Journal of Educational Research in Mathematics
2020, Special Issue, 001 ~ 013

Proof in Mathematics Education, 1980-2020: An Overview

Gila Hanna*†, Christine Knipping**

*Professor, University of Toronto, Canada
**Professor, Universität Bremen, Germany

ABSTRACT

This paper looks at the evolution of ideas on the role of proof in mathematics education from 1980 to 2020, examining in particular the contributions of both theoretical and empirical research to the teaching of mathematical proof. In so doing, it describes some of the major epistemological themes that emerged in the last forty years, primarily in the philosophy of mathematics and in mathematical education, and informed both the mathematics curriculum and research in mathematics education. The paper also discusses selected research studies that shed light on the opportunities and limitations students face when they engage in proof. Finally, it describes briefly a number of developments in proof technology and the potential of automated theorem provers for enhancing the teaching of proof.

Key Words: Proof, Reasoning and proof, Epistemological perspectives, Empirical research, Proof assistants

BACKGROUND: 1950-1980: Emphasis on axiomatic structure – the New Math and the backlash

The successful launch of Sputnik by the USSR in 1957 was an impetus for a re-examination in North America of the mathematics curriculum. Several university-based commissions addressed curriculum reform and produced materials that were disseminated during the 1950s and 1960s (NCTM, 1963). Notably, Max Beberman of the University of Illinois chaired a committee that proposed to reform the high-school mathematics curriculum by introducing the changes that came to be known as the "New Math". Beberman (1958) had identified areas of incompetence among first-year students of mathematics, such as poor computational facility, poor conceptual background, unfamiliarity with proof and its structure (except in geometry), and ignorance of contemporary applications in natural science, social science, and technology. The New Math set out to upgrade students' competencies in mathematics through a curriculum emphasis on set theory, on axiomatic structure, on rigorous proof, and on working with numbers in bases other than 10. Likewise, the School Mathematics Study Group (SMSG), another of the commissions tasked with reforming the mathematics curriculum, prepared a new eleventh-grade algebra text with a stress on structure and proof, arguing that "Proof is emphasized throughout in order that the student may gain some idea of the nature of a valid mathematical argument. ... Accuracy in the statement and use of postulates, definitions, and theorems is emphasized." (SMSG, 1962, p.9).

A backlash against the New Math was not long in coming. Many of the New Math curriculum developments fell by the wayside within a few years of their introduction. Many teachers found the New Math too abstract and too difficult to teach, and that the purely formal approach added next to nothing to clarity. Parents also criticized its approach, because it was too

† corresponding author
Email: gila.hanna@utoronto.ca, knipping@math.uni-bremen.de

Please cite this article as: Hanna, G. & Knipping, C. Proof in mathematics education, 1980-2020: An Overview. *Journal of Educational Research in Mathematics* (2020). https://doi.org/10.29275/jerm.2020.08.sp.1.1

different from the mathematics they were familiar with. In the end the New Math became so contentious that the American composer of humorous songs, Tom Lehrer (1965), who was also a mathematician, wrote a satirical song about it. In a more serious vein, Kline (1973) documented the problems arising from the New Math in his much-discussed book, *Why Johnny can't add: the failure of the New Math.*

THE PERIOD 1980-2000: The ongoing value of proof – epistemological and empirical issues

In psychology, this period saw the decline of behaviorist theories of learning and teaching, based on the simple transmission of information by teachers to passive learners, and the emergence of constructivist theories, based on cognitive approaches to learning centered on the active participation of learners in interpreting, organizing, and storing information. More and more researchers and teachers adopted the key insight that learning is an active rather than a passive process, that while information can be passively received, understanding can only come from the active engagement of the learner in making meaningful connections between prior knowledge and new knowledge (Cobb & Steffe, 1983; Vergnaud, 1987). An examination of the mathematics education research since the early 1980s reveals that all papers and reports followed the lead of psychology and adopted a constructivist theory of learning, albeit in various forms and shades (Gutiérrez & Boero, 2006).

In this period the teaching of mathematics also saw a growing recognition that concepts of mathematical proof have varied widely from one period and place to another, and that standards of rigour in particular have changed markedly over time (Davis & Hersh, 1983; Grabiner, 1986). Mathematics educators did not abandon the accepted definition of a proof as a finite sequence of propositions, each of which is an axiom or follows from preceding propositions by the rules of logical inference, but they acknowledged that in practice proofs can be informal, typically consisting of a mixture of natural language and formulae, and still retain their validity. This reflected the norm in mathematical practice, where such proofs are considered to be rigorous even when they omit routine logical inferences, and where in fact they are valued because they often have the advantage of conveying greater clarity, insight, and understanding.

The recognition that mathematicians themselves make use of informal proof and give priority to clarity and understanding had an influence on the mathematics curriculum. For example, the National Council of Teachers of Mathematics (NCTM) published An Agenda for Action (NCTM 1980) that made eight recommendations for school mathematics. Recommendation 2 notably asserted that “Basic skills in mathematics be defined to encompass more than computational facility”. It included six sub-recommendations, two of which are relevant to reasoning and proof: (2.5) “Teachers should provide ample opportunities for students to learn communication skills in mathematics. They should systematically guide students to read mathematics and to talk about it with clarity” and (2.6) “The higher-order mental processes of logical reasoning, information processing, and decision making should be considered basic to the application of mathematics.” (NCTM, 1980, p. 2).

