and learners generate instructional explanations, they must typically rely upon experience and intuition rather than research-based advice.

Our goal in this paper is to consider the quality of instructional explanations from the perspectives of undergraduate mathematics lecturers and students. There are two broad approaches to analysing explanation quality that can be adopted. The first, which we describe as the top-down approach, is to evaluate the quality of instructional explanations using pre-existing general frameworks that attempt to describe the features that high-quality instructional explanations may have (e.g., Wittwer & Renkl, 2008). An alternative, which we describe as the bottom-up approach, would be to collect a corpus of instructional explanations, develop an empirical method by which mathematicians and students can assess their quality, and then study the features that high- and low-quality explanations have. In light of the lack of frameworks for instructional explanation quality rooted in the mathematics classroom, and in light of the strong possibility (discussed below) that mathematical explanations may be disanalogous to non-mathematical explanations (such as those in science), we adopted the bottom-up approach. That is, instead of attempting to characterise the quality of instructional explanations in mathematics in a top-down fashion, using principles of general frameworks, we aim to empirically explore this notion of explanation quality as it exists among mathematics lecturers and students. Our hope is that such bottom-up exploration can contribute to the more general characterisation of the quality of instructional explanations in mathematics.

More specifically, we address two main questions. First, can mathematicians reliably judge the quality of explanations? In other words, when asked to assess the quality of different mathematical explanations, do mathematics lecturers tend to agree with each other?

Second, do the judgements about explanation quality made by lecturers coincide with those of their intended audience, undergraduate students? Before introducing the study we conducted to address these issues, we first briefly review the existing literature on explanations in mathematics.

2 Philosophical and educational perspectives on mathematical explanations

At least in the area of proof-based mathematics, mathematics education researchers have traditionally turned to the philosophy of mathematics for insights into a notion of explanation in mathematical practice that could be useful to determine what makes a proof explanatory in the mathematics classroom (Hanna, 1990; Hersh, 1993; Weber, 2010). Philosophers have devoted a great deal of attention to understanding what it means to say that A explains B. For instance, some accounts in the general philosophy of science literature rely on either statistical associations or causal mechanisms. For instance, Salmon (1971, 1984) suggested that A explains B if B is consistently correlated with A, or if there is a causal history that connects B and A (cf. Hempel & Oppenheim, 1948). So, we can say that wearing shoes in the wrong size explains why our feet are hurting because there is a causal connection between the two events. But, while causal and statistical accounts may work well in scientific contexts, they seem to fail in mathematics. Mathematical concepts are not related causally, as there is no temporal order in the universe of mathematical definitions, theorems, and proofs. The fact that the square root of 2 is irrational is not located at a particular point in time. Neither are mathematical facts related statistically, as they take no probabilities other than 0 or 1 (i.e., a mathematical statement is either true or false). Consequently, many scientific accounts of explanation do not seem to easily apply in mathematical contexts (Colyvan, 2012; Mancosu, 2001).

If mathematical explanations are not scientific explanations, what are they? This question has generated significant interest. Some regard the lack of causal and correlational relations as a reason to deny that mathematical explanations exist, at least in a manner analogous to scientific explanation (Resnik & Kushner, 1987; Zelcer, 2013). This approach seems at odds with practice. Studies of mathematical language show that research mathematicians do use explanatory words when communicating with one another (Mejía-Ramos et al., 2019). And it is certainly at odds with educational practice. As noted above, explanations are central to mathematical teaching and learning, and many mathematics education researchers have emphasised the importance of engaging students with mathematical proofs that explain theorems, rather than those which merely demonstrate that theorems are true (e.g., Bell, 1979; Hanna, 2000). The desire to distinguish explanatory from non-explanatory proofs also rules out the proposal that A explains B if B can be logically deduced from A. Under such an account, all valid deductive arguments would be equally explanatory. The apparent ubiquity of explanations in mathematical discourse leads most philosophers to reject the suggestion that mathematical explanations do not exist (Weber & Frans, 2017) or that they are simply logical deductions (Colyvan, 2012). Instead, they offer two broad categories of account, which Delarivière et al. (2017) described as the *ontic* and *epistemic* approaches to mathematical explanation.[1]

[1] A similar distinction exists in the philosophy of science literature on scientific explanation.

Ontic accounts focus on the objective properties of purported explanations. For instance, Steiner (1978) suggested that mathematical explanations are arguments that refer to characterising properties of relevant concepts. Ontic accounts also include Kitcher's (1984) proposal that mathematical explanations are characterised by the way in which they unify a range of different mathematical concepts, and Lange's (2014) suggestion that mathematical explanations are characterised by their use of certain types of salient features (e.g., symmetries) of the explained result. What all ontic accounts share is the belief that the (non-)explanatoriness of a mathematical proof—and it is typically proofs that these accounts focus on, a fact that was critiqued as 'proof chauvinism' by D'Alessandro (2020)—can be assessed without considering its actual or potential audience. Despite the fact that ontic accounts are audience independent, mathematics education researchers have attempted to use them to address the question of what makes a proof explanatory in the mathematics classroom (Hanna, 1990; Reid, 1995).

In contrast, epistemic and, similarly, functional accounts start with the assumption that explanations are communicative acts that aim to cause understanding (Delarivière et al., 2017; Inglis & Mejía-Ramos, 2021). For example, Wilkenfeld (2014) defined explanations to be those things that generate understanding "under the right circumstances and in the right sort of way" (p. 3367). In order to unpack exactly what the right circumstances and sorts of ways are, Wilkenfeld drew heavily on the epistemology literature (e.g., Grimm et al., 2016).

Although the ontic and epistemic approaches differ in emphasis, they are not necessarily contradictory. Inglis and Mejía-Ramos (2021) offered a functional account which they suggested could encompass the ontic accounts of Steiner (1978), Kitcher (1984), and Lange (2014). Specifically, they argued that modern theories of human cognitive architecture imply that mathematical arguments, which make use of characterising properties, which unify concepts, or which involve Lange-style saliency, will typically help a reader to increase their level of understanding of the domain(s) in which the explanation is situated. In other words, the ontic properties identified by Steiner, Kitcher, and Lange are all likely to contribute to an argument's epistemic explanatoriness.

The fact that philosophers disagree about whether ontic or epistemic approaches to mathematical explanation are more promising raises the possibility that mathematicians too may adopt differing perspectives on explanatoriness. If this were correct, then we might expect there to be between-mathematician variation in the types of explanations that they deem most explanatory. Assessing the extent to which this is the case is one goal of the study reported in this paper. Adopting an epistemic approach to mathematical explanation requires specifying what type(s) of understanding it is that explanations try to foster. Inglis and Mejía-Ramos (2021) emphasised that, for them, explanations aim to increase *objectual understanding*, not merely *propositional understanding*. This distinction can be illustrated by comparing the statements "Tessa understands that there are five continents" (propositional understanding) and "Honali understands topology" (objectual understanding). Whereas objectual understanding admits degrees, propositional understanding does not. It would be straightforward to find someone whose understanding of topology was either higher or lower than Honali's, but it does not seem possible to understand that there are five continents more or less than anyone else. Although there are various different accounts of precisely how objectual understanding should be characterised (e.g., Grimm, 2006; Kelp, 2016; Kvanvig, 2003), a common theme is that the relationship between previously disconnected information is crucial. For instance, Kvanvig (2003) wrote that "the grasping of relations between items of information is central to the nature of understanding" (p. 197).

