

Five Decades of Mathematics Education Research

Matthew Inglis
Loughborough University

Colin Foster
University of Leicester

Mathematics educators have been publishing their work in international research journals for nearly 5 decades. How has the field developed over this period? We analyzed the full text of all articles published in *Educational Studies in Mathematics* and the *Journal for Research in Mathematics Education* since their foundation. Using Lakatos's (1978) notion of a research programme, we focus on the field's changing theoretical orientations and pay particular attention to the relative prominence of the experimental psychology, constructivist, and sociocultural programmes. We quantitatively assess the extent of the "social turn," observe that the field is currently experiencing a period of theoretical diversity, and identify and discuss the "experimental cliff," a period during which experimental investigations migrated away from mathematics education journals.

Keywords: History of mathematics education; Methodology of scientific research programmes; Topic modeling

Research on mathematical thinking and learning has a long history, but it is reasonable to suggest that the field of mathematics education research started a new phase in 1968. It was in this year that *Educational Studies in Mathematics* (*ESM*) published its first issue, and 2 years later, in 1970, the *Journal for Research in Mathematics Education* (*JRME*) followed suit (Kilpatrick, 1992). Accepting 1968 as a starting date for the modern research field implies that mathematics educators have been publishing research in international journals for nearly half a century. How has the field changed during this period? Our goal in this article is to answer this question by reporting on a study in which we analyzed the full text of all articles published by *ESM* and *JRME* since their founding. This approach allowed us to identify the main theories, methods, and domains that mathematics education researchers have focused on over the last 5 decades and how their relative prominence has changed during this period.

Unsurprisingly, there have been several earlier characterizations of the development of mathematics education as a research field. For example, Hanna and Sidoli

This research was partially supported by a Royal Society Worshipful Company of Actuaries Research Fellowship. We are grateful to Dor Abrahamson, Lara Alcock, Daniel Battey, Barbara Jaworski, Brendan Larvor, Steve Lerman, John Mason, Juan Pablo Mejía-Ramos, Dirk Schlimm, Anne Watson, and Keith Weber for providing useful feedback on earlier drafts of this work.

(2002) celebrated the 50th volume of *ESM* by conducting a statistical analysis of the keywords assigned to *ESM* articles in the ERIC bibliometric database. They found that the research community's interest in geometry and space had declined since the 1970s and that its interest in problem solving peaked during the 1980s. Using internal *ESM* data used to assign reviewers to papers, Hanna and Sidoli also established that there had been a growth in the number of papers that focused on social issues in the teaching and learning of mathematics. This finding echoed Lerman, Xu, and Tsatsaroni's (2002) more detailed analysis of papers published during the 1990s. Lerman et al. analyzed a sample of papers published in *ESM* from 1990 to 2001 and found a substantial increase in the proportion that used social theories. In the first half of the decade, social approaches were present in 9% of papers compared with 34% in the second half. Tsatsaroni, Lerman, and Xu (2003) later extended this analysis to papers published in *JRME* and the *Proceedings of the International Group for the Psychology of Mathematics Education (PME)*, finding similar results. It was this data set that led Lerman (2000) to claim that during the 1990s, there had been a "social turn in mathematics education research." He argued that the decade had seen a shift in focus from the cognitive to the social: from theories that focus on individuals' thought processes to "theories that see meaning, thinking, and reasoning as products of social activity" (p. 23).

Lerman's (2000) description of the social turn proved extremely influential. Wagner (2015) argued that by giving the social turn a name, Lerman described the phenomenon but also helped shape it. One sign of Lerman's influence is that other researchers have adopted his language to advocate for other changes in the discipline. For instance, Gutiérrez (2013) suggested that mathematics education should make a "sociopolitical turn" and include a greater focus on issues of social justice and equity. However, unlike Lerman and the social turn, Gutiérrez was not claiming that mathematics education research had already gone through a sociopolitical turn; rather, she adopted Lerman's language to argue that it should.

Despite the significance of Lerman's characterization of the social turn, some have questioned whether it continued through the 2000s. For example, based on their personal impressions of presentations at PME conferences, Gates and Jorgensen (2015) maintained that the social turn is absent from much of the mathematics education research literature. Similarly, Jablonka and Bergsten (2010) hypothesized that the trend Lerman identified may not have continued beyond the period he studied. Quantitatively assessing the extent of the social turn is one goal of the study reported in this article.

Alongside Lerman et al.'s (2002) and Hanna and Sidoli's (2002) contributions, there have been numerous other attempts to empirically map shifts in mathematics education research (e.g., Sierpinska & Kilpatrick, 1998). Largely, these have focused on particular topics such as gender, class, and race (Chassapis, 2002; Lubienski & Bowen, 2000); particular theories (e.g., Pais & Valero, 2012); or particular geographical regions (e.g., Boero & Szendrei, 1998; Lai & Loo, 1992; Schoenfeld, 2016). These papers have typically analyzed database keywords (e.g., Chassapis, 2002; Lubienski & Bowen, 2000), reported interviews with leading researchers (e.g., Kieran, 1994),

given personal introspective accounts of their impressions of changes in the field (e.g., J. W. Wilson, 1994), or offered historical accounts (e.g., Kilpatrick, 1992, 2014).

Here, we offer a broader analysis of changes in mathematics education research over the past 5 decades. Rather than focusing on particular topics, time periods, or geographical areas, we analyzed the full text of all articles published in *ESM* and *JRME* since their inception 5 decades ago. Using this inclusive approach, we were able to identify the key topics that mathematics education researchers have focused on since 1968 and track how these topics have changed over time. Before presenting our method and results, we first situate our work using Lakatos's (1978) methodology of scientific research programmes.¹

Theory Change: The Methodology of Scientific Research Programmes

How do academic disciplines make progress? Philosophers of science have proposed a variety of accounts, including Popper's (1959) theory of progress through falsification and Kuhn's (1962) suggestion that progress occurs through periods of normal science being punctuated by rapid paradigm shifts. Here we situate our work theoretically using Lakatos's (1978) notion of a scientific research programme, which can be seen as a modification of Kuhn's account (Larvor, 1998). Whereas Kuhn felt that different research approaches—paradigms, in his language—were incommensurate, Lakatos argued that this view implied that there are no rational methods by which one can choose between different paradigms and, therefore, that scientific progress was “*a matter for mob psychology*” (p. 91). He saw his methodology of scientific research programmes as being a means by which to maintain Popper's belief that science is a rational process while retaining Kuhn's much greater fidelity to history (Larvor, 1998).

The main idea in Lakatos's (1978) account is that the base descriptive unit of research is not, as Popper (1959) argued, an individual research hypothesis or even an individual theory but rather a research programme. Such a programme is, in Lakatos's sense, a historically connected series of theories that all share the same “hard core”: a collection of key assumptions and beliefs accepted by those who work within the programme. For instance, the hard core of the Newtonian research programme included the notion of gravitational action at a distance together with Newton's laws of motion. The hard core is the programme's defining characteristic: It must be defended against falsification because if the hard core were to be modified, the programme itself would have been abandoned.

Research programmes often encounter difficulties in the form of anomalous empirical observations. Indeed, Lakatos (1978) said that programmes “grow in a permanent ocean of anomalies” (p. 6). He proposed that they deal with these in one of two ways. Often, anomalies are simply ignored: If the programme is successfully achieving its goals, researchers may simply treat anomalies as open questions to be dealt with later. Alternatively, the programme may use what Lakatos called a “protective belt.” This consists of a large collection of auxiliary

¹ As is traditional when discussing Lakatos's work, we use the British spelling of *programme*.

hypotheses that supplement the hard core and that can be used to prevent it from being falsified. The protective belt, unlike the hard core, can be modified or abandoned without doing serious damage to the programme. When some empirical anomaly—a potential challenge to the programme’s hard core—is observed, a modification is made to the protective belt. This allows the hard core to survive intact and the programme to continue. To illustrate this idea, Lakatos gave the example of a scientist observing a planet moving in a fashion inconsistent with Newton’s laws. Rather than abandoning the hard core of the Newtonian research programme, the scientist would examine hypotheses from the protective belt, perhaps by changing his or her assumptions about atmospheric refraction or even by proposing an as-yet-unobserved planet (Linton, 2004). What the scientist certainly would not do is abandon his or her beliefs in the hard core.

The third component of a research programme is its “heuristic,” the collection of methods and problem-solving techniques that researchers within the programme use to make progress. For instance, the Newtonian research programme’s heuristic involved modeling empirical observations and making predictions using a set of sophisticated mathematical techniques. The heuristic is tied to the programme, and it is not always straightforward to separate a programme’s hard core from its heuristic. Indeed, Lakatos (1978) suggested that this distinction could in some cases merely be “a matter of convention” (p. 181). For example, the measurement of response times is an important part of the heuristic of the cognitive psychology research programme, but this is because of assumptions from the programme’s hard core (the temporal nature of information processing).

Using the notions of the hard core, protective belt, and heuristic, Lakatos (1978) attempted to explain scientific progress and theory change by considering a discipline as a collection of competing programmes. He distinguished between two types of research programme. “Progressing” programmes are those that regularly generate surprising new results and research directions. Such programmes may be so successful that they can legitimately ignore anomalies, or they may deal with anomalies by modifying their protective belts in such a way that their heuristics are able to use the modifications to productively generate more new results. In contrast, a “degenerating” programme rarely makes novel discoveries or predictions and dedicates its protective belt to the post hoc accommodation of anomalous observations. Lakatos suggested that research programmes are abandoned when researchers give up trying to accommodate anomalies into a degenerating research programme and instead join a rival programme that is progressing.

In our discussion, we talk, as did Lakatos, about “science” and “scientific” research programmes. However, it is clear that Lakatos intended his ideas to apply to disciplines beyond the hard sciences. Indeed, he used his account to analyze the weaknesses of both Marxism and Freudianism (Lakatos, 1978), and others have applied his ideas to educational and psychological research (e.g., Dienes, 2008; Gilbert & Swift, 1985; Inglis, 2015; Taber, 2007).

Research programmes can be considered at different levels. Indeed, Lakatos (1978) pointed out that “even science as a whole can be regarded as a huge research

programme with Popper's supreme heuristic rule: 'devise conjectures which have more empirical content than their predecessors'" (Lakatos, 1978, p. 47). However, more commonly, an academic discipline is seen as consisting of several rival research programmes that compete for researchers' attention by attempting to demonstrate that they are progressing (Gillies, 2007; Larvor, 1998). It is this latter use to which we put Lakatos's notion in this article. Although Lakatos emphasized the role of competition between theories, he intended this to be a constructive form of competition in which each programme is spurred to progress as a result of challenges from rival programmes. Lakatos even suggested that individual researchers could work within more than one research programme in order to expedite this process (Lakatos, 1978, p. 112).

Lakatos (1978) argued for the accuracy of his way of thinking about scientific progress by analyzing episodes from history (e.g., Niels Bohr's work on light emission). His method was to produce what he called "rational reconstructions" of how ideas developed, with historical details and the biographies of those involved relegated to footnotes (Larvor, 1998). The aim of such a reconstruction is to provide "a rational explanation of the growth of objective knowledge," not to offer a detailed historical account (Lakatos, 1970, p. 91). We note that some aspects of Lakatos's methodology of scientific research programmes have been criticized, primarily by Feyerabend (1981, 1993), and toward the end of the article, we argue that these criticisms are not relevant to our use of Lakatos's work.

Our goal in the next section is to demonstrate that Lakatos's ideas provide a helpful structure within which to understand the development of mathematics education research. To this end, we offer a rational reconstruction of the social turn identified by Lerman (2000).

The Social Turn in Mathematics Education: A Rational Reconstruction

Our goal in this section is to offer a sketch of how existing accounts of the social turn in mathematics education research (e.g., Clements & Ellerton, 1996; Lerman, 2000; Lerman, Xu, & Tsatsaroni, 2002; Mousley, 2015; Sakonidis, 2015) can be reinterpreted in terms of Lakatos's (1978) methodology of scientific research programmes. To this end, we offer a brief rational reconstruction of the social turn. We begin in the 1980s, a period during which constructivism was the dominant research programme in mathematics education. We first describe its hard core, heuristic, and protective belt.

The constructivist research programme in mathematics education developed out of Piaget's (1952) work on child development. A key assumption that made up the hard core was that knowledge construction is an individual process designed to maximize one's ability to make sense of the world. Individuals receive sensory input, filter it, and then actively organize it into mental schemas. Learning takes place when learners construct new knowledge by creating new schemas or reorganizing existing schemas. Knowledge is not passively received; it is actively created by the individual (e.g., Cobb, Yackel, & Wood, 1988; Noddings, 1990; Thompson, 2014; von Glasersfeld, 1990).

The constructivist heuristic also originated in Piaget's (1952) work. It emphasized careful study of how students make sense of new mathematical ideas, typically using clinical interviews in which participants were asked to orally reflect on their thought processes (e.g., Ginsburg, 1981; Swanson, Schwartz, Ginsburg, & Kossan, 1981). As befitted the programme's emphasis on individuals' constructions of personal schemas, clinical interviews were almost always conducted with a single participant who was typically observed making sense of a previously unseen open-ended mathematical task.