1. The influence of Polya and Lakatos

The work of Polya and Lakatos had a tremendous influence on researchers’ conceptions of mathematics in general and of proof in particular. Polya (1954, 1981, 1985) maintained that fostering the art of guessing and making significant use of analogy and induction help students learn proving. As he said “Certainly, let us learn proving, but also let us learn guessing” (1954, p. vi). Notably, Polya’s (1985) well-known four steps of problem solving (understand the problem, devise a plan, carry out the plan, and look back) apply quite well to constructing a proof. Indeed, with his stress in the learning of proof on the role of plausible reasoning and problem solving, along with his focus on a heuristic presentation style, Polya came to be regarded as the father of the modern emphasis in mathematics education on both plausible reasoning and problem solving (Stanic & Kilpatrick, 1988).

In his book *Proofs and Refutations* Lakatos (1983) stated that mathematics is a quasi-empirical science: Its methods are similar to those of the empirical sciences, though it is not one itself. He strongly criticized the deductive approach in mathematics, emphasizing informal methods and claiming that mathematical proof consists of an intuitive, inductive or analogical argument always open to criticism. He challenged the value of formalism in mathematics and instead suggested that the body of mathematical truths grows through a process of proof and refutation, in which proofs are fashioned by attempts, critiques, and gradual improvements. His method is based on the following steps: 1) a primitive conjecture, 2) a tentative proof (a thought experiment or argument), 3), global refutation (by a counterexample), 4) a re-examination of the proof (looking for the guilty lemma), 5) a check on other theorems in case the guilty lemma is there. Lakatos also demonstrated, using the Euler example of $V - E + F = 2$, where V, E and F denote the number of vertices, edges and faces respectively for all regular polyhedral, that it is productive to have conjectures and proofs open to negotiation in discussion groups where students and teachers are equal discussants.

Along with a growing interest in the work of Polya and Lakatos and a renewed curriculum emphasis on heuristics, explanation, justification, and proof, during these years before the turn of the century there was a spate of research papers on the teaching and learning of proof at all grade levels. This increased focus on proof spawned several theoretical frameworks and gave rise to many discussions and even to heated debates. Beginning in the early 1980s, at international congresses such as ICME, PME, and CERME, special working groups were formed to explore the role of proof in the curriculum, its relation to other forms of explanation, and how well students were actually learning to prove. Despite their differing approaches, however, the various working groups were united in their acceptance of two premises:

- Proof must be part of any mathematics curriculum that aims, as it should, to reflect mathematics itself and thus the important role of proof within it, and
- The most significant potential contribution of proof in the classroom is in the promotion of mathematical understanding, a role it plays in mathematical theory and practice as well.

2. Selected epistemological perspectives – (some episodes 1980-2000)

The conceptualization of the process of proving was another topic deliberated by researchers in mathematics education at almost every international meeting and in scholarly journals, often under the influence of what educators had seen playing out in the classroom. One influential study was that by Balacheff (1988). He observed 28 secondary-school students (age 13-14) performing a proof task: "Give a way of calculating the number of diagonals of a polygon once the number of vertices is known." He asked the students to work in pairs and instructed them to present a unanimous conclusion. Based on the 14 responses, Balacheff suggested a taxonomy of student proving with four stages: 1) naïve empiricism, 2) crucial experiment, 3) generic example, and 4) thought experiment. He recognized that the transition by students from a pragmatic justification to a conceptual one (i.e., to an acceptable logical proof) is a crucial challenge in the teaching of proof, and that educators need to pay greater attention to its complexity.

Clearly, the influence of Lakatos' model of the process of proving is present in this taxonomy. In a later paper, however, Balacheff concluded that Lakatos' model of negotiation in discussion groups may impede progress in the classroom. He stated: "…in some circumstances social interaction might become an obstacle, when students are eager to succeed, or when they are not able to coordinate their different points of view, or when they are not able to overcome their conflict on a scientific basis" (Balacheff, 1991, p. 188).

One difficulty in the classroom, according to Harel & Sowder (1998), is that students do not necessarily take the term "proof" to mean the standards of mathematical proof accepted by mathematicians and mathematics educators. Initially driven to understand why students' work often falls short of these standards, they suggested a categorization of proof into different "proof schemes" that are deliberatively psychological and student-centered. Harel and Sowder found it useful to employ the distinction between "ascertaining" (aiming to remove a doubt about the truth of a conjecture) and "persuading" (convincing someone else that a conjecture is true). In addition, they proposed a taxonomy of the types of argument that are often thought of as proof, comprising three categories:

1. External conviction scheme that relies on a teacher's authority, or on the superficial form of the argument,
2. Empirical proof scheme based on specific examples, inductive reasoning or on direct measurements in the case of geometric proofs, and
3. Deductive proof scheme where a conjecture's truth is determined by logical deduction.

Harel and Sowder also pointed out "that these schemes are not mutually exclusive; people can simultaneously hold more than one kind of scheme" (p. 244). In a later publication Harel (2008, p. 271) stated that the first two categories are "undesirable ways of thinking" though they have some pedagogical value, because they can help "to generate ideas or to give insights".

Hanna (1983, 1989a) discussed the origins of the emphasis in mathematics on formal proof and more recent views of its role, and also explored the factors at play in the acceptance of a proof by practicing mathematicians, in particular the social process involved. She concluded that the development of mathematics, along with comments offered by mathematicians, suggest that most mathematicians accept a new theorem where some combination of the following factors are present:

1. They understand the theorem, the concepts embodied in it, its logical antecedents, and its implications; and there is nothing to suggest it is not true;
2. The theorem is significant enough to have implications in one or more branches of mathematics (and is thus important and useful enough to warrant detailed study and analysis);
3. The theorem is consistent with the body of accepted mathematical results;
4. The author has an unimpeachable reputation as an expert in the subject matter of the theorem;
5. There is a convincing mathematical argument for it (rigorous or otherwise), of a type they have encountered before (Hanna, 1989a, pp. 21-22).