Given the importance of explanations for educational practice, and given the unresolved philosophical debates about what exactly mathematical explanations are (and even whether or not they exist), it is perhaps surprising that more attention has not been devoted to understanding explanation quality in educational settings. One exception to this rule is Wittwer and Renkl's (2008) general framework for exploring the effectiveness of (not necessarily mathematical) instructional explanations. They defined instructional explanations as being explanations given in educational contexts designed for the purpose of teaching, and gave four normative criteria by which their quality can be interrogated.

1. *Explanations should take account of learners' existing knowledge.* If the goal of an explanation is to increase objectual understanding, and if increasing objectual understanding involves developing relationships between previously disconnected knowledge schemas, then it seems self-evident that instructional explanations should be designed with reference to the learner's existing knowledge (e.g., Leinhardt & Steele, 2005). This assumption has been studied empirically. For example, Duffy et al. (1986) compared the verbal explanations given by more and less effective fifth-grade teachers. They found that the more effective teachers would spontaneously adapt their explanations in light of interactions with students, in order to respond to student misunderstandings. The association between an instructor's knowledge of their students' knowledge levels and the quality of explanations they produce is causal. Wittwer et al. (2008) experimentally manipulated instructors' beliefs about their students' knowledge levels in an online tutoring environment. They found that those instructors given inaccurate information about their students' existing knowledge produced less appropriate explanations, as assessed by the extent to which students increased their knowledge.
2. *Explanations should focus on concepts and principles.* Because increasing objectual understanding involves constructing new relationships between knowledge, explanations which focus on generalisable concepts or principles are likely to help learners integrate more of their existing knowledge. Research has shown that principle-oriented explanations promote learners' mathematical understanding since they integrate conceptual and procedural information, thus making them particularly tangible for novice students (Lachner et al., 2019). It is well documented, however, that teachers often omit conceptual explanations when explaining procedures, which is detrimental to learning (Lachner & Nückles, 2016; Lachner et al., 2019). There are large between-instructor differences in the extent to which explanations focus on generalisable principles. For example, Perry (2000) compared the explanations offered by teachers in first- and fifth-grade classrooms in Japan, Taiwan, and the USA. She found that teachers in the Asian classrooms offered more explanations, but also that their explanations were much more likely to generalise beyond the specific problem being discussed. Perry (2000) suggested that this difference in explanation quality might partly account for differences in performance on international comparisons found between the USA and Asian countries.
3. *Explanations should be integrated into the learner's ongoing cognitive activities.* In line with the emphasis that constructivist accounts of learning place on active processing (e.g., Fiorella & Mayer, 2015), if learners actively engage with the information in an explanation—rather than just listen or read it—a higher level of objectual understanding is likely to be reached. For instance, Webb et al. (1995) investigated seventh-grade students' activity after receiving instructional explanations from their teachers. The extent to which students continued to work on problems after receiving an explanation,

by either solving the problem or explaining how to solve the problem, was strongly predictive of scores on a subsequent assessment of their learning.

4. *Explanations should not replace learners' knowledge-construction activities.* The last of Wittwer and Renkl's (2008) criteria for effective explanations concerned when explanations should *not* be provided. If the provision of an explanation leads to lower levels of engagement with the to-be-learned material, then there is likely to be less learning, regardless of the quality of the provided explanation. For instance, Roy et al. (2017) found that providing undergraduates with verbal explanations of steps in a proof led to lower comprehension retention than if no such explanations were provided. They hypothesised that the extra support the explanations provided disrupted the students' engagement with the proof, reducing the extent to which they were able to integrate their learning with existing knowledge (Alcock et al., 2015).

It is worth noting two aspects about Wittwer and Renkl's (2008) framework. First, their four criteria are audience dependent (even the second criterion is ultimately related to learners' understanding in terms of, and knowledge construction around, these concepts and principles). In this sense, this framework is more closely related to epistemic approaches of mathematical explanation in the philosophy of mathematics (albeit, specific epistemic approaches could operationalise "understanding" differently, e.g., purely in terms of abilities as opposed to cognitive structures and activities). In contrast, from an ontic perspective, these criteria (particularly the first, third, and fourth) are simply irrelevant in the characterisation of mathematical explanation (i.e., from an ontic perspective, there would be nothing wrong if a given mathematical explanation meets those criteria, but those criteria would certainly not be part of the defining features of mathematical explanation). Second, it is important to note that while Wittwer and Renkl's (2008) framework provides us with general "evidence-based conjectures about how to use instructional explanations to effectively support understanding and learning" (p. 51), it would be difficult to use this framework to compare the quality of specific instructional explanations in a given mathematics classroom (e.g., the individual criteria are somewhat general, and the framework does not assign differential weights to each of them).

The current study employs a novel approach to the investigation of explanation quality in mathematics. In an attempt to complement the top-down application of general characterisations of explanation quality from the philosophy of mathematics and educational psychology literature, we investigated the viability of a bottom-up investigation into the notion of explanation quality as it exists among mathematics lecturers and students. A natural first question in this bottom-up approach concerns the level of agreement in mathematics lecturers' and students' assessments of explanation quality in mathematics. The goal of the study reported in this paper is to tackle this question. Specifically, we asked whether university mathematics lecturers and undergraduate students are able to reliably judge explanation quality. In turn, we argue that this empirical investigation has implications for both the philosophy of mathematics and mathematics education. If philosophical accounts of mathematical explanation wish to reflect mathematical practice (Hamami & Morris, 2020; Van Bendegem, 2014), then it is important for philosophers to understand what mathematicians themselves consider to be explanatory. Studies seeking to assess the reliability of mathematicians' judgements in this domain therefore have the potential to constrain existing and future philosophical accounts of mathematical explanation (we return to this point in the discussion). In the case of

mathematics education, and indeed educational psychology more generally, results from the current study may help rule out one possible reason why instructors often produce explanations that do not effectively support learning (Chi et al., 2001; Lachner & Nückles, 2016; Lachner et al., 2019; VanLehn et al., 2003; Wittwer et al., 2008), namely that no one in the classroom (neither instructors nor students) can reliably distinguish high-quality explanations from low-quality ones (i.e., that assessing explanation quality generates large between-instructor, between-student, and between-instructor-student disagreements).

3 Creating a corpus of explanations

To achieve our goal of investigating the reliability of mathematicians' and undergraduates' judgements of explanation quality, we first created a corpus of mathematical explanations. Specifically, we aimed to collect mathematicians' responses to a prompt for an explanation of a mathematical concept. To this end, we designed a short online survey.

Participants in the survey, who received invitations to participate from a member of the research team by email, were all research-active mathematicians working at universities in New Zealand, the USA, Canada, Australia, Singapore, Belgium, and Finland. After giving consent to participate, participants provided us with some demographic information about their research area, and then clicked through to a page with the following prompt:

> Imagine that a math major on your linear algebra course comes to your office hours and says that they are confused. They explain that although they have seen the definition, they do not understand what an abstract vector space is, or what it is for. What explanation would you give the student in response? (Feel free to use tex or pidgin tex in your response.)