Various flavors of constructivism were developed, including von Glasersfeld's (1990) brand of radical constructivism. Radical constructivists were comfortable with the programme's hard core and its focus on individual sense making but went further by emphasizing that the goal of constructing schemas is merely to organize the individual's experience of the world and not to discover an objective external reality (e.g., von Glasersfeld, 1991). One way of interpreting these additional assumptions of radical constructivism is to see them as part of constructivism's protective belt. By adopting the radical position advocated by von Glasersfeld, a constructivist mathematics educator could avoid conflicts with the new fallibilist philosophies of mathematics that had been interpreted as denying the existence of objective mathematical knowledge (e.g., P. J. Davis & Hersh, 1980; Kitcher, 1983; Lakatos, 1976). However, the radical perspective was not part of the constructivist hard core. One could be what von Glasersfeld (1990) called a "trivial constructivist" and still happily endorse the constructivist hard core and heuristic.

By the end of the 1980s, the radical version of the constructivist research programme was dominant in mathematics education (R. B. Davis, 1990), but its emphasis on individual knowledge construction had difficulty accounting for several findings that seemed to suggest that learning had an important social component. For example, Lave (1988) found that adults' arithmetic strategies seemed to be contingent on the social setting in which they were performed. She found that although adults appeared to make few arithmetic errors when shopping for groceries, they performed relatively poorly on paper-and-pencil tests of arithmetic (but see Greiffenhagen & Sharrock, 2008, for an alternative interpretation). Similarly, Carraher (1988) found that children performed quite differently when conducting mathematics in school and "in the street." Further evidence that mathematical thinking was contingent on social factors came from Walkerdine's (1988, 1989) finding that negative attitudes toward, and lower attainment in, mathematics seemed to be more common in socially disadvantaged children and girls. These and other similar observations can be seen as a challenging anomaly to the constructivist research programme. How could the programme, for which the hard core and heuristic emphasized individual activity, deal with research findings that appeared to show that mathematical thinking and learning could not be fully understood without considering social contexts?

The notion of *social* constructivism (Ernest, 1991) that, following Vygotsky (1978), emphasized the role of social settings, history, and culture in forming individual knowledge can be seen as a rescue hypothesis that was added to constructivism's

protective belt to defend the programme against anomalies of the sort reported by Lave (1988), Carraher (1988), and Walkerdine (1988, 1989). Indeed, Mousley (2015) described social constructivism as “a compromise position” (p. 154) designed to explain “how the notion of individual cognition could remain viable in the context of social group interaction” (pp. 154–155). Although the role of social interaction in learning was emphasized by social constructivists, using constructs such as taken-as-shared knowledge, the construction of knowledge was still a fundamentally individual pursuit. The constructivist hard core remained.

As discussed earlier, Lakatos (1978) saw the addition of the rescue hypotheses as being a route through which we are able to judge whether a programme is progressing or degenerating. What did the social constructivist rescue hypothesis reveal about the constructivist programme? Opinions differed. Drawing on wider debates within the social sciences (for a review, see Bruner, 1996), Lerman (1996) argued that it created an internal contradiction. In Lakatos’s terms, Lerman was suggesting that constructivism was a degenerating programme. His point was simple. If knowledge was constructed by individuals through an idiosyncratic internal process, how could knowledge become shared—and known to be shared—within social groups? In other words, how could constructivism, with its focus on individual knowledge, explain intersubjectivity? He directly critiqued the social constructivist rescue hypothesis: “I suggest that the extension of radical constructivism toward a social constructivism, in an attempt to incorporate intersubjectivity, leads to an incoherent theory of learning” (p. 133). Lerman was arguing that the rescue hypothesis was incompatible with the constructivist hard core and that the research programme was degenerating to such an extent that it should be discarded: “Mathematics education would benefit from abandoning constructivism as a view of how people learn” (p. 133).

As well as arguing that the constructivist programme should be abandoned, Lerman (1996) also discussed the research programme that he felt should replace it. Although constructivism’s hard core emphasized that knowledge is constructed by individuals, the sociocultural research programme supported by Lerman instead had a hard core that assumed that thinking, reasoning, and knowledge were all products of social activity. Vygotsky (1986) characterized the distinction: “In our conception, the true direction of the development of thinking is not from the individual to the social, but from the social to the individual” (Vygotsky, 1986, p. 36). Clearly, this difference in the hard core also led to a substantially different heuristic. If thinking is constituted in social interactions, then the individual clinical interview is not likely to be a suitable way to study thinking. Instead, the sociocultural research programme’s heuristic included a much greater emphasis on observations of classroom discourse (e.g., Goos, Galbraith, & Renshaw, 2002).

As Lakatos would have predicted, Lerman’s (1996) argument was rejected by those who wished to defend the constructivist research programme (e.g., Steffe & Thompson, 2000), and they continued to conduct and publish constructivist research that was widely read (e.g., Steffe & Ulrich, 2014; Thompson, 2014). However, when Lerman (2000) claimed that there had been a “social turn in

mathematics education research,” he was asserting that the sociocultural research programme was growing—in terms of the number of researchers and its influence—at the expense of the constructivist research programme. As discussed above, Lerman based this claim on analyzing a sample of research papers from the 1990s (Lerman et al., 2002). The extent to which the social turn continued through the 2000s is a matter of debate (e.g., Gates & Jorgensen, 2015; Jablonka & Bergsten, 2010).

Topic Modeling as a Method

The social turn is an example of theory change in which a research programme was apparently abandoned by a subset of its followers. However, Lakatos’s notion of a research programme indicates that there are other ways in which programmes can develop. For instance, their domain of applicability might shift (the content being studied could change) or they might modify their heuristic (the main methods that they use could develop). To fully understand how the discipline of mathematics education has changed since *ESM* and *JRME* began publishing, we require a method that allows developments in the hard core, heuristic, and domain content to be identified. In this section, we introduce topic modeling, a method that allows such developments to be identified through the analysis of the language used in research papers.

The rationale for our use of a linguistic approach is that a research programme’s hard core, its heuristic, and the domain content that it is used to analyze all have characteristic linguistic features. For instance, we would expect a research paper that reports a constructivist analysis of geometry learning to contain words such as *triangle*, *circle*, and *angle* but also words such as *schema*, *constructivism*, and *interaction*.

Topic modeling is a computational method designed to summarize large collections of texts by a small number of conceptually connected topics or themes (Blei, Ng, & Jordan, 2003; Grimmer & Stewart, 2013). The aim is to discover the main themes that are present in a large unstructured collection of documents by analyzing the patterns with which words co-occur. One way of understanding topic modeling is to imagine how documents could be created from a preexisting set of topics. A topic is defined by a probability distribution over words. So, in one topic, the word *angle* would have a high weighting, and in another, it would have a low weighting, and similarly for *schema*, *triangle*, and so on. We can imagine creating a document by selecting a distribution over topics. For instance, a given document might be composed of 40% of words from Topic 1, 15% from Topic 2, 0% from Topic 3, and so on. Given this setup, documents of a given length can be created simply by selecting words from the topics with the appropriate frequency. For instance, every time a word is selected for our document, there would be a 40% chance of it coming from Topic 1, and within Topic 1, there would be some chance of it being *angle*, some chance of it being *triangle*, and so on. This method uses the so-called “bag of words” model of text, which dramatically simplifies language by ignoring both word order and “stop words” (words such as *the* and *a* that are topic independent).

Topic modeling can be thought of as carrying out this text construction process in reverse. The method starts with the documents, assumes that they were constructed via this process, and identifies which topics would be most likely to have produced them. Topic modeling is a computationally demanding task that relies on latent Dirichlet allocation algorithms that identify the topics that best fit the documents (Blei et al., 2003). The method is somewhat analogous to a quantitative version of grounded theory; there are no preconceived ideas about the topics that will emerge, and individual words are tagged with codes that identify the topics with which they are associated. Once the modeling process has occurred, we can study the composition of each document. For instance, we may discover that Document 10 is made up of 4% Topic 1 words, 60% Topic 2 words, and so on.

One difficulty with the topic-modeling approach is that one must specify in advance the number of topics the algorithm should find. By so doing, the researcher can determine the granularity of the analysis. One method to decide on a suitable number of topics is to assess how well the topic model fits the texts using a measure known as perplexity. The lower the *perplexity* of a model with a given number of topics, the better the model's fit (Blei et al., 2003). Perplexity is calculated by fitting a topic model to a subset of the texts and then assessing its fit on the remaining texts. This process is repeated for models with different numbers of topics. It is always possible to reduce the perplexity of a topic model by increasing the number of topics, but, at some point, the gain in fit will be offset by the increased difficulty of interpreting the larger number of topics. Jacobi, van Atteveldt, and Welbers (2016) proposed that the number of topics to retain should be assessed using a method analogous to Cattell's (1966) scree test in the context of factor analyses. By calculating the perplexity of models with different numbers of topics, one can assess if there is a point at which the reduction in perplexity appears to "level off." But Jacobi et al. emphasized that as with factor analyses, one major criterion for selecting the number of topics when producing a topic model is the interpretability of the resulting topics.

The topic-modeling approach has several advantages over traditional approaches to studying a field's historical development. First, the approach is extremely inclusive. It would be unrealistic for a researcher to read and analyze every *ESM* and *JRME* paper ever published, but the topic-modeling approach can take account of this number of texts, implying that important historical trends are unlikely to be missed. Second, the approach is relatively neutral; because the analysis is conducted algorithmically, it does not prioritize one historical trend over another. However, this neutrality comes at a cost. The results of our analysis are purely descriptive; the topic-modeling method identifies phenomena that must then be interpreted, a task that we attempt later in the article. Naturally, these interpretations are more subjective and open to criticism than the topic-modeling analysis itself.

Identifying the Topics

In the nearly 5 decades since publication of the first issues of *ESM* and *JRME*, the two journals have established themselves as the leading international venues

for research in mathematics education. Indeed, in 2012, a project by the European Mathematics Society and the European Society for Research in Mathematics found that *ESM* and *JRME* were the only two journals given the highest possible quality rating by at least two thirds of the expert mathematics educators sampled (Toerner & Arzarello, 2012). Consistent findings were reported by Nivens and Otten (2017) and Williams and Leatham (2017). We therefore focused our analysis on *ESM* and *JRME*, but we highlight that this does restrict our conclusions to English-language mathematics education research.

We downloaded every “article” published between 1968 and 2015 by *ESM* and *JRME* from the journals’ websites. These articles included everything published within the journals and stored as PDF files on the websites, including research papers, editorials, book reviews, and calls for papers. These PDF files were converted to plain text using ABBYY FineReader OCR Pro (Version 12.1.4), and “noncontent,” such as copyright statements or watermarks, was removed. Our final data set consisted of 1,933 files (9.49 million words) from *JRME* and 2,062 files (14.48 million words) from *ESM*.

We used MALLET (Version 2.0.8RC2), a UNIX command-line topic-modeling tool (McCallum, 2002), to calculate possible topic models. We first removed all the “stop words”—very common English words, such as *the*, *is*, and *a*, that would not be topic specific—on MALLET’s default list. Inspection of the perplexity graph (discussed above), shown in Figure 1, suggested that 35 topics seemed to be a reasonable choice, and an inspection of the topic models generated by different numbers of topics suggested that the overall message from the data did not seem to be sensitive to varying the number of topics slightly.

We interpreted each topic that the algorithm identified using two different approaches. We first studied the words that were highly characteristic of each topic (in the sense that, when a word from this topic was inserted into a new document during our counterfactual document creation process, these highly characteristic words had a high probability of being selected). For instance, the words with the highest probabilities in the first topic identified by the algorithm were *proof*, *proofs*, *mathematics*, *mathematical*, *reasoning*, *argument*, *students*, *arguments*, *statement*, *deductive*, and *proving*. From this, it seemed clear that this topic is concerned with proof and argumentation, and we gave it the name “proof and argumentation.” Second, for each topic, we studied those papers that had particularly high proportions of included words. For instance, the paper with the highest proportion of words from the “proof and argumentation” topic (64% excluding stop words) was Weber’s (2008) article about how mathematicians validate proofs. By studying these papers, we were able to further understand the nature of the topics, which contributed to our choice of names.

Some of the topics were not related to the content of research. For instance, four topics were concerned with journal administration (e.g., announcements of special issues, advice to prospective authors, or lists of editorial board members), and another three consisted of non-English words (*ESM* has published articles in both French and German). Clearly, articles written in French tend to have more

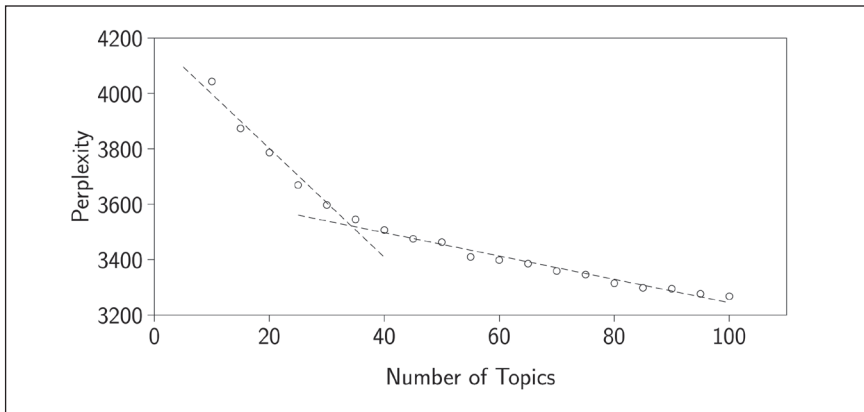


Figure 1. The perplexity of topic models with varying numbers of topics. The dashed lines show our interpretation of where the graph “levels off.”

linguistic similarity to each other than to those written in English, regardless of their academic content. We do not discuss these topics further, although it would be worthwhile to establish whether similar topics emerge from an analogous study of non-English language mathematics education research.