Given that these factors all rank higher than the mere existence of a rigorous proof, even among mathematicians, Hanna maintains that it is certainly reasonable in the teaching of proof to stress factors such as understanding, explanation, and significance (Hanna, 1989a, 1989b, 1990).

The work of Fischbein (1982, 1987, 1999) contributed a great deal to clarifying the relationship between reasoning in everyday life and mathematical reasoning, which requires a higher degree of rigor. In his view, "… the intuitive structures are essential components of every form of active understanding and of productive thinking. They may constitute the only or the main forms of knowledge if the corresponding analytical structures are lacking (as is the case during the period of intuitive thinking, in Piagetian terminology)." (1982, p.10). In discussing intuitions and schemata in mathematical reasoning (Fischbein, 1999), identified five ways in which everyday reasoning may differ from mathematical reasoning, in process or in result:

1. A statement is accepted intuitively and no proof is requested;
2. A statement is accepted intuitively, but in mathematics it is also formally proved (coincidence between intuitive acceptance and a logically-based conclusion);
3. A statement is not intuitive or self-evident and may be accepted only on the basis of a formal proof;
4. A conflict appears between the intuitive interpretation (solution) of a statement and the formally-based response, and
5. Two conflicting intuitions may appear.

Fischbein (1982) investigated in particular those situations in which a formal derivation of a theorem does not end up being convincing, because it is not associated with a feeling of "it must be so" (p. 11). In his view, it is essential to recognize that the intuitive and analytical forms of knowledge are deeply interrelated. The implication is that mathematics educators should in their teaching be aware of "… the need for a complementary intuitive acceptance of the absolute predictive capacity of a statement which has been formally proved" (p. 18).

3. Empirical research on reasoning and proof: Selected representative studies 1980-2000

In the 1980s and 1990s a large number of studies investigated classrooms in which mathematical reasoning and justification were topics. Most of them focused on assessing the reasoning skills of students and on determining how well they were able to master proving.

Senk (1985) carried out a survey of 2699 students in grades 9, 10, and 11 in thirteen USA public schools, reporting that the students' achievement in proof tasks was poor, even though they had studied proof writing in geometry classes. The survey showed a rather low level of achievement in proof writing. About 25 percent of the students displayed no competence, another 25 percent could only do trivial proofs, and 20 percent could do partial proofs only. The remaining 30 percent mastered proofs, but only ones similar to what they had seen in a standard textbook (p.454). Senk concluded that her data "… illustrate the current, severe mismatch between the intentions of the geometry curriculum in high school and what students actually learn."

Some studies did indicate that elementary-level students were able to engage in mathematical reasoning, develop conjectures and use both inductive and deductive evidence (Maher & Martino, 1996; Yackel & Cobb, 1994). However, Chazan (1993), who interviewed high-school geometry students on their views of empirical evidence and mathematical proof, found that many students could not differentiate between legitimate deductive proof and empirical evidence.

Similarly, Coe and Ruthven (1994) interviewed a group of advanced-level secondary school students aged 16 to 19, and concluded that, although all students were able to state that the function of proof was to establish the truth of a conjecture, few actually used a deductive method. Most students used empirical strategies (based on one or more examples or on a direct measurement), not deductive strategies as was expected. In a large-scale study of 14 and 15 year-old students studying proof in algebra, Healy and Hoyles (2000) also found that students used predominantly empirical arguments in proof construction.

Other researchers looked at the curriculum organization and its capacity to design classroom situations conducive to conveying a coherent notion of mathematical proof. Dreyfus and Hadas (1996) proposed creating situations in which the students can “see” the need for proof. They suggested activities that lead students to a surprising result, one requiring an explanation, and thus to the need for a proof. Hoyles (1997), for example, argued that it is essential to focus on curricular organization, when teaching proof and proving. Blum and Kirsch (1991) advocated a curriculum organization that supported the use of “preformal” proofs, in which students try to build a chain of conclusions from premises that are valid but not stated formally. Students could then proceed from a preformal to a formal proof with the help of explicit discussions with teachers and peers.

THE PERIOD 2000-2020: – Emergent ideas – epistemological and empirical issues

In the field of psychology, the dominant constructivist theories had ascribed the construction of knowledge to the individual student. This period saw an increasing recognition of sociocultural factors in the cognitive development process. A major influence in this direction was the earlier work of the Russian author Vygotsky (1978), who had assigned the greatest importance to cultural activity in helping the individual build knowledge: "first, collective activity, then culture, the ideal, sign or symbol and finally individual consciousness." (Davydov, 1995, p. 16). Although Vygotsky’s writings about the sociocultural context of learning were known outside the Soviet Union as early as the 1970s, mathematics educators began to adopt them only in the late 1990s (Lerman, 2019). What Vygotsky argued was that communication and social exchanges with others are essential for the individual’s mental processes; he went as far as to state that all human behaviour is the result of a social interaction with others, and that concepts appear on the social plane before they are internalized by an individual (Radford & Roth, 2017).