They were able to respond using a free-text box. This prompt, which referenced the student's lack of objectual understanding of vector spaces, was designed to describe a realistic scenario in which a university lecturer might be asked for an explanation by a student. We were careful to set the context for the required explanation: a mathematics major taking a linear algebra course, who had seen the definition of a vector space, but who did not understand it and therefore had come to the lecturer's office hours session.

A convenience sample of twenty mathematicians participated. A variety of different explanations were offered. One participant offered a diagram as part of their explanation (which was sent directly to the researchers outside of the Web form). From the twenty explanations offered, we selected a corpus of ten. These ten were chosen to reflect the full variety of the explanations offered. In other words, we removed extremely similar explanations.

The ten selected explanations, which we edited lightly for clarity, are given in the Appendix. The original versions of all twenty explanations are available in the online dataset which accompanies this article.[2] The ten explanations varied considerably in content and other elements. For instance, some involved procedural aspects (e.g., Explanations 1, 7, and 8), whereas others were entirely conceptual (e.g., Explanation 2). Some explanations were primarily geometric (e.g., Explanations 9 and 10), whereas others were focused on

[2] https://doi.org/10.17028/rd.lboro.14213831.v1

building on numerical intuitions (e.g., Explanation 8). The main feature of many explanations was highlighting similarities between examples (Explanations 3, 4, 5, 6, and 9) and one used a non-example (of a set that is not a vector space) to illustrate the concept boundary (Explanation 1). Utility considerations were present in many explanations, accentuating the purpose of the definition and exemplifying its possible uses (e.g., Explanations 1, 4, 5, 6, and 9). Moreover, some explanations explicitly involved learner-instructor interactions such as posing questions and providing further information depending on the answers received.

It is unsurprising that some of the content-related aspects of the ten explanations obtained have been widely discussed in the mathematics education literature. Numerous studies have emphasised the importance of the use of examples in mathematical learning in order to develop conceptual understanding (e.g., Bills & Watson, 2008; Fukawa-Connelly & Newton, 2014; Tall & Vinner, 1981), with the role of non-examples deemed necessary to gain a coherent concept image (Fukawa-Connelly & Newton, 2014; Goldenberg & Mason, 2008; Tall & Vinner, 1981). The importance of geometrical interpretations to suplement explanations and the use of diagrams and graphs has also been identified and discussed widely, emphasising their functional role in promoting learning (e.g., Mejía-Ramos & Weber, 2019; Samkoff et al., 2012; Soto-Johnson & Troup, 2014).

4 Comparative judgement

To assess the extent to which mathematicians and undergraduates can reliably assess the quality of explanations, we adopted a comparative judgement approach (Bisson et al., 2016; Jones et al., 2019, 2015). Comparative judgement approaches to assessment rely upon the finding that people are more accurate when comparing items than they are when asked to evaluate an item in isolation (Thurstone, 1927a). For example, it is easier to decide which of two weights is the heavier than it is to estimate the weight of one in isolation. Thurstone (1927b) referred to this observation as the 'law of comparative judgement' and used it to assign values to a variety of stimuli. For instance, in one study, Thurstone (1927b) was able to measure the perceived seriousness of different criminal offences (libel, perjury, smuggling, etc.) by asking college students to engage in a series of paired comparisons and fitting the resulting data to his psychophysical model.

Modern uses of comparative judgement in assessment rely upon the Bradley-Terry model (Bradley & Terry, 1952). This assumes that each item i (indexed by a positive integer) has a numerical parameter β_i which captures its quality on some dimension of interest. In our case, this is 'explanatoriness'; in Thurstone's (1927b), the dimension was 'seriousness of offence'. Given two items i and j, then the probability that a judge regards i as being rated higher on the given dimension is given by

$$\mathbb{P}(i > j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}.$$

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 the β_i can be obtained. Jones et al. (2019) suggested that an average of 10 judgements per item usually suffices to produce an accurate estimate of each β_i.

Comparative judgement techniques are becoming common in educational assessment contexts. They have been used to assess the quality of student essays (Heldsinger &

Humphry, 2013) and laboratory reports (McMahon & Jones, 2015), as well as more nebulous constructs such as students' conceptual understanding (Bisson et al., 2016), students' problem solving skills (Jones & Inglis, 2015), and mathematicians' conceptions of mathematical proof (Davies et al., 2021). The method is particularly helpful when one wishes to assess constructs about which people are expected to have an intuitive understanding, but which they may not be able to fully articulate or use to make reliable absolute judgements (Pollitt, 2012).

Critically, comparative judgement can also be used to produce estimates of the reliability of judges' judgements. A variety of reliability coefficients can be produced from comparative judgement data, but here we focus our attention on intuitively straightforward split-half inter-rater reliability coefficients (Bisson et al., 2016). To calculate such a coefficient, one randomly selects two groups of judges from the total set (typically half in each group) and fits the Bradley-Terry model separately for each group. This produces two β_i estimates for each of the judged items, one derived from each group of judges. The correlation between these two sets of β_i produces an estimate of the extent to which the two groups of judges agree with each other about the construct being assessed. Repeating this procedure 1000 times—with a new random split each time—and taking the average of the correlations yields an overall estimate of the reliability of the judges. A coefficient close to 1 indicates that the judges tend to agree with each other about the construct being assessed, whereas a coefficient close to 0 indicates little or no between-judge agreement. This is the method we used in our study.

5 Method

Two groups of participants took part in the judging session, where they were asked to judge which of two explanations was the better. Eighteen research mathematicians from the University of Auckland participated, following an email to relevant colleagues in the department. None had participated in the earlier Web survey that generated the corpus of explanations. Each participant was asked to complete 20 judgements, but one did not and was removed from the analysis. One further participant responded extremely quickly to each comparison (a mean of 8.1 s per judgement, compared to the average of 31.9 s) and was also removed from the analysis. Therefore a total of 320 judgements from 16 mathematicians were included in the final analysis, with a mean of 33.3 s per judgement.

We also recruited undergraduates who had recently taken a linear algebra course of the type mentioned in the explanation prompt discussed in Section 3. These participants were studying mathematics at either the University of Auckland or Rutgers University. We aimed to collect the same number of judgements as we had collected from the mathematicians (320), but following feedback from several mathematician participants about the length of the study, we asked each undergraduate participant to make only ten comparisons. Of the 34 participants we recruited, two failed to complete their set of judgements and were removed from the analysis. The mean duration per judgement for the undergraduates was 49.7 s, and no undergraduate was excluded due to a very low mean judgement duration (the lowest was 17.3 s). Therefore, a total of 320 judgements from 32 undergraduates were included in the final analysis. Of the 48 participants, 12 identified themselves as female (1 mathematician, 11 undergraduates), and 36 as male. A majority of mathematicians, 12, described their research area as falling primarily within the domain of pure mathematics.