The remaining 28 topics were assigned names based on their defining words and the nature of the papers that had particularly high proportions of words from them. The topic names, together with each topic’s characteristic words and the paper with the highest proportion of words from each topic, are shown in Table 1. To enable readers to better understand these 28 topics, we have listed the 10 papers with the highest proportions of words from each topic in a spreadsheet available at <https://doi.org/10.6084/m9.figshare.4877429>.

We make several remarks about Table 1. First, readers will have noticed that some words define more than one topic (e.g., group). This highlights one advantage of topic modeling. When the word group appears near the words *set*, *elements*, and *operation*, it is likely to have a different meaning from when it appears near the words *treatment*, *experimental*, *pretest*, and *posttest*.

Second, although it was straightforward to identify most of the topics by studying their characteristic words and their most representative papers, in other cases, this was not clear. Here, we briefly justify our characterizations for those topics where this may not be obvious.

Papers made up of particularly high proportions of words from the “teaching approaches” topic typically reported on discussions or evaluations of particular classroom teaching strategies. For example, Leikin and Zaslavsky’s (1997) investigation of student interactions in small-group settings had a high proportion of words from this topic, as did Brookhart, Andolina, Zuza, and Furman’s (2004) discussion of student self-assessment.

The common theme among papers that had high proportions of words from the “didactical theories” topic was that they used or discussed theories from the

Table 1
The 28 Topics, Each With the Words That Best Characterize Them (in Order of Probability) and the Paper With the Highest Proportion of Words

Topic name (ordered alphabetically)	Characteristic words (top 20)	Paper with the highest proportion of words from the topic
Addition and subtraction	children number children's addition counting subtraction strategies child numbers strategy arithmetic problems mental task count facts instruction development tasks ten	The development of counting strategies for single-digit addition (Baroody, 1987)
Analysis	function concept limit definition calculus numbers number students negative infinite sequence mathematical infinity image formal process functions derivative point points	An empirical study of students' understanding of a logical structure in the definition of limit via the ϵ -strip activity (Roh, 2010)
Constructivism	mathematics learning mathematical knowledge development cognitive theory education activity research process understanding processes social individual press concepts view conceptual construction	Interaction or intersubjectivity? A reply to Lerman (Steffe & Thompson, 2000)
Curriculum (especially reform)	mathematics curriculum school achievement students teachers student national textbooks standards grade high assessment schools content level curricula reform data textbook	The impact of prior mathematics achievement on the relationship between high school mathematics curricula and postsecondary mathematics performance, course-taking, and persistence (Post et al., 2010)
Didactical theories	mathematical knowledge students teacher theoretical activity process teaching analysis research mathematics situations situation didactic didactical classroom springer case learning context	Introduction teaching situations as object of research: Empirical studies within theoretical perspectives (Laborde & Perrin-Glorian, 2005)
Discussions, reflections, and essays	question time make fact point part questions process made problem case view important kind work sense situation answer find paper	Letter to the editor (Roberts, 2001)
Dynamic geometry and visualization	geometry visual figure spatial triangle angle geometric angles task logo diagram diagrams properties shapes triangles shape level van sides tasks	Facility with plane shapes: A multifaceted skill (Warren & English, 1995)

Topic name (ordered alphabetically)	Characteristic words (top 20)	Paper with the highest proportion of words from the topic
Equity	mathematics education school American cultural countries students social Chinese schools culture teachers African equity children educational ethnomathematics country Japanese parents	Attention deficit disorder? (Silver, 2003)
Euclidean geometry	geometry line point points figure circle plane lines geometrical angle triangle space straight parallel geometric fig figures theorem segment Euclidean	Inversive geometry (Coxeter, 1971)
Experimental designs	test study group scores research items table mathematics tests significant results groups variables analysis treatment ability performance item journal experimental	Interactions between structure-of-intellect factors and two methods of presenting concepts of modulus seven arithmetic: A follow-up and refinement study (Behr & Eastman, 1975)
Formal analyses	set concept number numbers elements concepts structure sets group order operations examples model operation relation system properties element relations objects	Checker games in operational systems as media for an inductive approach to teaching algebra (Steiner & Kaufman, 1969)
Gender	mathematics differences achievement girls gender boys anxiety attitudes sex school performance study research high females ability educational factors motivation males	Gender differences in a psychological model of mathematics achievement (Ethington, 1992)
History and obituaries	mathematics mathematical education book teaching history chapter mathematicians science university theory problems school historical geometry ideas curriculum educational development work	The epos of Euclidean geometry in Greek secondary education (1836– 1985): Pressure for change and resistance (Toumasis, 1990)
Mathematics education around the world	pupils mathematics school teaching learning schools year work level education secondary teachers teacher mathematical project pupil years primary children educational	Change in mathematics education since the late 1950's—Ideas and realisation West Indies (B. J. Wilson, 1978)

Topic name (ordered alphabetically)	Characteristic words (top 20)	Paper with the highest proportion of words from the topic
Multilingual learners	language mathematics English text reading writing mathematical word words texts linguistic languages learners bilingual learning written discourse meaning classrooms read	Using two languages when learning mathematics (Moschkovich, 2007)
Novel assessment	tasks task assessment mathematical students mathematics modelling knowledge model cognitive competence results models level solutions quality performance springer information study	Modes of modelling assessment—A literature review (Frejd, 2013)
Observations of classroom discussion	teacher classroom mathematical discourse interaction discussion analysis episode activity class learning interactions group social ideas talk work participation thinking it's	Mathematical micro- identities: Moment-to- moment positioning and learning in a fourth-grade classroom (Wood, 2013)
Problem solving	problem problems solving solution solve mathematical problem-solving strategies word solutions information model strategy solved processes structure study process table correct	Recall of mathematical problem information: Solving related problems (Silver, 1981)
Proof and argumentation	proof proofs mathematics mathematical reasoning argument students arguments statement deductive proving examples number true mathematicians theorem statements prove logical argumentation	How mathematicians determine if an argument is a valid proof (Weber, 2008)
Quantitative assessment of reasoning	correct items errors reasoning item responses answer number table subjects answers proportional grade incorrect ratio intuitive numbers error response type	The development of proportional reasoning and the ratio concept Part I— Differentiation of stages (Noelting, 1980)
Rational numbers	fractions number fraction division numbers unit multiplication rational parts decimal units knowledge understanding fractional pieces operations partitioning part scheme multiplicative	Michael's fraction schemes (Saenz-Ludlow, 1994)

Topic name (ordered alphabetically)	Characteristic words (top 20)	Paper with the highest proportion of words from the topic
School algebra	algebra algebraic equation function equations functions graph students graphs expressions representations expression computer variable arithmetic linear symbolic values understanding representation	Evolution of a teaching approach for beginning algebra (Banerjee & Subramaniam, 2012)
Semiotics and embodied cognition	mathematical objects semiotic gestures signs meaning object sign gesture fig language mathematics radford activity springer hand space line body metaphor	Grounded blends and mathematical gesture spaces: Developing mathematical understandings via gestures (Yoon, Thomas & Dreyfus, 2011)
Sociocultural theory	mathematics education social school practices practice work mathematical knowledge research identity cultural activity power learning context discourse people theory everyday	Symbolising the real of mathematics education (Pais, 2015)
Spatial reasoning	length area unit figure reasoning units rate number measurement pattern height change time measure relationship quantities cubes volume generalization distance	Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom (Battista, 1999)
Statistics and probability	probability data sample statistics statistical reasoning chance average sampling distribution dice thinking outcomes responses grade level probabilistic variation events population	A framework for assessing and nurturing young children's thinking in probability (Jones, Langrall, Thornton, & Mogill, 1997)
Teachers' knowledge and beliefs	teachers teacher mathematics teaching knowledge beliefs lesson classroom learning content practice mathematical education professional student thinking lessons school preservice instruction	Preservice teachers' sources of decisions in teaching secondary mathematics (Bush, 1986)
Teaching approaches	students student class group mathematics school time study questions groups work level classes year problems asked learning instruction high activities	Teacher benefits from using impromptu writing prompts in algebra classes (Miller, 1992)

continental European tradition, such as Chevallard's (1999) anthropological theory of didactics (ATD) or Brousseau's (2006) theory of didactical situations (TDS). For instance, Barbé, Bosch, Espinoza, and Gascón's (2005) analysis, using ATD, of the teaching of functions in Spanish high schools had a high proportion of words from this topic. As would be expected, as well as having many words from the "didactical theories" topic, this paper also had a high proportion from the "analysis" topic.

Papers that had high proportions of words from the "experimental designs" topic typically used experimental designs with random allocation of participants to conditions. Of the 10 papers with the highest proportions of words from this topic, nine used random allocation at the participant level to investigate a variety of different research questions.

Papers that had high proportions of words from the "quantitative assessment of reasoning" topic typically used large samples to document reasoning behavior. The majority of the 10 papers with the highest proportions of words from this topic investigated aspects of proportional reasoning, but there was also a survey of responses to the Wason selection task (Adi, Karplus, & Lawson, 1980) and an investigation of adults' reasoning about natural numbers (Vamvakoussi, Van Dooren, & Verschaffel, 2013).

All the papers with high proportions of words from the "sociocultural theory" topic used social theories to analyze aspects of mathematics learning. Indeed, two of the papers with the highest proportions from this topic were from a special issue titled "Social Theory and Research in Mathematics Education" (Morgan, 2014; Pais & Valero, 2014).

Of the 10 papers with the highest proportion of words from the "constructivism" topic, seven were authored by Cobb or Steffe (e.g., Cobb, Yackel, & Wood, 1992; Steffe & Kieren, 1994), and the remaining papers either used, discussed, or critiqued constructivist approaches to mathematics education.

Papers with high numbers of words from the "formal analyses" topic consisted of attempts to provide formal mathematical analyses of educational tasks or theories. For instance, Wittmann's (1973) attempt to provide an algebraic model of an aspect of Piagetian theory had a large proportion of words from this topic.

Finally, there was a topic—"discussions, reflections, and essays"—that seemed to represent reflective, nonempirical discussions of various issues. Papers that had high proportions of words from this topic often discussed meta-level research issues. Examples included letters to the editor, a discussion of citation practices in mathematics education (Leatham, 2015), and various editorials (e.g., Williams, 2007).

Finally, we attempted to verify that our interpretations of the topics were reasonable by studying the linear combinations of "topic proportions" for particularly highly cited papers from each journal. For instance, our topic model suggested that Tall and Vinner's (1981) paper on students' concept images of limits and continuity was mostly made up of the "analysis" topic (which provided

54% of the paper's words²) and the "discussions, reflections, and essays" topic (24% of the paper's words). Similarly, the model suggested that Yackel and Cobb's (1996) paper, which introduced the notion of sociomathematical norms, was largely made up of the "observations of classroom discussion" (42%) and "constructivism" (27%) topics. Both of these results seemed consistent with the content of the respective papers.

Changes Over Time

To look at how the prominence of each topic has changed over time, we used the linear combinations of topic proportions to calculate the mean proportion of words from each topic published by each journal in each year. For example, we took the proportion of words from the "proof and argumentation" topic in every paper published by *ESM* in 2015 and calculated the mean. This revealed that, normalized by paper length, 2.7% of the words published in *ESM* in 2015 were related to proof and argumentation. We did the same for *JRME* (0.22%) and for every year in our sample. We then calculated similar figures for the other 27 topics. This allowed us to track the extent to which each topic has been present in the two journals over the last 5 decades.

To help organize our discussion, we further classified our topics into five broad categories. Recall that Lakatos (1978) suggested that academic disciplines develop through the competition of research programmes, which can be characterized by their heuristics and hard cores (and, to a lesser extent, their protective belts). As we noted earlier, this means that there are at least two ways of tracking the development of an academic discipline. One is to consider the relative prominence of topics related to research programmes' heuristics and hard cores; a second is to look at the substantive content to which the research programmes have been applied. We therefore categorized each of our topics as being primarily related to a research programme's hard core, a research programme's heuristic, or substantive domain content. Because mathematics education research covers a broad range of content, to help structure our discussion, we further divided the domain content topics into three subcategories relating to "mathematical content," "mathematical processes," and "teachers and learning environments."