Vygotsky’s sociocultural perspective did not entirely replace the formerly dominant individual constructivist theory, and in fact these two viewpoints came to be seen as complementary. But there were some serious critiques of Vygotsky, notably by Williams (2016). In the context of helping students overcome their alienation from school mathematics, Vygotsky had assigned importance to the distance between the actual development level of learners and their potential level of development when under the guidance of adults or capable peers, a metric which he termed the “zone of proximal development.” Williams did not believe this factor played a role of any significance. Radford & Roth (2017) responded to Williams by pointing out that he had based his criticism on an individualistic conception of learning and had failed to take into account the crucial role Vygotsky had correctly attributed to its social and cultural aspects. The work of Vygotsky clearly continued to have a great influence on researchers in mathematics education into the 2000s, as they investigated in detail the ways in which individual cognitive growth is influenced by the social interaction in small groups of students working together.

In mathematics, as in psychology, many developments during the years 2000-2020 had their roots in research and discussion that took place in the 1990s. For example, mathematicians increasingly challenged the vision of proof presented by Jaffe and Quinn (1993), who had argued that journals should only publish rigorously proven results that provide every detail of the arguments. It is appropriate, in their view, to engage in experimentation and speculation, but it is not acceptable for a mathematician to publish proof sketches or other speculative results. Jaffe and Quinn had cautioned against weakening the standards of proof, and they had objected in particular to Thurston’s published proof of “Haken manifolds” (see Dunfield & Thurston, 2003), describing it, with its hints, as beautiful and insightful but definitely insufficient as a proof.

Fifteen prominent mathematicians (Atiyah et al, 1994) responded to Jaffe and Quinn, some supporting and others opposing their views. Thurston's (1994), response was a long essay titled "On proof and progress in mathematics" (republished in a mathematics education journal FLM, 1995). Thurston (1994) said that "we [mathematicians] are not trying to meet some abstract production quota of definitions, theorems and proofs. The measure of our success is whether what we do enables people to understand and think more clearly and effectively about mathematics" (p 163), and "…we should recognize that the humanly understandable and humanly checkable proofs that we actually do are what is most important to us, and that they are quite different from formal proofs." (p. 172). Thurston's comments were in line with the emphasis mathematics educators had always put on mathematical understanding. They were seen as supporting the classroom use of proofs that foster understanding, and in particular the educational value of those proofs that fulfill both functions, not only persuading students that they are true but also making it clear why they are true.

The focus on understanding in mathematics education was also assisted by the availability and increased affordability of powerful computers and the development of graphic software, which provided teachers with a variety of new teaching materials and suggested new approaches in mathematics teaching. The classroom use of dynamic geometry software (DGS) in particular gave an impetus to mathematics exploration and made it much easier to suggest and test conjectures. DGS, graphic calculators, and computational tools all came to have a profound effect on the development of students' skills, not only in exploring geometric figures but also in modelling, hypothesizing, validating, and constructing proofs. When in 2000 the National Council of Teachers of Mathematics (NCTM) in the United States published their statement of basic principles and standards for the modern teaching of mathematics (NCTM, 2000), it placed the use of technology among its six "basic precepts that are fundamental to a high-quality mathematics education". It recommended the integration of technology into teaching because "Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning" (NCTM, 2000, Executive summary, p. 2).

1. Selected epistemological perspectives – (some episodes 2000-2020)

The early years of this century saw an increased focus on the use of proof in the mathematics classroom, but at the same time made clear that there are differing interpretations of the aims, rationale and basis for teaching proof. Each of the interpretations focuses on different issues. Each emphasizes particular aspects of what is important to convey to students about proof and proving and what features of proof will contribute to the enhancing of learning. The three key interpretations presented here perceive the educational usefulness of mathematical proof from the perspective of theoretical physics, from that of cognitive development, and from that of adherence to the explicit principles of formal logic.

The perspective of theoretical physics is illustrated by the work of Jahnke (2007, 2010), who saw the "old" didactic tradition as one that holds an outmoded epistemological view of the role of mathematical axioms and proofs. He advocated introducing to the classroom a "new" way of discussing mathematical proof (while preserving elements of the old), an approach that would encourage students to explore how mathematical propositions are often mirrored by laws of physics and thus by the behaviour of physical objects.

In his chapter "The conjoint origin of proof and theoretical physics" Jahnke (2010) further points out the empirical origins of much of mathematics, expanding on his view that "Today, there is a general consensus that the axioms of mathematics are not absolute truths that can be sanctioned by intuition: rather, they are propositions on which people have agreed." (p. 28). As well, Jahnke asserts, following mainly A. Szabó's path-breaking historical studies of the concept of proof, that this view is not really new, since mathematical proof has always been understood primarily as a mode of rational discourse, and not only as a means to establish the validity of "dogmatic' claims. Its main role was to defend plausible presuppositions and to organize mathematical knowledge in an axiomatic-deductive way.

Jahnke goes on to recommend that teachers encourage students of proof to invent and test hypotheses through empirical investigations, in effect asking them to behave like theoretical physicists (2007, p. 82). As Jahnke (2007) sees it, "The epistemological motivation of proof is not to be founded on the idea that proofs in contrast to measurement provide absolute certainty, but on the idea that proofs open new and more complex possibilities of empirical corroboration." (2007, p. 83). He proceeded to test his approach in a classroom situation by carrying out an empirical study with grades eight and nine students (Jahnke & Wambach, 2013). The authors concluded that introducing the concept of 'hypothesis' as a central epistemological idea helped students to become “very conscious of the role of assumptions in the process of building a deductive theory” (p. 482).