Participants were invited to take part via an email from a member of the research team, which stated the following:

> We are interested in understanding how people assess the quality of mathematical explanations. We asked 10 mathematicians this question:
>
> Imagine that a math major on your linear algebra course comes to your office hours and says that they are confused. They explain that although they have seen the definition, they do not understand what an abstract vector space is, or what it is for. What explanation would you give the student in response?
>
> We are going to ask you to evaluate their responses so that we can understand what you value in an explanation.

If recipients of the email wished to participate, they clicked through to a website which explained the purpose of the study, took some demographic information (research area for the mathematicians, gender for all participants), and then asked participants to start judging. To record participants' judgements, we used the No More Marking comparative judgement platform (https://www.nomoremarking.com/). Participants were presented with two randomly chosen explanations side by side, and simply asked "which is the better explanation of what a vector space is?". They responded by selecting either the explanation on the left or the explanation on the right. Participants were instructed that, if they were unsure, they should go with their "gut instinct". When participants had completed their allotted judgements (20 for mathematicians, 10 for undergraduates), they exited the judging platform. (Raw data, materials, and analysis scripts are available online.[3])

The data analysis method involved comparative judgement approach, which was based on the Bradley-Terry model, described in Section 4. The method produced β estimates for each judged item, capturing the perceived explanatoriness of each explanation, separately for each group. These βs are unitless, so can only be interpreted in relation to other βs on the same scale. The Bradley-Terry model also produces a standard error associated with each β, which captures the precision with which the β has been estimated.

6 Results

Each group produced highly reliable β estimates. The split-half inter-rater reliability coefficient was $r = .781$ for the mathematicians and $r = .796$ for the undergraduates, indicating that the within-group agreement about the quality of the various explanations was high.

The βs for each explanation produced by each group are shown in Fig. 1. Overall, the correlation between the two groups was high, at $r = .85$, indicating that the two groups largely agreed with each other about which explanations were better and which were worse.

Inspecting Fig. 1 indicates that both groups agreed that Explanation 2—which attempted to explain vector spaces in terms of enriched Abelian groups—was the least explanatory. The two groups also both disliked Explanation 10, which attempted to use the geometry of the academic's office to introduce the notion of linearly independent vectors, and then generalised to abstract vector spaces.

[3] https://doi.org/10.17028/rd.lboro.14213831.v1

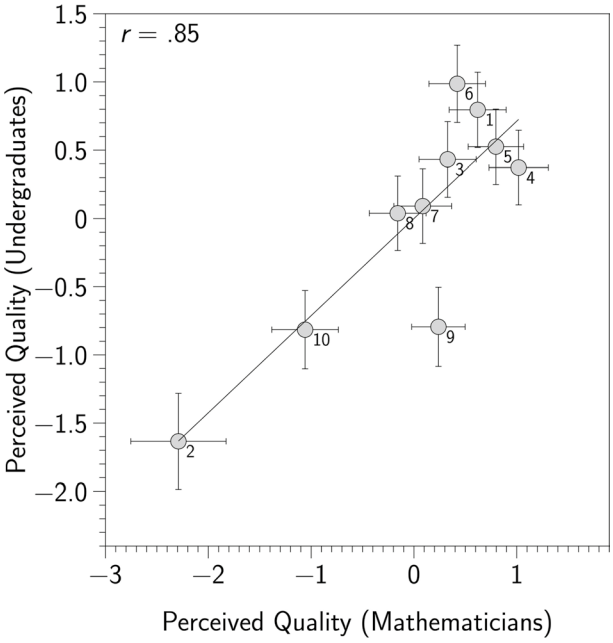


Fig. 1 The perceived quality of each explanation (the β estimates) produced by the mathematicians and undergraduates. *Note*. Error bars show ±1 standard error. The numbers next to each point indicate the explanation represented by that point (see the Appendix for the full list of explanations). Note that the units on the x and y scales are not comparable

We operationalised disagreement between the two groups as being where an explanation's β estimate was over two standard errors away from the regression line shown in Fig. 1. There were only two examples of disagreement. The mathematicians perceived Explanation 9 to be more explanatory than the undergraduates. This focused on using geometric analogies and then pointed out that vector spaces are useful to understand the properties of solutions to partial differential equations (PDEs) using geometric ideas, images, and pictures. While this utility consideration would make sense to a mathematician, most students taking a linear algebra course would not be familiar with the theory of PDEs. Finally, towards the top end of the scale, the undergraduates rated Explanation 6 as being the most explanatory, whereas this was only the fourth most explanatory for the mathematicians. Explanation 6 focused on connecting the notion of a vector space to the properties of mathematical objects that undergraduates could be reasonably expected to be very familiar with (addition and multiplication of numbers, combining functions, adding matrices, etc.). It went on to explain that the purpose of the notion of a vector space is to extract the commonality between these various concepts. Notably, it was the only explanation to start with an informal hand-wavy definition: "a vector space is just a collection of objects, together with a way to combine those objects, and a list of rules that govern how we combine them", thereby, perhaps, earning appreciation from undergraduates.

Despite these two disagreements, the overall message of our study was one of strong agreement. Using a comparative judgement technique, both the mathematicians and undergraduates were able to reliably assess the quality of these mathematical explanations, in the sense that the members of each group tended to agree with other members of the same group about the explanatoriness of each explanation. Furthermore, the two groups agreed with each other. The mathematicians were largely able to predict how the undergraduates would assess explanation quality, and vice versa.

7 Discussion

7.1 Summary of main findings

Our goal in this paper was to conduct a bottom-up investigation of the quality of mathematical explanations. To this end, we created a small corpus of mathematical explanations, and examined whether mathematicians and undergraduates were able to reliably assess their quality using a comparative judgement approach. We found that both groups showed reasonably high levels of reliability, in the sense that the split-half inter-rater reliability coefficients were well above 0.7 in both cases.

Given this, we asked whether the two groups agreed with each other: in other words, do mathematicians tend to assess explanatoriness in a similar fashion to undergraduates? We found that the two groups' assessments of the explanations in our corpus were strongly correlated, suggesting that—notwithstanding some differences—the two groups have a shared understanding of what makes a high- and a low-quality mathematical explanation.

7.2 Study limitations

Several limitations of our study may have influenced our results. The first concerns the methodological decisions made in the design of the explanations by the participating mathematicians. Some of the shorter explanations produced by the participating mathematicians (such as Explanations 5, 8, 9, and 10) present as outlines (plans) of an explanation, stating the intentions but omitting details. For example, Explanation 9 starts with "I would give geometric explanations and analogies; and show with pictures how the geometric explanations works for all vectors." This leaves room for various interpretations around the implementation of such an intent, thus translating into more subjective judgements. Given this, it is perhaps surprising that we found such high levels of between-participant agreement.

Second, no information was available about the domain-specific knowledge of the participating undergraduates, other than that they were mathematics majors who recently completed a linear algebra course. Potentially large differences in participants' levels of domain knowledge or mathematical experience could perhaps have influenced their judgements, which we might expect to suppress observed agreement levels. Future studies could productively investigate whether individual differences influence the type and nature of students' (and mathematicians') judgements of explanatoriness.