As discussed above, Lakatos (1978) himself accepted that the distinction between a programme's hard core and heuristic is sometimes a matter of convention, so this distinction is necessarily subjective. In contrast, we found it relatively straightforward to identify those topics related to mathematical content, mathematical processes, and teachers and learning environments. One ambiguity came from the "quantitative assessment of reasoning" topic, which seemed to be related to both a heuristic (large-scale surveys) and a mathematical process (reasoning). For the purposes of the discussion and figures that follow, we categorized this topic as being related to mathematical processes. Similarly, the "history and obituaries" topic seemed not easily to fit within one of our three domain content subcategories; in the discussion and figures below, we have included it within the "teachers and learning environments" section.

² For the remainder of the article, when we cite the percentage or proportion of words from a particular topic, we are excluding stop words (particularly common words such as *the*, *it*, or *a*) from the denominator.

Our discussion of these changes over time is organized into two main sections. First, by considering the topics categorized as “mathematical content,” “mathematical processes,” and “teachers and learning environments,” we consider how the prominence of different research foci has changed in the last 5 decades. Second, we discuss the changing theoretical perspectives adopted by *ESM* and *JRME* authors by considering topics related to the hard cores and heuristics of different research programmes.

Changes in Domain Content

Figures 2, 3, and 4 show the changes in mean topic proportions per year for topics related to the “mathematical content,” “mathematical processes,” and “teachers and learning environments” subcategories, respectively. The points on these graphs show the mean proportion of words from each paper published during the given year that were from the given topic. In each figure, graphs are ordered according to positive correlations between year and mean topic proportion (taking the average for each journal), with the highest at the top. For instance, the correlations between the mean topic proportions and year for the “school algebra” topic (shown at the top of Figure 2) were $r = +.49$ for *ESM* and $r = +.60$ for *JRME*, whereas the equivalent correlations for the “Euclidean geometry” topic (shown at the bottom of Figure 2) were $r = -.61$ for *ESM* and $r = -.44$ for *JRME*.

We make several remarks about these figures, first concerning those topics related to mathematical content. Figure 2 shows that there has been a substantial decline in interest in Euclidean geometry since the 1970s. This has especially been the case in *ESM*, which devoted a great deal of attention to this topic in its early years. However, the fall in prominence of Euclidean geometry in *JRME* has also been substantial but is somewhat disguised by the scale on the vertical axis (the correlations between year and topic proportion were $r = -.61$ and $r = -.44$ for *ESM* and *JRME*, respectively). In contrast, both “school algebra” and “analysis” have received gradually increasing levels of interest, whereas the extent to which “rational numbers” and “statistics and probability” have been discussed has not substantially changed since the 1970s. The “addition and subtraction” topic shows a peak of interest during the 1980s, but the two journals have apparently devoted less attention to it in recent years.

Two of the “mathematical processes” topics, shown in Figure 3, also show peaks during the 1980s. There seems to have been somewhat more interest in problem solving during that decade than there is today, and the same is true for the “quantitative assessment of reasoning” topic, especially in *JRME*. The “proof and argumentation” topic seems to have steadily increased in prominence since 2000, but the remaining “mathematical processes” topics show no strong trends.

Several topics related to teaching and learning environments, shown in Figure 4, have received an increasing level of interest over the past 5 decades. The most notable is the increased focus on teacher knowledge and beliefs, which has become steadily more prominent in both journals since the 1980s, perhaps following Shulman’s (1986) influential work. The “curriculum (especially reform)” topic has also grown in prominence since 1990; perhaps unsurprisingly, given NCTM’s

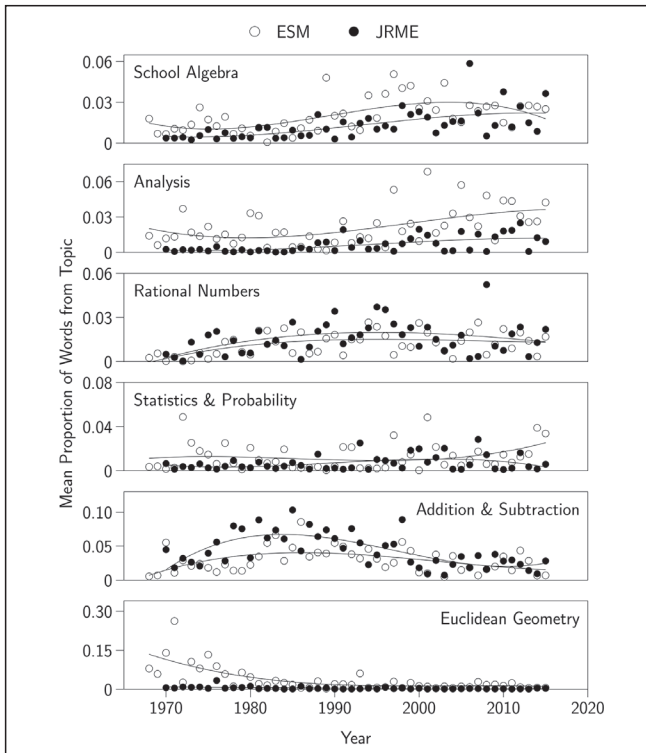


Figure 2. The mean proportion of words (excluding stop words) from “mathematical content” topics published by each journal per year. Lines show cubics of best fit. Note that because these graphs are designed to show within-topic changes over time, they have different y-axis scales.

involvement in curriculum reform efforts, the bulk of this increase has been in *JRME*. Similarly, in recent years, there appears to have been an increased focus on designing and evaluating novel assessment methods in both journals.

There is also a group of topics shown in Figure 4 that have not shown substantial changes in the extent to which they have been discussed. The “multilingual learners,” “teaching approaches,” “equity,” and “history and obituaries” topics all fall into this group. The “gender” topic shows a notable peak: *ESM* published almost no articles with words from this topic until the late 1970s, a situation that changed substantially during the 1980s. Since that decade, however, both *ESM* and *JRME* appear to have published fewer articles focused on gender. Finally, discussions of “mathematics education around the world” appear to have become less of a priority than they once were.

Changes in Research Programmes

Figure 5 shows changes in prominence over time for those topics categorized as being related either to a research programme’s hard core or to its heuristic.

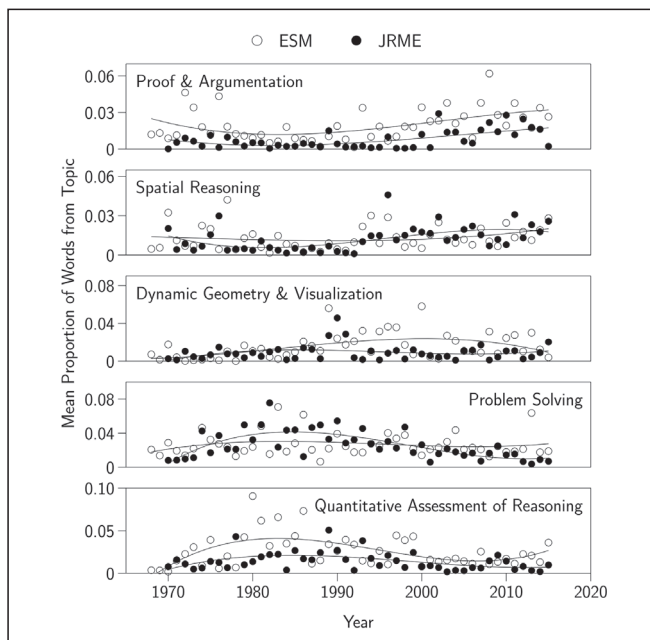


Figure 3. The mean proportion of words (excluding stop words) from “mathematical processes” topics published by each journal per year. Lines show cubics of best fit. Note that because these graphs are designed to show within-topic changes over time, they have different y-axis scales.

Again, the graphs are ordered by the strength of the correlation between year and mean topic proportion. In the sections that follow, we discuss the three main trends that can be seen in these figures. First, we consider the social turn as identified by Lerman (1996). Second, we identify and discuss the increased level of theoretical diversity seen in the discipline since the late 1990s. Finally, we examine the dramatic decline in experimental methods seen since the 1970s.

Before proceeding to discuss these three issues, we briefly note two other trends. First, the prominence of the “formal analyses” topic has declined significantly since the 1970s. In articles published by *ESM* in the 1970s, between 10% and 20% of words were from this topic, which relates to formal mathematical analyses of educational theories or situations. In contrast, very few researchers since the mid-1980s appear to have adopted this heuristic. The identification of the existence, and subsequent decline, of this “formal analyses” topic—which appears not to be widely commented on in existing historical accounts of the field’s development—gives some credibility to our earlier suggestion that the inclusive nature of the topic-modeling approach allows us to identify trends that other methods may miss. Second, we note that the “discussions, reflections, and essays” topic has shown a steady decline in prominence in *ESM*. In contrast, *JRME* seems to have maintained a steady—but lower—rate of publication of words from this topic.

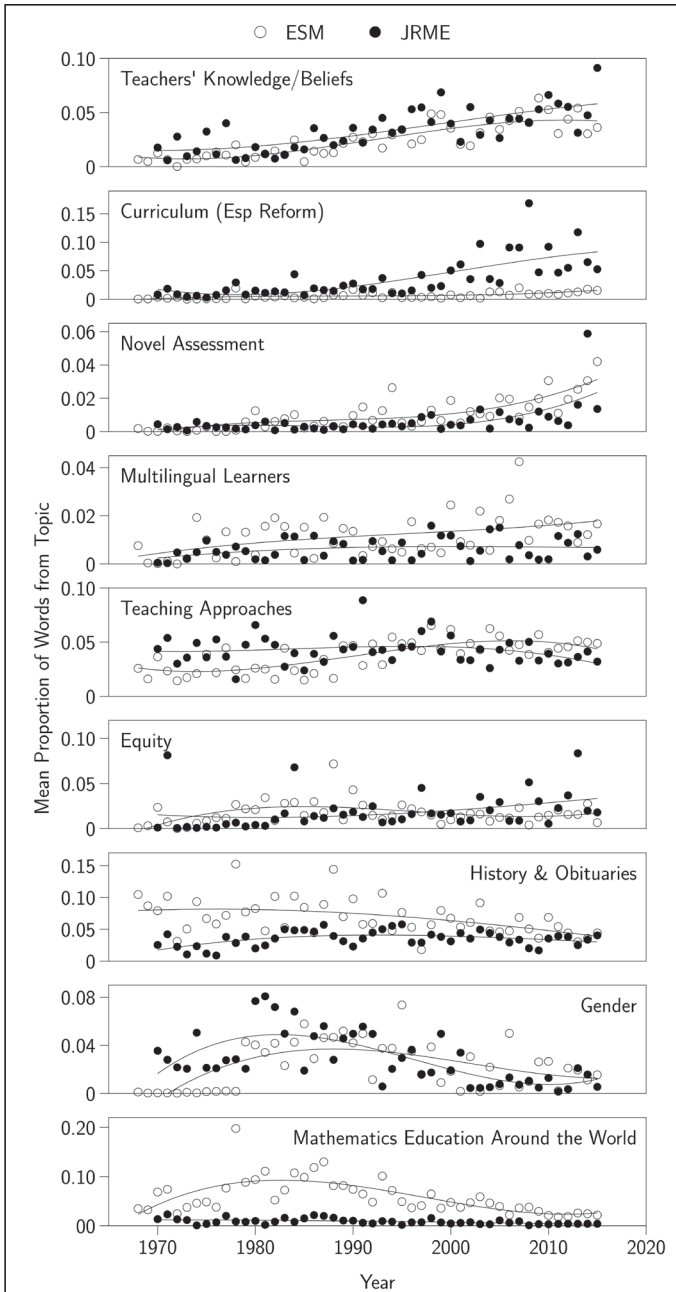


Figure 4. The mean proportion of words (excluding stop words) from “teaching and learning environments” topics published by each journal per year. Lines show cubics of best fit. Note that because these graphs are designed to show within-topic changes over time, they have different y-axis scales.