David Tall, on the other hand, focusing on the cognitive aspects of proof, considered the growth of proof in recent years and synopsized his position in a personal communication as follows:

“The notion of mathematical proof involves a number of different aspects, including the structure of mathematics in differing mathematical contexts, the cultural views of different communities of practice, the cognitive and affective development of each individual, and the structure and operation of the human brain. Given the wide range of differences between individuals, it is not possible to formulate a single notion of mathematical proof suitable for all. In the ICMI study chapter on the cognitive development of proof, (Tall et al, 2011), researchers focused on different aspects of proof including geometry, the symbolism of arithmetic and algebra and formal proof in pure mathematics and logic. This relates to an initial analysis in terms of visual, *symbolic* and *formal axiomatic* proof supported by the increasingly sophisticated use of language. In *How Humans Learn to Think Mathematically* (Tall, 2013) the visual aspect was broadened to include all forms of sensory input and its mental processing, referring to this as embodied. Growth in embodiment and symbolism was interpreted as growing in sophistication from a practical level, where properties are seen to occur together, related coherently, to a theoretical level where selected properties are used as definitions and other properties are deduced from the definitions. There is a fundamental change to an *axiomatic formal* level based on set theoretic axioms where all properties are proved as theorems using propositional logic. Some theorems, called *structure theorems*, prove properties that can be interpreted in embodied and symbolic forms to offer insights that lead to more sophisticated formal theory. For individuals with appropriate experience this sophisticated theory can be more meaningful while others may find it more complicated and meaningless. My current approach for university students studying formal mathematics to progress from intuition to rigour is formulated in the latest edition of *Foundations of Mathematics* (Stewart & Tall, 2014).” (D. Tall, personal email communication, January 20, 2020).

Differing in their approach to proof from both Jahnke and Tall, Durand-Guerrier et al (2012) emphasized the role of formal logic in mathematical reasoning and proving. While they noted the value of logic in both syntactic and semantic mathematical discourse, their focus was on the logical organization of statements as a prerequisite for both constructing and understanding a proof. In their view, the model-theoretic approach introduced by Tarski, as well as an explicit knowledge of the predicate calculus, are necessary elements in teaching proof, since inference concepts and rules (such as *modus ponens, modus tollens*, syllogism, and disjunction) cannot be expected to be part of the common knowledge of students.

The authors added that four additional rules of inference are essential to the learning of proof, namely universal instantiation, existential instantiation, existential generalization, and universal generalization (Durand-Guerrier et al, 2012, p. 374). They argued that all the mentioned rules must be included in the mathematics curriculum, mainly because they are essential in helping students understand and construct proofs. “Having some level of understanding of the fundamental rules of predicate logic helps students check mathematical statements that are in doubt, avoid invalid deductions, and comprehend the basic structures of both mathematical proof (direct and indirect) and disproof by counterexample.” p. 374.

Empirical research carried out by Epp (2003) and Durand-Guerrier (2003; 2008) has shown that students at the undergraduate level have difficulties with the reasoning required to determine the truth or falsity of mathematical statements. As they note, this situation only serves to reinforce their position that “it is important to view logic as dealing with both the syntactic and the semantic organization of mathematical discourse.” (Durand-Guerrier et al., p. 385).

2. Empirical research on reasoning and proof: Selected representative studies 2000-2020

Dynamic geometry environments (DGE): Since dynamic geometry environments (DGE) have become available, a very large number of research papers have been published investigating their effects on teaching (De Villiers, 2004); Mariotti, 2002). In several countries the introduction of DGE to the curriculum at the elementary and secondary levels seems to have been very well received by both teachers and students. The new graphical tools offered to students strong and convincing evidence that certain propositions (of geometric figures) are indeed true. The limitation of DGE was that it could not show the students why these propositions are true. Accordingly, many research papers explored the extent to which the use of DGE could help students develop their abilities in mathematical reasoning.

Notably, a set of papers published in a special issue of "Educational Studies in Mathematics" (edited by Jones, Gutiérrez & Mariotti, 2000) discussed the novelty of dynamic environments as well as their potential advantages and pitfalls. They were based, of necessity, on experience with the two DGEs widely in use in the late 1990s and the early 2000s: *Cabri* (Laborde, 1990) and Geometer Sketchpad (Jackiw, 1991).

Hadas et al. (2000) reported on a teaching experiment carried out in a grade eight classroom that led then to conclude that the design of judicious activities in conjunction with DGE does encourage students to seek and provide deductive explanations. They developed a collection of "innovative activities intended to cause surprise and uncertainty" (p. 128), in order to overcome the tendency of students to seek conviction through empirical evidence only and thus to create a need for deductive explanations.

Jones (2000) reported on a longitudinal study, also with grade eight students, showing that the students made big strides towards precise mathematical explanations. This research showed "the mediational impact of using dynamic geometry software" (p.80). Marrades and Gutiérrez (2000) reached a similar conclusion through their investigation of ways in which dynamic geometry software can be used to improve students' proof skills. They analyzed the variety of justifications students offered when trying to construct proofs with the use of a DGE, and arrived at the tentative conclusion that the software "may well help secondary school students understand the need for abstract justifications and formal proofs (p. 119)."

The mid-2000s saw the advent of more advanced DGE, in the form of Geogebra (Hohenwarter, 2002), an interactive geometry program with algebraic capabilities. Since it became accessible to students and teachers in secondary schools, teaching with DGE software has become a useful tool for the design of effective mathematical instruction in several countries. Subsequently, there was a proliferation of research investigating all aspects of mathematics learning, in addition to proof and proving, using such dynamic software. Most studies provided strong evidence of the effectiveness of the software.