Moreover, Explanations 1–10 might not have been sensitive enough to capture differences in participants' perceptions as they were not systematically designed to test variations. Systematically varying aspects of explanations might be worthwhile in future investigations in order to test factors that characterise explanation quality in mathematics. Furthermore, future studies could benefit from increasing the relatively small sample size used in our investigation.

7.3 Validity

Although we found strong evidence of the reliability of mathematicians' and undergraduates' judgements about explanation quality, a question remains about whether those explanations which our participants perceived to be the most explanatory are actually the most explanatory. Answering this question requires us to—in a top-down fashion—independently specify what explanatoriness consists in.

As noted in Section 2, philosophers typically adopt either ontic or epistemic accounts of mathematical explanation. What would theorists from each of these camps think about the explanations in our corpus? For ontic theorists, this seems a difficult question. As D'Alessandro (2020) noted, ontic theorists have typically assumed that explanations in mathematics must involve proofs, that the only things which can be explanatory in mathematics are explanatory proofs. D'Alessandro (2020) critiqued this view, which he labelled 'proof chauvinism'. Clearly, the explanations in our corpus are not proofs, so it is difficult to know what Steiner (1978) or Kitcher (1984) would make of them.[4] One approach would be for ontic theorists to deny that our explanations are in fact explanations, and perhaps instead to consider them to be mere motivations of the vector space definition. While this approach would have the advantage of being consistent with their conceptualisation of explanation, it has the significant disadvantage of being inconsistent with mathematical practice. The mathematicians who generated the explanations in our corpus were asked "what explanation would you give the student?" They were not asked how they would "motivate the definition." If we wish our conceptions of mathematical explanation to be consistent with mathematical practice, they ought to be applicable to the kinds of things mathematicians produce when asked to explain.

Epistemic theorists have an easier time when considering the validity of our participants' explanatory judgements. If explanations are defined to be those things which generate understanding, then better explanations are going to be those explanations which generate more understanding. Wittwer and Renkl's (2008) framework provides a top-down method for us to consider the extent to which each of the explanations in our corpus is likely to generate understanding for a typical undergraduate, and therefore allows us to interrogate the validity of our participants' judgements.

We consider each of Wittwer and Renkl's (2008) criteria in turn. Recall that each explanation is given in the Appendix.

1. *Explanations should take account of learners' existing knowledge.* It seems clear that the lowest scoring explanation, Explanation 2, falls foul of this criterion. Explanation 2 references Abelian groups, fields, enrichments, rings, and algebras over a ring. Given the context specified in the prompt—an undergraduate taking an introductory linear algebra course—it seems very unlikely that the explanation's recipient would be familiar with many of these more advanced mathematical concepts. In a similar vein, Explanation 9, referring to the theory of PDEs, does not score highly in the ranking.
2. *Explanations should focus on concepts and principles.* A vector space is an abstract notion that captures a range of mathematical objects that all behave in the same way, in the sense that if a property follows from the vector space axioms, then all vector spaces will have that property. This extraction of common mathematical structure seems a central concept/principle which accounts for why mathematicians are interested in vector spaces. Notably, this idea is entirely absent from Explanation 2, is not clearly included in Explanation 10, and is not clearly expressed in Explanations 8 and 9. These four explanations were rated particularly poorly by our participants. In contrast, all the other explanations reference this key concept/principle in one form or another. Explanations 1, 4, 5, and 6—all high scoring for both groups—are extremely explicit about this. For

[4] As D'Alessandro (2020) noted, Lange's (2014) view acknowledged that theorems as well as proofs could be explanatory. Nevertheless, D'Alessandro (2020) still considered Lange (2014) to be a proof chauvinist, on account of his view that theorems are explanatory by virtue of them having a certain kind of proof.

example, Explanation 5 included the line "by proving (or knowing) something about a vector space we know it about all of those examples [discussed earlier in the explanation]."

3. *Explanations should be integrated into the learner's ongoing cognitive activities.* Good explanations encourage active engagement with the recipient. Interestingly, Explanations 1, 4, and 5—three of the high-scoring explanations for both groups—all included some kind of interactive element. For example, Explanation 4 involved rhetorical questions ("Now you might ask in what ways are [real polynomials and $\mathbb{R}^n$] similar?"), and Explanation 1 involved multiple possibilities for the student receiving the explanation to contribute ("Is W a vector space? Let's check: closed under addition—yes."). Explanation 5 emphasised how the explanation would be contingent on how the student responded ("I would [⋯] point out the common properties. Once the student is happy with those, we would say that "something" with all of those properties is called a vector space."). In contrast, the lowest scoring explanations did not seem to require the student to interact with the explanation or the explainer at all.
4. *Explanations should not replace learners' knowledge-construction activities.* The final criterion proposed by Wittwer and Renkl (2008) concerned when it would be preferable for instructors to withhold explanations from students, and so does not directly apply to the context specified in our vignette.

In sum, these considerations suggest that the judgements made by the mathematicians and undergraduates in our study were not inconsistent with Wittwer and Renkl's (2008) framework. Indeed, in some important ways, the higher-scoring explanations seemed to meet the criteria, and the lower-scoring explanations seemed not to. However, these discussions represent, at best, highly indirect evidence of validity. Future studies which directly compare explanatory judgements of different explanations with the extent to which those explanations generate student understanding would be extremely worthwhile. If comparative judgement could be established as both a valid and reliable way of assessing the quality of instructional explanations in mathematics, the method could be harnessed to help improve both classroom instruction and instructional materials such as textbooks and lecture notes.

7.4 Developing explanatory skills

Supposing that mathematicians are able to reliably and validly judge the quality of mathematical explanations, this raises a puzzling question. Why do some mathematicians produce low-quality explanations when prompted? Consider Explanation 2, the lowest-ranked explanation. Of the 62 comparisons made by mathematicians involving Explanation 2, it 'lost' 92% (in the sense that the explanation it was paired with was deemed the better in 92% of pairings). Given this level of consensus about its low quality, why did the mathematician who wrote it consider it to be appropriate?

One answer might be to suggest that producing high-quality explanations may be considerably harder than assessing the quality of explanations. Analogously, it is possible for critics to assess the quality of novels, even though they may not be able to produce a novel themselves. If this is right, then it raises the question of how teachers and lecturers can be helped to produce better explanations.

We believe that the type of comparative judgement session that we used here for research purposes could serve as a starting point for productive form of professional development for teachers and lecturers. Comparison has been shown to be an effective pedagogical strategy in

other contexts. Rittle-Johnson et al. (2020) recently reviewed the evidence concerning how comparing different problem-solving strategies for the same mathematical problem can support students' learning (see also Alfieri et al.'s (2013) meta-analysis). It is typically suggested that promotion of analogical reasoning is the mechanism behind this effect: by comparing two examples, students are able to create an analogy between them, which helps them to attend to structural similarities and differences. This, in turn, may facilitate transfer to new situations (Gentner et al., 2003; Gick & Holyoak, 1983). If this is correct, then this mechanism would seem to be as applicable in the context of teachers and lecturers comparing different instructional explanations (along with explicating and reflecting on their differences) as it is in the context of students comparing different problem-solving strategies.