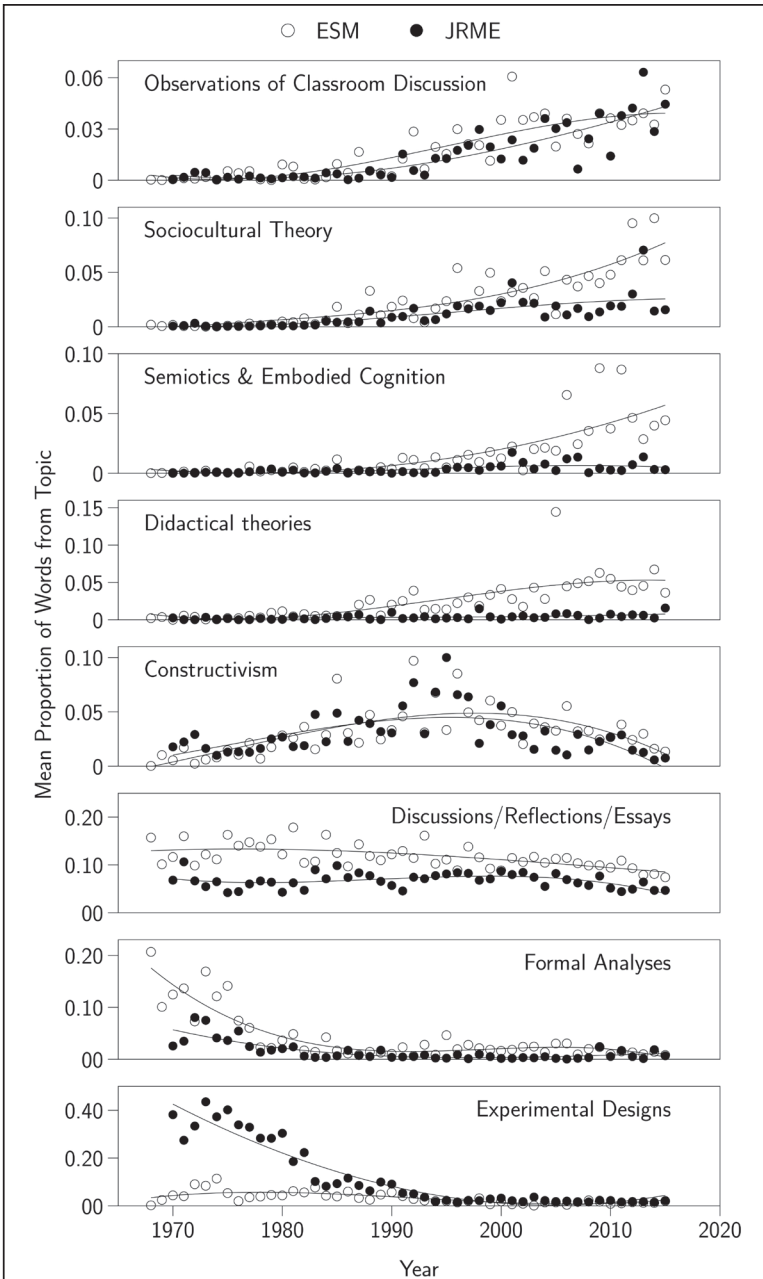


Figure 5. The mean proportion of words (excluding stop words) from the eight hard-core and heuristic topics published by each journal per year. Lines show cubics of best fit. Note that because these graphs are designed to show within-topic changes over time, they have different y-axis scales.

The social turn. Earlier in the article, we offered a rational reconstruction of the social turn in which we suggested that the emergence of the sociocultural research programme in the 1990s was in reaction to what Lerman (1996) considered to be a degenerating constructivist programme. By Lerman’s account, if the social turn had become embedded into the mainstream mathematics education literature at the expense of the constructivist programme, then over time we would have expected to have seen a gradual increase in prominence of the “sociocultural theory” topic and a gradual decrease in prominence of the “constructivism” topic.

This is precisely what our data show. Figure 6 shows the proportions of words from each topic by year collapsed across journals. The “constructivism” topic shows a large peak of interest from the late 1980s to mid-1990s before declining in prominence from the late 1990s onwards. At around the same time that the field’s interest in constructivism declined, the proportion of words from the “sociocultural theory” topic began to increase and is now at a roughly similar level to that of the “constructivism” topic at its peak (when it was widely seen as dominating mathematics education research). The field’s interest in constructivism is now at the low levels seen in the 1970s. Although some have questioned the extent to which the social turn represents a lasting change to the discipline (e.g., Gates & Jorgensen, 2015; Jablonka & Bergsten, 2010), this analysis suggests that it was a significant development, which has had lasting consequences.

Although the sociocultural and constructivism hard cores are distinct, it is notable that the sociocultural research programme shares a similar heuristic to the social branch of the constructivist programme, namely, a focus on observing classroom interactions. Indeed, this heuristic change—from individual clinical interviews to observations of interactions—is one way of characterizing the shift from radical constructivism to social constructivism. Given this, it is perhaps unsurprising to see that Figure 5 shows a complementary steady increase in the proportion of words coming from the “observations of classroom discussion” topic since the 1990s. This trend started at around the time that social constructivism was identified by Ernest (1991) and has continued through the social turn.

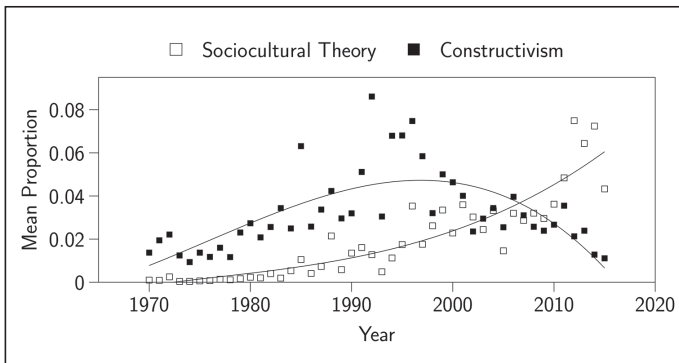


Figure 6. The social turn. The mean proportion of words from the “sociocultural” and “constructivist” topics by year collapsed across journals. Lines are cubics of best fit.

Theoretical diversification. “Sociocultural theory” is not the only topic from the hard core and heuristic category that has seen a substantial increase in prominence since 2000. Both the “semiotics and embodied cognition” and the “didactical theories” topics have shown a similar development over the same period, particularly in *ESM*. We might, therefore, characterize mathematics education as currently being in a phase of theoretical diversity in which many research programmes are competing for researchers’ attention.

Notably, both the “semiotics and embodied cognition” and “didactical theories” topics cover multiple research programmes. Although both the semiotics and the embodied cognition research programmes have hard cores that emphasize the importance of noncognitive factors to mathematical thinking, they differ in where their emphasis is placed. Semioticians focus on the roles of signs and symbols in mathematical thought (e.g., Presmeg, Radford, Roth, & Kadunz, 2016), whereas embodied cognition researchers focus on the role of the body outside of the brain (e.g., Núñez, Edwards, & Matos, 1999). Of course, some researchers conceptualize certain aspects of bodily movement—gestures or gazes, for instance—as being signifying acts (e.g., Radford, 2003; Roth, 2012), which provides a natural link between the two research programmes and which may explain why “semiotics and embodied cognition” emerged as one topic. Similarly, at least two research programmes are contained within our “didactical theories” topic: the ATD of Chevallard (1999) and the TDS of Brousseau (2006). Therefore, if anything, our analysis underestimates the extent to which the field has diversified.

This theoretical diversification has been noted before (e.g., Bikner-Ahsbahs & Prediger, 2010; Sriraman & English, 2005). Whereas some researchers have argued that a diversity of theoretical approaches is valuable for allowing multiple perspectives to be brought to the same phenomena (e.g., Reid & Knipping, 2010; Simon, 2009), others have highlighted dangers. For instance, Dreyfus (2006) complained that mathematics educators “tend to invent theories, or at least theoretical ideas, at a pace faster than we produce data to possibly refute our theories” (p. 78).

Dreyfus’s (2006) argument can be interpreted within Lakatos’s (1978) framework. Recall that Lakatos suggested that progress in an academic discipline comes about when research programmes compete for researchers’ attention. One way of assessing a programme is by considering how it incorporates anomalies into its protective belt. Those programmes that are progressing will eventually attract more attention than those that are degenerating. Therefore, a degree of programme diversity is a sign of a maturing academic discipline, as long as these programmes are challenging each other for researchers’ attention by identifying anomalies and evaluating the resulting modifications to protective belts. Via this process, researchers can decide whether a particular programme is progressing or degenerating and allocate their attention accordingly. As long as this process occurs, some programme diversity is a strength (cf. Cobb, 2007). However, if, as Dreyfus presumably felt, not all programmes in mathematics education are regularly challenged in this way, this may lead to a lack of progress.

One attempt to deal with the field's recent theoretical diversity, and concerns about a lack of anomalies, is to "network theories" (e.g., Bikner-Ahsbahs & Prediger, 2010; Kidron, Lenfant, Bikner-Ahsbahs, Artigue, & Dreyfus, 2008). Through this process, researchers attempt to understand, synthesize, and perhaps even unify different theoretical approaches. The aim is to find connections as far as is possible and useful but not further than that. Whether complete unification is possible would seem to turn on whether different theoretical approaches are genuinely different programmes, in the sense that they have incompatible hard cores, or whether they have similar hard cores that, for historical reasons, are merely expressed using different terminology. In other words, these networking attempts could be a helpful method of determining whether or not different theoretical approaches are genuinely different research programmes.

Regardless of whether unification is possible, Lakatos (1978) emphasized that understanding and engaging with different research programmes is likely to contribute to identifying and evaluating anomalies. Although it is possible to identify anomalies from within a research programme, Lakatos wrote that "it is only *constructive criticism* which, with the help of rival research programmes, can achieve real success" (p. 92) and that "*the sooner competition starts, the better for progress*" (p. 69). Feyerabend (1985, 1993) also emphasized the advantages that accrue to a field that has multiple competing research programmes: "*The best criticism is provided by those theories which can replace the rivals they have removed*" (Feyerabend, 1985, p. 110). The social turn provides an example of this: Lerman's (1996) critique of social constructivism was heavily informed by his knowledge of sociocultural theory, the rival programme.

The experimental cliff. Although the social turn was an important development in the history of mathematics education research, it is not the most striking trend seen in our data. During the 1970s, the "experimental methods" topic was dominant to an extent not seen in any other topic during any subsequent period. Indeed, between 30% and 40% of words published in *JRME* in the 1970s were from this topic. Although this trend is most striking in *JRME*, a similar pattern can be seen in *ESM*. The correlations between the topic's proportion and year are $r = -.86$ for *JRME* and $r = -.63$ for *ESM*. Across both journals, the mean proportion of words in each paper from the "experimental methods" topic was 22% in the 1970s, 10% in the 1980s, 3% in the 1990s, 1.7% in the 2000s, and 1.5% in the 2010s. We refer to this development as the "experimental cliff."

It is notable that there appears to have been no recent upturn in the prominence of this topic despite educational policymakers actively trying to encourage, through targeted funding initiatives, the use of experimental designs in education research. For instance, the U.S. Department of Education's Strategic Plan 2002–2007 set an explicit target, stating that 75% of funded projects that address causal questions should use randomized experimental designs (U.S. Department of Education, 2002), and, in 2011, the UK government created a large new source of educational research funding, the Education Endowment Foundation, that was

required to fund only randomized experimental designs. These official encouragements have yet to have any effect on the frequency with which experimental designs appear in *ESM* or *JRME*.

Experimental designs are a subset of quantitative research methods. Although the “experimental designs” topic has substantially decreased in prominence, it is notable that we did not identify a “qualitative research methods” topic showing an upward trend of a similar magnitude. Our explanation for this observation is that articles that report qualitative methods tend to be more linguistically diverse than those that report quantitative methods. Consistent with this suggestion, although we found no unified topic focused on qualitative methods, we did find topics concerned with classroom discourse, sociocultural theory, semiotics and embodied cognition, and didactical theories. All of these topics reflect research programmes that typically (but not exclusively) use qualitative research methods, and all have shown increases since the 1970s.

Our discussion of the experimental cliff falls into four main sections. First, we characterize the research programme most closely identified with the experimental method, which, following Cobb (2007), we refer to as “experimental psychology.” Second, drawing on contemporary sources, we consider reasons for this trend. Third, we report a bibliometric analysis of the fate of the experimental psychology programme and conclude that it is in good health, albeit absent from the mathematics education literature. Finally, we draw on Lakatos’s (1978) account to reflect on the possible implications of this absence of experimental work for progress in the field.

Which research programmes in mathematics education are most closely associated with the experimental method? Different authors have used different terminology. Thompson (1982), for instance, contrasted constructivism with what he called “environmentalism.” He characterized this latter research programme as being focused on understanding mathematics learning through the experimental manipulation of the environment rather than, as a constructivist would, the consideration of students’ internal mental constructions. Thus, environmentalists would typically try to see how some manipulation of the environment affected student or teacher behavior. Cobb (2007) offered a similar characterization but referred to the research programme as “experimental psychology,” contrasting it with “cognitive psychology,” a term he used to describe what we have been calling constructivism. Clements and Ellerton (1996) used different terminology again, referring to the research programme as “the ‘scientific’ approach,” and Kilpatrick (1992), following Begle (1969), used the term “experimental science.”

Here we refer to this research programme using the term “experimental psychology” and characterize it as having the aim of testing or generating theories of human behavior, typically by manipulating experimental stimuli in ways designed to reveal underlying psychological processes or associations (e.g., Mook, 1983). These processes could be cognitive in nature, social in nature, developmental in nature, or simply behavioral. Because of the large range of psychological processes that can be and are studied using experimental methods, we believe that

Cobb's (2007) use of the term "cognitive psychology" to refer to constructivism is misleading. Most cognitive psychologists would consider their research area to be the subset of experimental psychology that focuses on understanding internal cognitive processes rather than other types of behavior. Support for this assertion comes from the observation that journals such as the *Quarterly Journal of Experimental Psychology* or the *Journal of Experimental Psychology: General* publish many cognitive studies (and from the observation that journals such as *Cognition* or *Cognitive Psychology* publish few constructivist studies).