Research on proof at the undergraduate level: Empirical research on proof at the undergraduate level has grown at an exponential rate since the beginning of the 2000s. It is not possible to cover here all the facets of this research, therefore we will only report on a few papers that deal with proof comprehension and discuss the extent to which undergraduate students are able to provide acceptable proofs.

Selden and Selden (2013) see proof construction as having two facets. One is the approach of problem-solving, which relies primarily on informal but plausible reasoning. The other approach is that of what might be termed formal rhetoric, the "part of a proof that depends on unpacking and using the logical structure of the statement of the theorem, associated definitions, and earlier results" (p. 308). Based on this dichotomy, the authors developed an instruction intervention that has been helpful to students in generating the insights needed to construct a proof.

Hemmi (2006) investigated undergraduate students' learning of proof by becoming part of a community of practice consisting of their instructors, their peers, and the university environment. She introduced the concept of transparency that emphasizes the difference between a deductive argument and simple empirical evidence. She found that a crucial requirement for the learning of proof is to make all the components of a proof, such as, the logical structure and the significant main ideas, visible to the students without ignoring less visible (or invisible) elements, such as, explaining and justifying. She therefore advocated finding a right balance between the visible and invisible elements to teach proof efficiently.

The "Proof comprehension research group" based at Rutgers University has published a very large number of papers on topics related to proof comprehension. The group's research topics include the assessment of reading comprehension of mathematical proofs, the similarities and differences in the ways in which expert and novices read proofs, and more. Some researchers investigated the process of proving in a framework consisting of the analysis of solely two approaches to proving, a predominantly syntactic style and a predominantly semantic style (Alcock & Inglis, 2009; Alcock & Weber, 2010; Weber & Alcock, 2004). Weber and Mejia-Ramos (2009), however, found that this framework turned out to be too coarse when they tried to use it to it analyze students' proofs, and so suggested what they saw as a more refined framework for the evaluation of proof productions. They suggested adding a dimension called "inferences", so that "for each inference, we can then determine the extent to which, and the ways in which, semantic or syntactic reasoning contributed to this inference" (p.214). Zazkis et al. (2016) examined the proof construction activities of undergraduates when they attempted to write a symbolic proof based on a graphic argument; they found that students were able to produce correct syntactic proofs from graphic arguments using tactics, which the authors term "elaborating", "syntactifying", and "rewarranting".

Raman (2003) made a somewhat analogous distinction between proofs based on procedural ideas and proofs based on a key idea, with the former being based on purely syntactic manipulation and the latter being based on an insight gained from an informal, private way of understanding a concept. Raman advocates making key ideas a more central part of the teaching of proof. Yan (2017) expanded Raman's concept of key idea to include the role of main ideas in proofs and the concept of the width of a proof (Gowers, 2007; Hanna & Mason, 2014) in helping a learner understand and remember a proof. Yan's research revealed that the concept of key idea was more complex and "more subjective than initially anticipated at the start of this research". It turned out "that the identification of key ideas is not only context dependent but also learner dependent" (p.131).

The past four decades have seen an enormous number of research studies on proof and proving in mathematics education that made important advances in our knowledge on the teaching of proof. For recent extensive reports on proof in mathematics education research and teaching, see Hanna and de Villiers (2012), Mariotti et al (2018), Stylianides et al. (2016), and Stylianides et al. (2017).

LOOKING AHEAD: New directions in proof

Over the last few decades, mathematical practice has become increasingly digitised. A prime example is the growing use of a range of proof assistants that, with varying degrees of autonomy, can translate informal proofs into formal ones and verify their validity. A major reason for this growth is that mathematicians are looking for early and thorough verification of their own work, since they have come to see the traditional refereeing process as slow and unreliable. In fact, using such computer programs, they can now check the correctness of a formalized proof at a level of certainly that they could not possibly achieve when examining informal proofs (Avigad & Harrison, 2014). According to Wiedijk (2008), computer programs have successfully checked the validity of the proofs of several well-known theorems, such as the Fundamental Theorem of Algebra (2000), Jordan's Closed Curve Theorem (2005), the Fundamental Theorem of Calculus (1996), the Four Colour Theorem (2004), and the Prime Number Theorem (2008). Field medalist Vladimir Voevodsky strongly supports the use of proof assistants:

> *And I now do my mathematics with a proof assistant. I have a lot of wishes in terms of getting this proof assistant to work better, but at least I don't have to go home and worry about having made a mistake in my work. I know that if I did something, I did it, and I don't have to come back to it nor do I have to worry about my arguments being too complicated or about how to convince others that my arguments are correct. I can just trust the computer. (Voevodsky, 2014, p. 9).*

In light of these new developments in mathematical practice in the area of proof and verification, mathematics educators need to develop new approaches to the teaching of proof that capitalize on the newly available technology. A major reason is that without such new approaches the use of proof in the classroom may come to reflect an outmoded view of mathematics as a discipline. If students are to gain an understanding of mathematics as it is currently practiced, they will have to acquire a familiarity with state-of-the-art tools such as computer-based proof assistants.

Another reason for new classroom approaches is that the use of proof technology may present a new opportunity for fostering mathematical understanding, similar to that which has already been seen in the use of other technology such as dynamic geometry programs. Interactive and automated theorem provers (ITP and ATP) are designed to provide a guarantee of correctness. The question is how and to what degree these tools can go beyond providing a guarantee and be useful in providing explanation as well. Focusing on students' understanding of mathematics, and of proof in particular, while making use of automated proving, will be a promising avenue of research. Clearly, mathematics educators will have several challenging open questions to consider when looking ahead.