If comparative judgement were to be harnessed to develop professional development materials for teachers and lecturers who want to develop their ability to produce instructional explanations, then there is a large literature on which to draw. For instance, Rittle-Johnson et al. (2020) pointed out that including various instructional supports, such as presenting examples side by side (as in our study), including cues to guide attention to important similarities and differences, and including self-explanation prompts at relevant points, all facilitate learning from comparison.

7.5 Implications for the philosophical accounts of mathematical explanation

Finally, we briefly comment on the implications of our results for philosophers interested in mathematical explanation. We first re-emphasise that all the mathematicians in these studies were either willing to produce non-proof explanations, or to compare non-proof explanations. It therefore seems clear that any account of explanation in mathematics—or at least any account that wishes to remain faithful to the practices of mathematicians—must, as argued by D'Alessandro (2020), be able to account for explanations which are not proofs. Our view is that epistemic accounts such as Delarivière et al.'s (2017) and Inglis and Mejía-Ramos's (2021) can do this with relative ease. In contrast, the challenge for ontic accounts seems much greater.

More generally, the evidence presented here suggests that mathematicians and undergraduates have a shared conception of what constitutes an effective explanation in mathematics, at least in the context of these rather simple explanations. Hence, this finding serves as foundation for designing and undertaking further investigations. Future work should explore whether this remains the case when different types of explanations, including proofs, are considered. If it does, then we see the goal of philosophical accounts of mathematical explanation as being to produce an accurate description of what this shared conception actually is. Perhaps, this will necessitate a consideration of distinct taxonomic groups (such as definitions, theorems, proofs, problem-solving procedures) within which the classification of mathematical explanations with respect to their quality can be achieved. Having a method available—comparative judgement—which seems to be able to reliably measure mathematicians' explanatory judgements will allow existing and future accounts to be tested empirically.

Appendix: Corpus of explanations

Explanation 1

I would say that in its abstract form a vector space is a concept that generalises physical entities we know as lines ($\mathbb{R}^1$), planes ($\mathbb{R}^2$), and 3-d space ($\mathbb{R}^3$). Considering these as sets of vectors, with the usual addition and multiplication by a real constant, we can confirm that they satisfy all ten properties of a vector space. There are a variety of spaces which are not geometrically recognisable as vector spaces, but which have an analogous structure. For example: Consider V—the set of all 2×2 matrices with real entries, with the standard addition and scalar multiplication.

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix} \in V$$

Hence, V is closed under addition (satisfies the first property of a vector space). Also, V is closed under scalar multiplication (check), and all other eight properties. Hence, V is a vector space. Now, consider W—the set of all 2×2 matrices with integer entries, with the standard addition and scalar multiplication. Is W a vector space? Let's check: closed under addition—yes. There is a zero vector:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

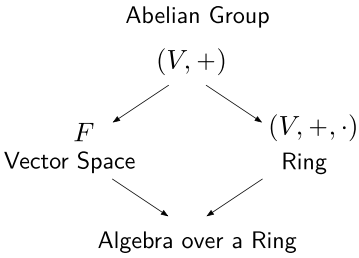
Is it closed under scalar multiplication? No, it is not. Here's a counterexample:

$$\frac{1}{2}\begin{bmatrix} 2 & -4 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 3 & \frac{1}{2} \end{bmatrix} \notin W.$$

Hence, W is not a vector space. Working with a vector space ensures that you have a 'nice' structure and allows for efficient ways of reaching conclusions. For example, you can conceptualise the set of solutions of a homogeneous linear ordinary differential equation as a vector space of functions.

Explanation 2

I'd say that a vector space is first and foremost an Abelian group. It also has the action of a field. A vector space is an enrichment of an Abelian group. There are two ways to enrich an Abelian group $(V, +)$: one way is by making it a ring $(V, +, \cdot)$, another way is by making it a vector space over a field F. Both of those enrichments can be enriched even further to become an algebra over a ring.



Explanation 3

I would like to avail myself to very simple but illuminating examples, and highlight their similarities, despite the difference in mathematical context:

1. The real line, with the usual addition, and multiplication by a real constant.
2. The 2-dimensional Cartesian plane, with the usual vector addition, and multiplication by a real number.
3. Increase the dimension n, and talk about $\mathbb{R}^n$.
4. Let A be a given set. Consider the power set of A and the field $\mathbb{Z}_2 := \{0, 1\}$, with vector addition as disjoint union, and scalar multiplication defined by the two equations (1) $1 \cdot X := X$ and (2) $0 \cdot X := \emptyset$ for all subsets X of A. Then $(\mathcal{P}(A), \dot{\cup}, \cdot)$ forms a vector space.

The similarities to be highlighted include the (carrier) set V which is the vector space, the vector addition, and the scalar multiplication.

Then comes the set of axioms to be satisfied.

Explanation 4

I would say "so far, most of the math you've learned is about $\mathbb{R}^n$. As you probably understand, the axioms that define an abstract vector space apply to $\mathbb{R}^n$. Now we want to understand what other mathematical objects satisfy those same axioms. For instance, the collection of all polynomials with real coefficients and degree less than or equal to n also satisfies those same axioms, and therefore there are ways in which studying these polynomials is the same as studying $\mathbb{R}^n$.

Now you might ask in what ways are they similar? Well, that's also the goal of studying abstract vector spaces—what are the properties of $\mathbb{R}^n$ that rely only on the axiomatic properties of being a vector space? For instance, one such property is having a collection of vectors (basis vectors) such that any element of $\mathbb{R}^n$ can be written as a linear combination of the vectors in this collection. Since that property of $\mathbb{R}^n$ only relies on the axiomatic properties of being a vector space, any other abstract vector space (such as the collection of all polynomials with real coefficients and degree less than or equal to n) must have that same property, and in that way is similar to $\mathbb{R}^n$."

Explanation 5

I would give some examples, $\mathbb{R}^2$ and $\mathbb{R}^3$, then the space of polynomials of degree less than or equal to n, and point out the common properties. Once the student is happy with those, we would say that "something" with all of those properties is called a vector space. By proving (or knowing) something about a vector space, we know it about all of those examples. The use of analogy and examples in mathematics is really important.

Explanation 6

I would say that a vector space is just a collection of objects, together with a way to combine those objects, and a list of rules that govern how we combine them. We see examples of this from the moment we first learn about addition and multiplication of numbers. We learn how to add and multiply numbers, and then we learn that there are certain rules that addition and multiplication of numbers satisfy. Then we eventually learn about functions and that there are ways to combine those functions. We can add them, multiply them, or compose them. We can also multiply them by numbers. Then we learn that there are certain rules that these operations satisfy. Some of these are the same rules that we saw with addition and multiplication of numbers. We have also leaned about vectors and matrices, and that there are similar operations that we can do with these, such as addition and various forms of multiplication, including multiplying by numbers. Then we learned that these operations satisfy many of the same rules that we observed with operations on numbers and functions.