If we understand the experimental cliff in mathematics education to be a dramatic move away from the experimental psychology research programme and toward the constructivist research programme, what was behind it? Kilpatrick (1992) pointed out that some researchers at the time felt that mathematics education research was not successfully influencing educational practice and suggested that this encouraged the exploration of alternatives. Clements and Ellerton (1996) attributed the change to two main factors. First, they argued that doubts had begun to form regarding the validity of null hypothesis significance testing, a critical part of the experimental method (e.g., Carver, 1978; Menon, 1993).³ Second, and what is more important, they suggested that there was a reaction against the dominance of the experimental psychology research programme and a desire for more diverse approaches to be permitted. Clements and Ellerton went as far as describing experimental psychology as "a straitjacket" (p. 74) from which mathematics education research had to emerge.

A great deal of discussion of this latter point took place in the pages of *JRME* during 1977–1979. Following a series of letters to *JRME* that debated the merits of experimental and nonexperimental research (Aiken, 1977; Fennema, 1978; Rappaport, 1977; Scott, 1977; Shaughnessy, 1978; Steffe, 1978; Wheeler, 1978), Lester and Kerr (1979) wrote a position statement that explicitly argued for greater methodological diversity. In a series of recommendations, they suggested that mathematics education doctoral students should be trained in both experimental methods and nonexperimental methods, such as clinical interviewing and the analysis of verbal protocols required to conduct constructivist research; that *JRME* must be willing to publish both experimental and nonexperimental research; and that researchers should openly debate their recommendation for an increase in the diversity of research programmes accepted in the field.

The experimental cliff shown in Figure 5 indicates that Lester and Kerr's (1979) call for nonexperimental work to become more acceptable was remarkably successful but at the cost of their call for methodological diversity. Rather than the constructivist

³ From a contemporary perspective, several of the critiques offered by these authors seem misplaced. For instance, Clements and Ellerton (1996) were incorrect to state that "almost any study can be made to show significant results if a sufficiently large sample is used" (p. 80): This would only be the case for designs in which participants are not randomly allocated into groups. Furthermore, recent research on the relationship between p values and Bayes factors indicates that it is an exaggeration to say that p values reveal "nothing" about the truth of the null hypothesis (e.g., Held, 2010; Sellke, Bayarri, & Berger, 2001).

research programme becoming an accepted approach alongside the experimental psychology programme, as Lester and Kerr were advocating, it replaced it.

However, we do not believe that the experimental psychology research programme degenerated to the point where it was abandoned entirely. Our impression is that there is a thriving research community conducting psychological research on mathematics learning using experimental methods; indeed, we have contributed to some of this work (e.g., Alcock et al., 2016). One indication of the current health of the experimental psychology programme, as it pertains to the learning of mathematics, is the recent foundation of the *Journal of Numerical Cognition*, the official journal of the Mathematical Cognition and Learning Society (Towse, 2015).

To empirically test our hypothesis that the experimental psychology research programme has continued to actively investigate mathematics learning, albeit only outside of mathematics education journals, we conducted an analysis of publications using the Scopus bibliographic database. Our goal was to identify trends in the quantity of published research on mathematics learning from the experimental psychology research programme outside of the mathematics education literature.

We searched for all psychology articles that contained the words *experiment* or *experimental* in any field and that contained *learning* and either *mathematical* or *numerical* in the abstract or title. Because Scopus includes the mathematics education journals *Mathematical Thinking and Learning* and *Journal of Mathematical Behavior* in its psychology category, we excluded these from our search.⁴ We calculated the number of articles per year that met these criteria alongside the total number of experimental psychology articles published each year (psychology articles that included *experiment* or *experimental* in any field, excluding the two mathematics education journals that Scopus includes in the psychology category). This allowed us to estimate the proportion of experimental psychology articles published per year that focused on mathematics learning.⁵ These proportions are plotted in Figure 7 (right axis), alongside the “experimental methods” topic proportions by year collapsed across both *ESM* and *JRME* (left axis). We have plotted the number of experimental psychology papers on mathematics learning as a percentage of the total number of experimental psychology papers published, which yields a relatively low percentage, but in absolute terms, the number of papers that met our (fairly restricted) search term was relatively large (89 in 2015 compared with the 72 articles published in *ESM* and *JRME* in that year).

⁴ Our formal Scopus search term was ((TITLE-ABS(mathematical OR numerical) AND TITLE-ABS(learning) AND SUBJAREA(psyc)) AND ALL(experimental OR experiment) AND (EXCLUDE(EXACTSRCTITLE,"Journal Of Mathematical Behavior") OR EXCLUDE(EXACTSRCTITLE,"Mathematical Thinking And Learning"))).

⁵ The results reported in this section are robust to minor variations in these search terms. For instance, a similar pattern is found when searching for articles from sources that have the words *psychology* or *cognitive* in their title rather than the psychology subject area; when removing the *experiment* or *experimental* criterion; or when searching for articles with *education* rather than *learning* in the title, abstract, keywords, and so on. Similarly, consistent results are obtained if *Mathematical Thinking and Learning* and the *Journal of Mathematical Behavior* are included in the psychology category.

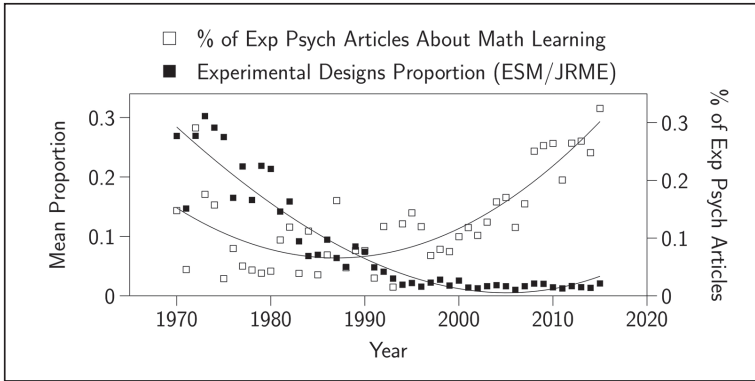


Figure 7. The experimental cliff. Trends over time for (a) the proportion of words from the “experimental methods” topic (collapsed across *ESM* and *JRME*), shown on the left axis, and (b) estimates of the proportion of experimental psychology articles that focus on mathematics learning, shown on the right axis. Lines are cubics of best fit.

Figure 7 indicates that over the same period in which the proportion of experimental psychology work published in *ESM* and *JRME* has been falling, the proportion of experimental work in general psychology journals that focuses on mathematics learning has been rising. Given this, perhaps the term “experimental migration” might be a more accurate characterization of what we have called the experimental cliff. In sum, there appears to be a great deal of research being conducted on mathematics learning within the experimental psychology research programme, but little of this work is being published in the two leading mathematics education research journals. From the perspective of Lakatos’s (1978) methodology of scientific research programmes, we argue that this situation is highly suboptimal.

As we have discussed, according to Lakatos’s account, competition between programmes is an effective method of driving scientific progress. This commonly happens when anomalies are identified that challenge the hard cores of established research programmes, and the programmes react by incorporating rescue hypotheses into their protective belts. Evaluating these rescue hypotheses gives one way for researchers to decide whether a programme is progressing or degenerating. Lakatos (1978) and Feyerabend (1985, 1993) both believed that the most effective challenges to research programmes come from the perspective of a rival programme. If this belief is correct,⁶ then knowledge of multiple programmes would be helpful for progression in the field. For instance, the anomalies that led to the social challenge to constructivism (Carraher, 1988; Lave, 1988; Walkerdine, 1988) were highly visible in the mathematics education community. Indeed, Carraher’s (1988) work was presented as a plenary at the annual PME conference

⁶ Recall that Kuhn (1962) felt that different research programmes are incommensurate, so he would have rejected this argument. In contrast, Lakatos (1978) argued that scientific progress is more than “a matter for mob psychology” (p. 91).

in 1988. If this work had remained isolated from mathematics education researchers, perhaps few would have appreciated the seriousness of the anomaly posed to the constructivist programme.

However, assuming that researchers primarily read articles from journals in which they also publish, then publishing work from different research programmes in disjoint sets of outlets is likely to be a substantial barrier to the development of this knowledge. Specifically, if knowledge of the experimental psychology research programme is to be developed among mathematics education researchers, then either some experimental psychology research must be published in mathematics education journals, or mathematics education researchers must make a concerted effort to read journals from outside the discipline. Equally, it is hard to see how experimental psychology researchers will develop knowledge of, say, the sociocultural programme unless research from this tradition is published in the journals that they primarily read or unless they widen the range of journals that they typically read.

What kinds of anomalies might be identified if there were more interactions between the research programmes? Here we give two indicative examples. Some work in the experimental psychology research programme has suggested that instructional approaches termed *direct* or *explicit instruction* may be more effective for student attainment than those relying on guided discovery and collaborative group learning (Gersten et al., 2009; Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004). On the face of it, this would appear to be an anomaly for the sociocultural research programme. If teacher-centered instructional approaches, sometimes involving scripted explanations and an emphasis on mastery of skills through sustained individual repetitive practice, are successful in building both procedural and conceptual mathematical knowledge, then how can it be that knowledge is constituted primarily in social groups rather than in individuals? To be clear, we do not suggest that this is an insurmountable anomaly for the sociocultural research programme, only that it is one that should be accommodated, if possible, in its protective belt.

As an example in the opposite direction, consider recent sociocultural work on micro-identities. Many sociocultural researchers have noticed that the mathematical identities students adopt can substantially influence how they learn mathematics (e.g., Martin, 2000; Sfard & Prusak, 2005). Traditionally, identities have been conceived as being relatively stable over time. However, Wood (2013) recently illustrated how students' identities may change over extremely short periods in response to relatively minor changes in context. She used the terms "micro-identity" and "macro-identity" to distinguish between identities that show moment-to-moment changes and those that are more stable over time. Importantly, Wood reported a case study of one student who demonstrated at least three different mathematical micro-identities during the course of a single lesson, and Wood argued that these developing identities influenced the quality of the student's learning. This observation poses an interesting challenge to the experimental psychology research programme in which researchers often assume that

the factors that affect behavior within a study are due to relatively stable individual traits of the participant, to the experimental stimuli, or to independently and identically distributed random noise. Wood's study implies that actions that appear innocuous from the researchers' perspective may systematically influence behavior in a way that cannot be assumed to be identical across participants or conditions, an observation that deserves attention.

To be clear, the benefits of more between-programme engagement is symmetric. We are not merely arguing that mathematics education would benefit from some sociocultural, didactical theory and semiotic researchers being exposed to experimental psychology research but also that experimental psychology researchers would benefit from exposure to the research programmes common in the modern mathematics education literature. If Lakatos's (1978) analysis of academic disciplines is correct, then exposure to multiple research programmes aids the effective competition between programmes and, therefore, progresses the discipline.

Theoretical Remarks

We conclude the article by discussing two main issues related to our empirical and theoretical analysis. First, we clarify our results by considering precisely what our topic-modeling approach allows us to conclude about a research programme's trajectory. Second, we discuss whether Feyerabend's (1981, 1993) criticisms of Lakatos's (1978) methodology of scientific research programmes apply to our own work.

Using Topic Modeling to Identify Programme Shifts

Our analysis used a topic-modeling approach to identify trends in the mathematics education literature with a particular focus on the research programmes that researchers have adopted over the last 5 decades. The approach works by analyzing the occurrence of words within papers. Although this method seems to have been successful at identifying topics that represent research programmes, the presence of a particular word associated with a particular research programme is not sufficient to conclude that the programme is progressing or degenerating. The same words are used to denote ideas from the hard core of a programme both during progressing periods and degenerating periods.

The implication of this observation is that we cannot directly conclude that a programme has degenerated in Lakatos's (1978) sense, only that interest in it has declined over time. Specifically, although degenerating research programmes will decline in prominence over time, we cannot conclude that if a research programme has declined in prominence, then it must have been degenerating. Programmes may lose (or gain) interest for reasons unrelated to whether they are degenerating (or progressing). Indeed, we have suggested that the experimental psychology programme did not degenerate but rather migrated. This raises the possibility that a similar phenomenon could account for other trends observed in our data. For instance, perhaps constructivist mathematics education research outputs have migrated to different (or new) journals since the social turn. Similarly, perhaps educational research on Euclidean geometry has migrated since the 1970s. Although

we have no reason to believe that these hypotheses are plausible, further research would be required to test them directly.

Feyerabend's Critique of Lakatos

Throughout this article, we have framed our analysis of trends in the mathematics education literature using Lakatos's (1978) methodology of scientific research programmes. However, Lakatos's account is not uncontested. The main critique comes from Feyerabend's (1981, 1993) "anarchist theory of knowledge." Feyerabend's criticisms of Lakatos fall into two categories. First, he argued that Lakatos's use of historical rational reconstructions as evidence was arbitrary in several different ways. He felt that Lakatos failed to justify (a) his decision only to reconstruct scientific episodes from a relatively restricted historical period (the last 200 years), (b) his choice of the specific episodes to reconstruct, and (c) the actual methods by which he conducted his reconstruction. Second, Feyerabend suggested that there were no reasons to support Lakatos's belief that his methodology of scientific research programmes could show that scientists behaved rationally in their choice of research programme. He wrote that "we can only say that one programme was *accepted* while the other *receded into the background*; we cannot add that the acceptance was rational or that a rational development took place" (Feyerabend, 1981, p. 220).