ACKNOWLEDGEMENTS

We are grateful to the reviewers for their helpful comments. We wish to acknowledge the generous support of the Social Sciences and Humanities Research Council of Canada.

References

Alcock, L., & Inglis, M. (2009). Representation systems and undergraduate proof productions: A comment on Weber. *Journal of Mathematical Behavior, 28*(4), 209-211.

Alcock, L., & Weber, K. (2010). Referential and syntactic approaches to proving: Case studies from a transition-to-proof course. *Research in Collegiate Mathematics Education,* 7, 101-123.

Atiyah, M. et al. (1994). 'Responses to "Theoretical Mathematics": Towards a cultural synthesis of mathematics and theoretical physics, by A. Jaffe and F. Quinn', *Bulletin of the American Mathematical Society, 30*(2), 178-207.

Avigad, J., & Harrison, J. (2014). Formally verified mathematics. *Communications of the ACM 57*(4), 66-75.

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216-235). London: Hodder and Stoughton.

Balacheff, N. (1991). The benefits and limits of social interaction: The case of mathematical proof. In A. J. Bishop, E. Mellin-Olsen, & J. van Dormolen (Eds.), *Mathematical knowledge: Its growth through teaching* (pp. 175-192). Dordrecht, Netherlands: Kluwer.

Beberman, M. (1958). *An emerging program of secondary school mathematics*. Cambridge: Harvard University Press.

Blum, W., & Kirsch, A. (1991). Preformal proving: Examples and reflections. *Educational Studies in Mathematics, 22*(2), 183-203.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics, 24*(4), 359-387.

Cobb, P., & Steffe, L. (1983). The constructivist researcher as teacher and model builder. *Journal for Research in Mathematics Education, 14*(2), 83-94. doi:10.2307/748576

Coe, R., & Ruthven, K. (1994). Proof practices and constructs of advanced mathematics students. *British Educational Research Journal, 2*(1), 41-53.

Davis, P. J., & Hersh, R. (1983). *The mathematical experience*. Harmondsworth: Penguin Books.

Davydov, V. V. (1995). (Translated by Stephen T. Kerr). The influence of L. S. Vygotsky on education theory, research, and practice. *Educational Researcher, 24*(3), 12–21. https://doi.org/10.3102/0013189X024003012

De Villiers, M. (2004). Using dynamic geometry to expand mathematics teachers' understanding of proof. *The International Journal of Mathematical Education in Science and Technology, 35*(5), 703-724.

Dreyfus, T., & Hadas, N. (1996). Proof as an answer to the question why. *Zentralblatt für Didaktik der Mathematik (ZDM), 28*(1), 1-5.

Dunfield, W., & Thurston, W. (2003). The virtual Haken conjecture: experiments and examples. *Geom. Topol. 7*, 399-441.

Durand-Guerrier, V., Boero, P., Douek, N., Epp, S.S., & Tanguay, D. (2012). Examining the role of logic in teaching proof. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 369-389). Dordrecht: Springer.

Fischbein, E. (1982). Intuition and proof. *For the learning of mathematics 3*(2), 8-24.

Fischbein, E. (1987). *Intuition in science and mathematics*. Dordrecht: Kluwer.

Fischbein, E. (1999). Intuitions and schemata in mathematical reasoning. *Educational Studies in Mathematics, 38*(1-3), 11-50.

Gowers, W. T. (2007). Mathematics, memory, and mental arithmetic, In M. Leng, A. Paseau, & M. Potter (Eds.), *Mathematical knowledge* (pp. 33-58). Oxford, UK: Oxford University Press.

Grabiner, J. V. (1986). Is mathematical truth time-dependent? In T. Tymoczko (Ed.), *New directions in the philosophy of mathematics: An anthology* (pp. 201-213). Boston: Birkhäuser.

Gutiérrez, A., & Boero, P. (Eds.), (2006). *Handbook of research on the psychology of mathematics education: Past, present and future*. Rotterdam/Taipei: Sense Publishers.

Hadas, N., Herschkowitz, R., & Schwarz, B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environment. *Educational Studies in Mathematics, 44*(1&2), 127-150.

Hanna, G. (1983). *Rigorous proof in mathematics education*. Toronto: OISE Press.

Hanna, G. (1989a). More than formal proof. *For the Learning of Mathematics, 9*(1), 20-25.

Hanna, G. (1989b). Proofs that prove and proofs that explain. In *Proceedings of the Thirteenth International Conference on the Psychology of Mathematics Education*, Vol. 2, 45-52. Paris.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21*(1), 6-13.

Hanna, G. (2003). Journals of mathematics education, 1900-2000. In Coray, D. and Furringhetti, F.

(Eds.), *One Hundred Years of L'Enseignement Mathématique : Moments of Mathematics Education in the 20th Century* (pp. 67-84). Geneva: L'Enseignement Mathématique.

Hanna, G., & de Villiers, M. (Eds.), (2012). *Proof and proving in mathematics education*, New York, NY: Springer.

Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In Bishop, A. et al (Eds.), *International handbook of mathematics education* (pp. 877-905). Dordrecht: Kluwer Academic Publishers.

Hanna, G., & Mason, J. (2014). Key ideas and memorability in proof. *For the Learning of Mathematics, 34*(2), 12-16.