The purpose of the abstract notion of a vector space is to extract all of these common concepts, objects with operations that satisfy a certain set of rules, and then we use these rules as a foundation for proving other things, things that must be true for all objects with operations that satisfy this list of rules. Then we explore other things in mathematics and in the world around us that share these same features, objects with operations that satisfy the same list of rules. When we find something that matches these features, we know that they automatically must satisfy all of the extra properties and theorems that we proved about abstract vector spaces, without the need to prove them again in this specific instance.

Explanation 7

I would say that a vector in 2-dimensional space, such as (1, 2), records a movement from a starting point to an ending point. We have a special point, which we call the origin, with coordinates (0, 0). The vector (1, 2) is like an arrow starting at the origin, and going to its end point, which is one unit right horizontally, and 2 units up vertically from the origin. Vectors are often used to describe the force applied to an object, such as the force being applied by gravity.

It is useful to add vectors together. Take $(3, -5)$ and $(-1, 2)$. Each of these corresponds to an arrow starting at the origin. To find the sum $(3, -4) + (-1, 2)$, we slide the $(-1, 2)$ arrow so that its starting point is now at the ending point of the $(3, -4)$ arrow (which changing its direction or length). We now look at the new ending point of the $(-1, 2)$ arrow, and we see that it is equal to $(2, -2)$, which is $(3 + (-1), (-4) + 2)$. One reason for adding vectors in this way is that it shows how forces compound on an object.

So if gravity is applying a vector of $(3, -5)$ to an object, and an engine is applying a vector of $(-1, 2)$ to the same object, the combined force that the object feels is $(-2, 2)$.

We can also multiply vectors by numbers, so 2 times $(-1, 2)$ is just $(2 \times (-1), 2 \times 2) = (-2, 4)$. This corresponds to the force becoming twice as large.

When we work in an abstract vector space we are still thinking about vectors, addition, and multiplication. But now our vectors do not have explicit coordinates. So instead of talking about vectors with coordinates $(-1, 2)$ and (3, 4), we might just have abstract vectors, which we label with something like u and v. Instead of adding and multiplying coordinates, we just have rules that tell us what the abstract vector $u + v$ is.

The reason we care about abstract vector spaces is that by studying them, we can study all concrete vector spaces simultaneously, and learn how they behave.

Explanation 8

Here's what I would do. Firstly start with the concrete case of two dimensions. Then (using coordinates) talk about length and direction of vectors based at 0 and the properties of scaling a vector. Then explain these notions are independent of the origin. Then look at three dimensions and then begin to abstract these notions as being dimension independent. Now to addition. This is like reading a map: go this far in this direction (length and direction) and then this far in this other direction.

I would then show how this links up with coordinates in 2D and 3D before abstracting away from coordinates. Always remember that vectors add like friendly dogs. Nose to tail!

Explanation 9

I would give geometric explanations and analogies; and show with pictures how the geometric explanations works for all vectors.

Then I would say that the idea of a vector space makes use of the properties common for all vectors, it allows the use of geometric ideas, images or pictures for such quantities as solutions to a (linear) differential equation; the notion of an abstract vector space makes much easier to understand, for instance, the properties of solutions to a PDE (and most explanations of nature are made with PDEs...)

Explanation 10

I would say that it is the generalization of our physical three-dimensional space. I would point at a spot in the room where two walls and the ceiling meet and explain that any vector in our space is a linear combination of these three (linearly) independent vectors. An abstract vector space would be a proper notion to identify a structure similar to our space (in arbitrary dimensions).

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

This work was supported by the University of Auckland Faculty of Science Research and Development Grant (Project code: 3720159).

Data availability Raw data and materials supporting the findings of this study are publicly available at https://doi.org/10.17028/rd.lboro.14213831.v1.

Code availability The code used for analysis in this study is publicly available at https://doi.org/10.17028/rd.lboro.14213831.v1.

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval An approval was granted by the University of Auckland Human Participants Ethics Committee (Ref 3093) and by the Loughborough University Ethics Approvals (Human Participants) Sub-Committee.

References

Alcock, L., Hodds, M., Roy, S., & Inglis, M. (2015). Investigating and improving undergraduate proof comprehension. *Notices of the American Mathematical Society, 62*(7), 742–752. https://doi.org/10.1090/noti1263

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

Bell, A. W. (1979). The learning of process aspects of mathematics. *Educational Studies in Mathematics, 10*(3), 361–387. https://doi.org/10.1007/BF00314662

Bills, L., & Watson, A. (2008). Editorial introduction. *Educational Studies in Mathematics, 69*(2), 77–79. https://doi.org/10.1007/s10649-008-9147-z

Bisson, M.-J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 141–164. https://doi.org/10.1007/s40753-016-0024-3

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika, 39*, 324–345. https://doi.org/10.1093/biomet/39.3-4.324

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

Colyvan, M. (2012). *An introduction to the philosophy of mathematics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139033107

D'Alessandro, W. (2020). Mathematical explanation beyond explanatory proof. *British Journal for the Philosophy of Science, 71*, 581–603. https://doi.org/10.1093/bjps/axy009

Davies, B., Alcock, L., & Jones, I. (2021). What do mathematicians mean by proof? A comparative-judgement study of students' and mathematicians' views. *The Journal of Mathematical Behavior, 61*, 100824. https://doi.org/10.1016/j.jmathb.2020.100824

Delarivière, S., Frans, J., & Van Kerkhove, B. (2017). Mathematical explanation: A contextual approach. *Journal of Indian Council of Philosophical Research, 34,* 309–329. https://doi.org/10.1007/s40961-016-0086-2

Duffy, G. G., Roehler, L. R., Meloth, M. S., & Vavrus, L. G. (1986). Conceptualizing instructional explanation. *Teaching and Teacher Education, 2*(3), 197–214. https://doi.org/10.1016/S0742-051X(86)80002-6

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity*. Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Fukawa-Connelly, T.P., & Newton, C. (2014). Analyzing the teaching of advanced mathematics courses via the enacted example space. *Educational Studies in Mathematics, 87*(3), 323–349. https://doi.org/10.1007/s10649-014-9554-2

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393. https://doi.org/10.1037/0022-0663.95.2.393.

Gick, M.L., & Holyoak, K.J. (1983). *Schema induction and analogical transfer*. Academic Press. https://doi.org/10.1016/0010-0285(83)90002-6

Goldenberg, P., & Mason, J. (2008). Shedding light on and with example spaces. *Educational Studies in Mathematics, 69*(2), 183–194. https://doi.org/10.1007/s10649-008-9143-3

Grimm, S. R. (2006). Is understanding a species of knowledge? *British Journal for the Philosophy of Science, 57*(3), 515–535. https://doi.org/10.1093/bjps/axl015

Grimm, S.R., Baumberger, C., & Ammon, S. (Eds.). (2016). Explaining understanding: New perspectives from epistemology and philosophy of science. Taylor & Francis. https://doi.org/10.4324/9781315686110

Hamami, Y., & Morris, R. (2020). Philosophy of mathematical practice: A primer for mathematics educators. *ZDM-Mathematics Education*, *52,* 1113–1126. https://doi.org/10.1007/s11858-020-01159-5

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21*, 6–13. https://doi.org/10.1007/BF01809605

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics, 44*(1–2), 5–23. https://doi.org/10.1023/A:1012737223465