Is our use of Lakatos's (1978) methodology of scientific research programmes vulnerable to Feyerabend's (1981, 1993) criticisms? With respect to the first, our inclusive approach of analyzing every article published in *ESM* and *JRME* since they began publishing seems to provide a defense against the charge of arbitrariness. Whereas Lakatos chose specific historical incidents to analyze, we considered every article published by these journals, widely considered to be the two leading journals in the field. In view of *ESM* and *JRME*'s status, it seems probable that most of the field's important trends are represented in the articles that they publish. Nevertheless, we are vulnerable to at least one charge of arbitrariness, that our sample included only journals that publish in English. This important limitation should be borne in mind when considering our findings.

Feyerabend's (1981, 1993) second criticism does not seem to apply to our use of Lakatos's (1978) work. Whereas Lakatos was keen to show that scientists are behaving rationally when they choose to move from a progressing to a degenerating research programme, we do not require this assumption. Our goal was not to argue that those who took part in the social turn were being either rational or irrational when they did so; rather, we have used Lakatos's methodology of scientific research programmes as a descriptive—not normative—framework. In that sense, Feyerabend's main criticisms of Lakatos do not apply to our analysis. However, our suggestion that mathematics education would benefit from greater interaction between its research programmes does have a normative flavor. But Feyerabend (1993) advanced a similar argument. He wrote that

Some of the most important formal properties of a theory are found by contrast, and not by analysis. A scientist who wishes to maximize the empirical content of the views he

holds and who wants to understand them as clearly as he possibly can must therefore introduce other views; that is, he must adopt a *pluralistic methodology*. (p. 21)

In sum, we have argued that mathematics education would benefit from greater interaction between the experimental psychology and sociocultural research programmes. This is a proposal that Feyerabend would have wholeheartedly endorsed.

References

- Adi, H., Karplus, R., & Lawson, A. E. (1980). Conditional logic abilities on the four-card problem: Assessment of behavioral and reasoning performances. *Educational Studies in Mathematics*, 11(4), 479–496. doi:10.1007/BF00231218
- Aiken, L. R. (1977). Letter to the editor. *Journal for Research in Mathematics Education*, 8(3), 240. doi:10.2307/748529
- Alcock, L., Ansari, D., Batchelor, S., Bisson, M.-J., De Smedt, B., Gilmore, C., . . . Weber, K. (2016). Challenges in mathematical cognition: A collaboratively-derived research agenda. *Journal of Numerical Cognition*, 2(1), 20–41. doi:10.5964/jnc.v2i1.10
- Banerjee, R., & Subramaniam, K. (2012). Evolution of a teaching approach for beginning algebra. *Educational Studies in Mathematics*, 80(3), 351–367. doi:10.1007/s10649-011-9353-y
- Barbé, J., Bosch, M., Espinoza, L., & Gascón, J. (2005). Didactic restrictions on the teacher's practice: The case of limits of functions in Spanish high schools. *Educational Studies in Mathematics*, 59(1–3), 235–268. doi:10.1007/s10649-005-5889-z
- Baroody, A. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education*, 18(2), 141–157. doi:10.2307/749248
- Battista, M. (1999). Fifth graders' enumeration of cubes in 3D arrays: Conceptual progress in an inquiry-based classroom. *Journal for Research in Mathematics Education*, 30(4), 417–448. doi:10.2307/749708
- Begle, E. G. (1969). The role of research in the improvement of mathematics education. *Educational Studies in Mathematics*, 2(2–3), 232–244. doi:10.1007/BF00303460
- Behr, M., & Eastman, P. (1975). Interactions between structure-of-intellect factors and two methods of presenting concepts of modulus seven arithmetic: A follow-up and refinement study. *Journal for Research in Mathematics Education*, 6(3), 150–157. doi:10.2307/748756
- Bikner-Ahsbabs, A., & Prediger, S. (2010). Networking of theories—An approach for exploiting the diversity of theoretical approaches. In B. Sriraman & L. English (Eds.), *Theories of Mathematics Education* (pp. 483–506). Berlin, Germany: Springer-Verlag. doi:10.1007/978-3-642-00742-2_46
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Boero, P., & Szendrei, J. R. (1998). Research and results in mathematics education: Some contradictory aspects. In A. Sierpiska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (pp. 197–212). Dordrecht, the Netherlands: Kluwer.
- Brookhart, S. M., Andolina, M., Zusa, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics*, 57(2), 213–227. doi:10.1023/B:EDUC.0000049293.55249.d4
- Brousseau, G. (2006). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990*. Dordrecht, the Netherlands: Springer.
- Bruner, J. (1996). *Celebrating diversity: Piaget and Vygotsky*. Presented at the 'Vygotsky–Piaget Conference' at the Second Congress of Sociocultural Research, Geneva, Switzerland.
- Bush, W. (1986). Preservice teachers' sources of decisions in teaching secondary mathematics. *Journal for Research in Mathematics Education*, 17(1), 21–30. doi:10.2307/749316
- Carraher, T. N. (1988). Street mathematics and school mathematics. In A. Borbás (Ed.), *Proceedings of the 12th International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 1–23). Veszprém, Hungary: International Group for the Psychology of Mathematics Education.

- Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48(3), 378–399. doi:10.17763/haer.48.3.t490261645281841
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276. doi:10.1207/s15327906mbr0102_10
- Chassapis, D. (2002). Social groups in mathematics education research: An investigation into mathematics education-related research articles published from 1971 to 2000. In P. Valero & O. Skovsmose (Eds.) *Proceedings of the 3rd International Mathematics Education and Society Conference* (pp. 229–237). Copenhagen, Denmark: Centre for Research in Learning Mathematics.
- Chevallard, Y. (1999). Pratiques enseignantes en théorie anthropologique. *Recherches en Didactique des Mathématiques*, 19(2), 221–266.
- Clements, M. A., & Ellerton, N. F. (1996). *Mathematics education research: Past, present and future*. Bangkok, Thailand: UNESCO.
- Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 3–38). Charlotte, NC: Information Age.
- Cobb, P., Yackel, E., & Wood, T. (1988). Curriculum and teacher development: Psychological and anthropological perspectives. In E. Fennema, T. P. Carpenter, & S. J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 83–120). Madison, WI: Wisconsin Centre for Educational Research.
- Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, 23(1), 2–33. doi:10.2307/749161
- Coxeter, H. S. M. (1971). Inverse geometry. *Educational Studies in Mathematics*, 3(3–4), 310–321. doi:10.1007/BF00302300
- Davis, P. J., & Hersh, R. (1980). *The mathematical experience*. Harmondsworth, United Kingdom: Penguin.
- Davis, R. B. (1990). Discovery learning and constructivism. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 93–106). Reston, VA: National Council of Teachers of Mathematics.
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York, NY: Palgrave Macmillan.
- Dreyfus, T. (2006). Linking theories in mathematics education. In A. Simpson (Ed.), *Retirement as process and concept: A festschrift for Eddie Gray and David Tall* (pp. 77–82). Durham, United Kingdom: Durham University.
- Ernest, P. (1991). *The philosophy of mathematics education*. London, United Kingdom: Routledge Falmer.
- Ethington, C. (1992). Gender differences in a psychological model of mathematics achievement. *Journal for Research in Mathematics Education*, 23(2), 166–181. doi:10.2307/749499
- Fennema, E. (1978). Letter to the editor. *Journal for Research in Mathematics Education*, 9(5), 398. doi:10.2307/748780
- Feyerabend, P. (1981). *Problems of empiricism: Philosophical papers: Volume 2*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9781139171588
- Feyerabend, P. K. (1985). *Realism, rationalism and scientific method: Volume 1: Philosophical papers* (Vol. 1). Cambridge, United Kingdom: Cambridge University Press.
- Feyerabend, P. (1993). *Against method: Outline of an anarchist theory of knowledge* (3rd ed.). London, United Kingdom: Verso.
- Frejth, P. (2013). Modes of modelling assessment—A literature review. *Educational Studies in Mathematics*, 84(3), 413–438. doi:10.1007/s10649-013-9491-5
- Gates, P., & Jorgensen (Zevenbergen), R. (2015). Mapping the field and documenting the contribution. In P. Gates & R. Jorgensen (Zevenbergen) (Eds.), *Shifts in the field of mathematics education*, (pp. 1–14). London, United Kingdom: Springer.
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, 79(3), 1202–1242. doi:10.3102/0034654309334431

- Gilbert, J. K., & Swift, D. J. (1985). Towards a Lakatosian analysis of the Piagetian and alternative conceptions research programs. *Science Education*, 69(5), 681–696. doi:10.1002/sce.3730690510
- Gillies, D. (2007). Lessons from the history and philosophy of science regarding the Research Assessment Exercise. *Royal Institute of Philosophy Supplement*, 61, 37–73. doi:10.1017/S1358246107000148
- Ginsburg, H. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics*, 1(3), 4–11.
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, 49(2), 193–223. doi:10.1023/A:1016209010120
- Greiffenhagen, C., & Sharrock, W. (2008). School mathematics and its everyday other? Revisiting Lave's 'Cognition in Practice.' *Educational Studies in Mathematics*, 69(1), 1–21. doi:10.1007/s10649-008-9115-7
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. doi:10.1093/pan/mps028
- Gutiérrez, R. (2013). The sociopolitical turn in mathematics education. *Journal for Research in Mathematics Education*, 44(1), 37–68. doi:10.5951/jresmetheduc.44.1.0037
- Hanna, G., & Sidoli, N. (2002). The story of ESM. *Educational Studies in Mathematics*, 50(2), 123–156. doi:10.1023/A:1021162617070
- Held, L. (2010). A nomogram for P values. *BMC Medical Research Methodology*, 10, 21. doi:10.1186/1471-2288-10-21
- Inglis, M. (2015). Review of "APOS Theory: A framework for research and curriculum development in mathematics education, Arnon et al. (2014)." *International Journal of Research in Undergraduate Mathematics Education*, 1(3), 413–417. doi:10.1007/s40753-015-0015-9
- Jablonka, E., & Bergsten, C. (2010). Commentary on 'Theories of mathematics education: Is plurality a problem?' In B. Sriraman & L. English (Eds.), *Theories of mathematics education: Seeking new frontiers* (pp. 111–117). Dordrecht, the Netherlands: Springer. doi:10.1007/978-3-642-00742-2_12
- Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106. doi:10.1080/21670811.2015.1093271
- Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, 32(2), 101–125. doi:10.1023/A:1002981520728
- Kidron, I., Lenfant, A., Bikner-Ahsbahs, A., Artigue, M., & Dreyfus, T. (2008). Toward networking three theoretical approaches: The case of social interactions. *ZDM Mathematics Education*, 40(2), 247–264. doi:10.1007/s11858-008-0079-y
- Kieran, C. (1994). Doing and seeing things differently: A 25-year retrospective of mathematics education research on learning. *Journal for Research in Mathematics Education*, 25(6), 583–607. doi:10.2307/749574
- Kilpatrick, J. (1992). A history of research in mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 3–38). New York, NY: MacMillan.
- Kilpatrick, J. (2014). History of research in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 267–272). Dordrecht, the Netherlands: Springer. doi:10.1007/978-94-007-4978-8_71
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. doi:10.1207/s15326985ep4102_1
- Kitcher, P. (1983). *The nature of mathematical knowledge*. Oxford, United Kingdom: Oxford University Press.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667. doi:10.1111/j.0956-7976.2004.00737.x
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

- Laborde, C., & Perrin-Glorian, M. J. (2005). Introduction teaching situations as object of research: Empirical studies within theoretical perspectives. *Educational Studies in Mathematics*, 59(1–3), 1–12. doi:10.1007/s10649-005-5761-1
- Lai, K. H., & Loo, S. P. (1992). *State-of-the-art review of research in mathematics education in Malaysia*. Penang, Malaysia: Southeast Asian Research Review and Advisory Group.
- Lakatos, I. (1970). History of science and its rational reconstructions. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association: Vol. 1970* (pp. 91–136). doi:10.1086/psaprobienmeetp.1970.495757
- Lakatos, I. (1976). *Proofs and refutations*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9781139171472
- Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers Volume 1*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9780511621123
- Larvor, B. (1998). *Lakatos: An introduction*. London, United Kingdom: Routledge.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/CBO9780511609268
- Leatham, K. R. (2015). Observations on citation practices in mathematics education research. *Journal for Research in Mathematics Education*, 46(3), 253–269. doi:10.5951/jresmetheduc.46.3.0253
- Leikin, R., & Zaslavsky, O. (1997). Facilitating student interactions in mathematics in a cooperative learning setting. *Journal for Research in Mathematics Education*, 28(3), 331–354. doi:10.2307/749784
- Lerman, S. (1996). Intersubjectivity in mathematics learning: A challenge to the radical constructivist paradigm? *Journal for Research in Mathematics Education*, 27(2), 133–150. doi:10.2307/749597
- Lerman, S. (2000). The social turn in mathematics education research. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 19–44). Westport, CT: Ablex.
- Lerman, S., Xu, G., & Tsatsaroni, A. (2002). Developing theories of mathematics education research: The ESM story. *Educational Studies in Mathematics*, 51(1–2), 23–40. doi:10.1023/A:1022412318413
- Lester, F. K., & Kerr, D. R. (1979). Some ideas about research methodologies in mathematics education. *Journal for Research in Mathematics Education*, 10(3), 228–232. doi:10.2307/748813
- Linton, C. (2004). *From Eudoxus to Einstein: A history of mathematical astronomy*. Cambridge, United Kingdom: Cambridge University Press.
- Lubienski, S. T., & Bowen, A. (2000). Who's counting? A survey of mathematics education research 1982–1998. *Journal for Research in Mathematics Education*, 31(5), 626–633. doi:10.2307/749890
- Martin, D. B. (2000). *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency*. Mahwah, NJ: Erlbaum.
- McCallum, A. K. (2002). *MALLET: MACHine Learning for Language Toolkit*. Retrieved from <http://mallet.cs.umass.edu>
- Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal*, 5(1), 4–18. doi:10.1007/BF03217248
- Miller, L. D. (1992). Teacher benefits from using impromptu writing prompts in algebra classes. *Journal for Research in Mathematics Education*, 23(4), 329–340. doi:10.2307/749309
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4), 379–387. doi:10.1037/0003-066X.38.4.379
- Morgan, C. (2014). Social theory in mathematics education: Guest editorial. *Educational Studies in Mathematics*, 87(2), 123–128. doi:10.1007/s10649-014-9572-0
- Moschkovich, J. (2007). Using two languages when learning mathematics. *Educational Studies in Mathematics*, 64(2), 121–144. doi:10.1007/s10649-005-9005-1
- Mousley, J. A. (2015). Knowledge construction: Individual or social? In P. Gates & R. Jorgensen (Zevenbergen) (Eds.), *Shifts in the field of mathematics education* (pp. 151–170). London, United Kingdom: Springer.
- Nivens, R. A., & Otten, S. (2017). Assessing journal quality in mathematics education. *Journal for Research in Mathematics Education*, 48(4), 348–368. doi:10.5951/jresmetheduc.48.4.0348

- Noddings, N. (1990). Constructivism in mathematics education. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (JRME Monograph No. 4, pp. 7–18). Reston, VA: National Council of Teachers of Mathematics.
- Noelting, G. (1980). The development of proportional reasoning and the ratio concept Part I—Differentiation of stages. *Educational Studies in Mathematics*, 11(2), 217–253. doi:10.1007/BF00304357
- Núñez, R. E., Edwards, L., D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, 39(1–3), 45–65. doi:10.1023/A:1003759711966
- Pais, A. (2015). Symbolising the real of mathematics education. *Educational Studies in Mathematics*, 89(3), 375–391. doi:10.1007/s10649-015-9602-6
- Pais, A., & Valero, P. (2012). Researching research: Mathematics education in the political. *Educational Studies in Mathematics*, 80(1–2), 9–24. doi:10.1007/s10649-012-9399-5
- Pais, A., & Valero, P. (2014). Whither social theory? *Educational Studies in Mathematics*, 87(2), 241–248. doi:10.1007/s10649-014-9573-z
- Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press. doi:10.1037/11494-000
- Popper, K. (1959). *The logic of scientific discovery*. Abingdon, United Kingdom: Routledge.
- Post, T., Medhanie, A., Harwell, M., Norman, K., Dupuis, D., Muchlinski, T., . . . Monson, D. (2010). The impact of prior mathematics achievement on the relationship between high school mathematics curricula and postsecondary mathematics performance, course-taking, and persistence. *Journal for Research in Mathematics Education*, 41(3), 274–308.
- Presmeg, N., Radford, L., Roth, W.-M., & Kadunz, G. (2016). Semiotics in theory and practice in mathematics education. In N. Presmeg, L. Radford, W.-M. Roth, & G. Kadunz (Eds.), *Semiotics in mathematics education* (pp. 5–29). London, United Kingdom: Springer. doi:10.1007/978-3-319-31370-2
- Radford, L. (2003). Gestures, speech, and the sprouting of signs: A semiotic-cultural approach to students' types of generalization. *Mathematical Thinking and Learning*, 5(1), 37–70. doi:10.1207/S15327833MTL0501_02
- Rappaport, D. (1977). Letter to the editor. *Journal for Research in Mathematics Education*, 8(2), 160–161. doi:10.2307/748550
- Reid, D., & Knipping, C. (2010). *Proof in mathematics education: Research, learning and teaching*. Rotterdam, the Netherlands: Sense.
- Roberts, D. (2001). Letter to the editor. *Journal for Research in Mathematics Education*, 32(2), 227–227. doi:10.2307/749675
- Roh, K. H. (2010). An empirical study of students' understanding of a logical structure in the definition of limit via the ϵ -strip activity. *Educational Studies in Mathematics*, 73(3), 263–279. doi:10.1007/s10649-009-9210-4
- Roth, W.-M. (2012). Tracking the origins of signs in mathematical activity: A material phenomenological approach. In M. Bockarova, M. Danesi, & R. Núñez (Eds.), *Cognitive science and interdisciplinary approaches to mathematical cognition* (pp. 209–247). Munich, Germany: Lincom Europa.
- Saenz-Ludlow, A. (1994). Michael's fraction schemes. *Journal for Research in Mathematics Education*, 25(1), 50–85. doi:10.2307/749292
- Sakonidis, C. (2015). Tracing the advances in the field of mathematics education. In P. Gates & R. Jorgensen (Zevenbergen) (Eds.), *Shifts in the field of mathematics education* (pp. 59–74). London, United Kingdom: Springer. doi:10.1007/978-981-287-179-4_5
- Schoenfeld, A. H. (2016). Research in mathematics education. *Review of Research in Education*, 40(1), 497–528. doi:10.3102/0091732X16658650
- Scott, D. E. (1977). Letter to the editor. *Journal for Research in Mathematics Education*, 8(3), 239. doi:10.2307/748529
- Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71. doi:10.1198/000313001300339950
- Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, 34(4), 14–22. doi:10.3102/0013189X034004014

- Shaughnessy, J. M. (1978). Letter to the editor. *Journal for Research in Mathematics Education*, 9(5), 399–400. doi:10.2307/748780
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. doi:10.3102/0013189X015002004
- Sierpinska, A., & Kilpatrick, J. (Eds.). (1998). *Mathematics education as a research domain: A search for identity: An ICMI study*. Dordrecht, the Netherlands: Kluwer.
- Silver, E. (1981). Recall of mathematical problem information: Solving related problems. *Journal for Research in Mathematics Education*, 12(1), 54–64. doi:10.2307/748658
- Silver, E. (2003). Attention deficit disorder? *Journal for Research in Mathematics Education*, 34(1), 2–3.
- Simon, M. A. (2009). Amidst multiple theories of learning in mathematics education. *Journal for Research in Mathematics Education*, 40(5), 477–490.
- Sriraman, B., & English, L. (2005). Theories of mathematics education: A global survey of theoretical frameworks/trends in mathematics education research. *ZDM Mathematics Education*, 37(6), 450–456. doi:10.1007/BF02655853
- Steffe, L. P. (1978). Letter to the editor. *Journal for Research in Mathematics Education*, 9(5), 398–399. doi:10.2307/748780
- Steffe, L., & Kieren, T. (1994). Radical constructivism and mathematics education. *Journal for Research in Mathematics Education*, 25(6), 711–733. doi:10.2307/749582
- Steffe, L. P., & Thompson, P. W. (2000). Interaction or intersubjectivity? A reply to Lerman. *Journal for Research in Mathematics Education*, 31(2), 191–209. doi:10.2307/749751
- Steffe, L. P., & Ulrich, C. (2014). Constructivist teaching experiment. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 102–109). London, United Kingdom: Springer. doi:10.1007/978-94-007-4978-8_32
- Steiner, H. G., & Kaufman, B. A. (1969). Checker games in operational systems as media for an inductive approach to teaching algebra. *Educational Studies in Mathematics*, 1(4), 445–483. doi:10.1007/BF02396318
- Swanson, D., Schwartz, R., Ginsburg, H., & Kossan, N. (1981). The clinical interview: Validity, reliability and diagnosis. *For the Learning of Mathematics*, 2(2), 31–38.
- Taber, K. S. (2007). *Classroom-based research and evidence-based practice: An introduction*. London, United Kingdom: Sage. doi:10.4135/9781849208734
- 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. doi:10.1007/BF00305619
- Thompson, P. W. (1982). Were lions to speak, we wouldn't understand. *Journal of Mathematical Behavior*, 3(2), 147–165.
- Thompson, P. W. (2014). Constructivism in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 96–102). London, United Kingdom: Springer. doi:10.1007/978-94-007-4978-8_31
- Toerner, G., & Arzarello, F. (2012). Grading mathematics education research journals. *EMS Newsletter*, 86, 52–54.
- Toumasis, C. (1990). The epos of Euclidean geometry in Greek secondary education (1836–1985): Pressure for change and resistance. *Educational Studies in Mathematics*, 21(6), 491–508. doi:10.1007/BF00315941
- Towse, J. (2015). Finding a voice for numerical cognition. *Journal of Numerical Cognition*, 1(1), 1–6. doi:10.5964/jnc.v1i1.6
- Tsatsaroni, A., Lerman, S., & Xu, G.-R. (2003, June). *A sociological description of changes in the intellectual field of mathematics education research: Implications for the identities of academics*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. Retrieved from <https://eric.ed.gov/?id=ED482512>
- U.S. Department of Education. (2002). *U.S. Department of Education strategic plan 2002–2007*. Retrieved from <https://www2.ed.gov/about/reports/strat/plan2002-07/plan.doc>

- Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2013). Brief report. Educated adults are still affected by intuitions about the effect of arithmetical operations: Evidence from a reaction-time study. *Educational Studies in Mathematics*, 82(2), 323–330. doi:10.1007/s10649-012-9432-8
- von Glasersfeld, E. (1990). An exposition of constructivism: Why some like it radical. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (JRME Monograph No. 4, pp. 19–30). Reston, VA: National Council of Teachers of Mathematics.
- von Glasersfeld, E. (1991). Knowing without metaphysics: Aspects of the radical constructivist position. In F. Steier (Ed.), *Research and reflexivity* (pp. 12–29). London, United Kingdom: Sage.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1986). *Thought and language* (A. Kozulin, Trans., Ed.). Cambridge, MA: MIT Press.
- Wagner, D. (2015). A speech act in mathematics education—The social turn. In P. Gates & R. Jorgensen (Zevenbergen) (Eds.), *Shifts in the field of mathematics education* (pp. 75–90). London, United Kingdom: Springer.
- Walkerdine, V. (1988). *The mastery of reason*. London, United Kingdom: Routledge.
- Walkerdine, V. (1989). *Counting girls out*. London, United Kingdom: Virago.
- Warren, E., & English, L. (1995). Facility with plane shapes: A multifaceted skill. *Educational Studies in Mathematics*, 28(4), 365–383. doi:10.1007/BF01274079
- Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, 39(4), 431–459.
- Wheeler, D. (1978). Letter to the editor. *Journal for Research in Mathematics Education*, 9(3), 240. doi:10.2307/749003
- Williams, S. (2007). The story's the thing. *Journal for Research in Mathematics Education*, 38(4), 342–343.
- Williams, S. R., & Leatham, K. R. (2017). Journal quality in mathematics education. *Journal for Research in Mathematics Education*, 48(4), 369–396. doi:10.5951/jresmetheduc.48.4.0369
- Wilson, B. J. (1978). Change in mathematics education since the late 1950's—Ideas and realisation: West Indies. *Educational Studies in Mathematics*, 9(3), 355–379. doi:10.1007/BF00241037
- Wilson, J. W. (1994). Guest editorial. *Journal for Research in Mathematics Education*, 25(1), 2–3.
- Wittmann, E. (1973). The concept of grouping in Jean Piaget's psychology—Formalization and applications. *Educational Studies in Mathematics*, 5(2), 125–146. doi:10.1007/BF01421021
- Wood, M. B. (2013). Mathematical micro-identities: Moment-to-moment positioning and learning in a fourth-grade classroom. *Journal for Research in Mathematics Education*, 44(5), 775–808. doi:10.5951/jresmetheduc.44.5.0775
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477. doi:10.2307/749877
- Yoon, C., Thomas, M. O., & Dreyfus, T. (2011). Grounded blends and mathematical gesture spaces: Developing mathematical understandings via gestures. *Educational Studies in Mathematics*, 78(3), 371–393. doi:10.1007/s10649-011-9329-y

Authors

Matthew Inglis, Mathematics Education Centre, Loughborough University, Schofield Building, Loughborough, Leicestershire, LE11 3TU, United Kingdom; m.j.inglis@lboro.ac.uk

Colin Foster, School of Education, University of Leicester, 21 University Road, Leicester LE1 7RF, United Kingdom; colin.foster@leicester.ac.uk

Submitted April 29, 2017

Accepted September 6, 2017