Heldsinger, S.A., & Humphry, S.M. (2013). Using calibrated exemplars in the teacher-assessment of writing: An empirical study. *Educational Research, 55*, 219–235. https://doi.org/10.1080/00131881.2013.825159

Hempel, C.G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175. https://doi.org/10.1086/286983

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics, 24*(4), 389–399. https://doi.org/10.1007/BF01273372

Inglis, M., & Mejía-Ramos, J.P. (2021). Functional explanation in mathematics. *Synthese, in press, 1–24,*. https://doi.org/10.1007/s11229-019-02234-5

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

Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics, 89*, 337–355. https://doi.org/10.1007/s10649-015-9607-1

Jones, I., Swan, M., & Pollitt, A. (2015). 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

Kelp, C. (2016). Towards a knowledge-based account of understanding. S.R. Grimm, C. Baumberger, & S. Ammon (Eds.), Explaining under-standing: New perspectives from epistemology and philosophy of science (pp. 251-271). Routledge

Kitcher, P. (1984). *The nature of mathematical knowledge*. Oxford University Press. https://doi.org/10.1093/0195035410.001.0001

Kvanvig, J. L. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9780511498909

Lachner, A., & Nückles, M. (2016). Tell me why! Content knowledge predicts process-orientation of math researchers' and math teachers' explanations. *Instructional Science, 44*(3), 221–242. https://doi.org/10.1007/s11251-015-9365-6

Lachner, A., Weinhuber, M., & Nückles, M. (2019). To teach or not to teach the conceptual structure of mathematics? Teachers undervalue the potential of principle-oriented explanations. *Contemporary Educational Psychology, 58*, 175–185. https://doi.org/10.1016/j.cedpsych.2019.03.008

Lange, M. (2014). Aspects of mathematical explanation: Symmetry, unity, and salience. *Philosophical Review, 123*(4), 485–531. https://doi.org/10.1215/00318108-2749730

Leinhardt, G., & Steele, M.D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction, 23*(1), 87–163. https://doi.org/10.1207/s1532690xci2301_4

Lew, K., Fukawa-Connelly, T. P., Mejìa-Ramos, J. P., & Weber, K. (2016). Lectures in advanced mathematics: Why students might not understand what the mathematics professor is trying to convey. *Journal for Research in Mathematics Education, 47*(2), 162–198. https://doi.org/10.5951/jresematheduc.47.2.0162

Mancosu, P. (2001). Mathematical explanation: Problems and prospects. *Topoi, 20*(1), 97–117. https://doi.org/10.1023/A:1010621314372

McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice, 22*, 368–389. https://doi.org/10.1080/0969594X.2014.978839

Mejía-Ramos, J.P., Alcock, L., Lew, K., Rago, P., Sangwin, C., Inglis, M. (2019). Using corpus linguistics to investigate mathematical explanation. E. Fischer (Ed.), Methodological advances in experimental philosophy (pp. 239-264). Bloomsbury. https://doi.org/10.5040/9781350069022.ch-009

Mejía-Ramos, J.P., & Weber, K. (2019). Mathematics majors' diagram usage when writing proofs in calculus. *Journal for Research in Mathematics Education, 50*(5), 478–488. https://doi.org/10.5951/jresematheduc.50.5.0478

Perry, M. (2000). Explanations of mathematical concepts in japanese, chinese, and u.s. first-and fifth-grade classrooms. *Cognition and Instruction, 18*(2), 181–207. https://doi.org/10.1207/S1532690XCI1802n_02

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*, 281–300. https://doi.org/10.1080/0969594X.2012.665354

Reid, D. (1995). Proving to explain. L. Meira & D. Carraher (Eds.), Proceedings of the nineteenth annual conference of the international group for the psychology of mathematics education (Vol. 3, p. 137-143). Recife, Brazil.

Resnik, M.D., & Kushner, D. (1987). Explanation, independence and realism in mathematics. *British Journal for the Philosophy of Science, 38*(2), 141–158. https://doi.org/10.1093/bjps/38.2.141

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2020). How can cognitive-science research help improve education? The case of comparing multiple strategies to improve mathematics learning and teaching. *Current Directions in Psychological Science, 29*, 599–609. https://doi.org/10.1177/0963721420969365

Roy, S., Inglis, M., & Alcock, L. (2017). Multimedia resources designed to support learning from written proofs: An eye-movement study. *Educational Studies in Mathematics, 96*(2), 249–266. https://doi.org/10.1007/s10649-017-9754-7

Salmon, W. C. (1971). *Statistical explanation and statistical relevance*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt6wrd9p

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

Samkoff, A., Lai, Y., & Weber, K. (2012). On the different ways that mathematicians use diagrams in proof construction. *Research in Mathematics Education, 14*(1), 49–67. https://doi.org/10.1080/14794802.2012.657438

Soto-Johnson, H., & Troup, J. (2014). Reasoning on the complex plane via inscriptions and gesture. *The Journal of Mathematical Behavior, 36*, 109–125. https://doi.org/10.1016/j.jmathb.2014.09.004

Steiner, M. (1978). Mathematical explanation. *Philosophical Studies, 34*(2), 135–151. https://doi.org/10.1007/BF00354494

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics, 12*(2), 151–169. https://doi.org/10.1007/BF00305619

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286. https://doi.org/10.1037/h0070288

Thurstone, L. L. (1927). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology, 21*, 384–400. https://doi.org/10.1037/h0065439

Treagust, D., & Harrison, A. (1999). The genesis of effective scientific explanations for the classroom. J. Loughran (Ed.), Researching teaching: Methodologies and practices for understanding pedagogy (pp. 28-43). Falmer Press.

Van Bendegem, J.P. (2014). The impact of the philosophy of mathematical practice on the philosophy of mathematics. L. Soler, S. Zwart, M. Lynch, & V. Israel-Jost (Eds.), *Science after the practice turn in the philosophy, history, and social studies of science* (pp. 215-226). Routledge.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209–249. https://doi.org/10.1207/S1532690XCI2103_01

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406. https://doi.org/10.1037/0022-0663.87.3.406

Weber, E., & Frans, J. (2017). Is mathematics a domain for philosophers of explanation? *Journal for General Philosophy of Science, 48*(1), 125–142. https://doi.org/10.1007/s10838-016-9332-1

Weber, K. (2010). Proofs that develop insight. *For the Learning of Mathematics, 30*(1), 32–37.

Wilkenfeld, D. A. (2014). Functional explaining: A new approach to the philosophy of explanation. *Synthese, 191*(14), 3367–3391. https://doi.org/10.1007/s11229-014-0452-z

Wittwer, J., Nückles, M., & Renkl, A. (2008). Is underestimation less detrimental than overestimation? The impact of experts' beliefs about a layperson's knowledge on learning and question asking. *Instructional Science, 36*(1), 27–52. https://doi.org/10.1007/s11251-007-9021-x

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. https://doi.org/10.1080/00461520701756420

Zelcer, M. (2013). Against mathematical explanation. *Journal for General Philosophy of Science, 44*(1), 173–192. https://doi.org/10.1007/s10838-013-9216-6