Hanna, G., & Winchester, I. (Eds.), (1990). *Creativity, thought and mathematical proof.* Toronto: Ontario Institute for Studies in Education.

Harel, G. (2008). What is mathematics? A pedagogical answer to a philosophical question. In R. B. Gold, & R. Simons (Eds.), *Proof and other dilemmas: Mathematics and philosophy* (pp. 265-290). Washington, DC: Mathematical Association of America.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education*, III (pp. 234-283). Washington DC: Mathematical Association of America.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4), 396-428.

Hemmi, K. (2008). Students' encounter with proof: the condition of transparency. *ZDM–– The International Journal on Mathematics Education, 40*(3), 413-426.

Hohenwarter, M. (2002). GeoGebra - Ein Software system für dynamische Geometrie und Algebra der Ebene. Master thesis, University of Salzburg.

Hoyles, C. (1997). The curricular shaping of students' approaches to proof. *For the Learning of Mathematics, 17*(1), 7-16.

Jaffe, A. & Quinn, F. (1993). "Theoretical mathematics": Toward a cultural synthesis of mathematics and theoretical physics. *Bulletin of the American Mathematical Society, 29*(1) 1-13.

Jahnke, H. N. (2005). A genetic approach to proof. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 428-437). Sant Feliu de Guíxols.

Jahnke, H. N. (2007). Proofs and hypotheses. *ZDM – The International Journal on Mathematics Education, 39*(1–2), 79-86.

Jahnke, H. N., & Wambach, R. (2013). Understanding what a proof is: a classroom-based approach. *ZDM – The International Journal on Mathematics Education 45*(3), 4690-482. https://doi.org/10.1007/s11858-013-0502-x

Jones, K. (2000). Providing a foundation for a deductive reasoning: students' interpretation when using dynamic geometry software and their evolving mathematical explanations. *Educational Studies in Mathematics, 44*(1&2), 55-85.

Jones, K., Gutiérrez, A., & Mariotti, M. A. (Eds.), (2000). Proof in dynamic geometry environments. *Educational Studies in Mathematics, 44*(1-3). [Guest editorial for a PME special issue]

Kline, M. *(1973). Why Johnny can't add: The failure of the New Math. New York:* St. Martin's Press. ISBN 0-394-71981-6.

Laborde, J.-M. (1990). Cabri-Geometry [Software]. Université de Grenoble, France.

Lakatos, I. (1983*). Proofs and refutations*, 4th reprint. Cambridge: Cambridge University Press.

Lehrer, T. (1965). *That was the year that was*. Live album. https://www.youtube.com/watch?v=UIKGV2cTgqA

Lerman, S. (2019). Theoretical aspects of doing research in mathematics education: An argument for coherence. In Kaiser G., Presmeg N. (Eds.), *Compendium for early career researchers in mathematics education* (pp. 309-324). ICME-13 Monographs. Cham: Springer.

Maher, C. A., & Martino, A. M. (1996). The development of the idea of a mathematical proof: A 5-year case study. *Journal for Research in Mathematics Education, 27*(2), 194-214.

Mariotti, M. A. (2002). Influence of technologies advances on students' math learning. In L. English (Ed.), *Handbook of International Research in Mathematics* (pp. 695-724). New Jersey: Lawrence Erlbaum Associates.

Mariotti, M. A., Durand-Guerrier, V., Stylianides, G. J. (2018). Argumentation and proof. In T. Dreyfus, M. Artigue, D. Potari, S. Prediger & K. Ruthven (Eds.), *Developing research in* *mathematics education. Twenty years of communication, cooperation and collaboration in Europe* (pp. 75–89). Oxon, UK: Routledge.

Mariotti, M. A., & Pedemonte, B. (2019). Intuition and proof in the solution of conjecturing problems. *ZDM-– The International Journal on Mathematics Education, 51*(5), 759-777.

Marrades, R., & Gutiérrez, Á. (2000). Proofs produced by secondary school students learning geometry in a dynamic computer environment. *Educational Studies in Mathematics, 44*(1), 87-125. https://doi.org/10.1023/A:1012785106627

National Council of Teachers of Mathematics. (1963). *An analysis of new mathematics programs*. Washington, DC: NCTM.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: NCTM. [cit. 20150912]. Available from: https://www.nctm.org/uploadedFiles/Standards_and_Positions/PSSM_ExecutiveSummary.pdf

National Council of Teachers of Mathematics. (1980). *An Agenda for Action: Recommendations for school mathematics of the 1980s*. Reston, VA: NCTM.

Polya, G. (1954). *Mathematics and plausible reasoning*. Princeton, NJ: Princeton University Press.

Polya, G. (1981). *Mathematical discovery: On understanding, learning, and teaching problem-solving*. New York, NY: John Wiley & Sons.

Polya, G. (1985). *How to solve it: A new aspect of mathematical method*. Princeton and Oxford: Princeton University Press.

Radford, L. & Roth, WM. (2017). Alienation in mathematics education: a problem considered from neo-Vygotskian approaches. *Educational Studies in Mathematics, 96*(3), 367–380. https://doi.org/10.1007/s10649-017-9769-0

Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof? *Educational Studies in Mathematics, 52*(3), 319-325.

School Mathematics Study Group (SMSG). (1962). *Grades 7-12 texts and commentaries*. Yale: Yale University Press.

Selden, A., & Selden, J. (2013). Proof and problem solving at university level. *Montana Mathematics Enthusiast, 10*(1-2), 303-334.

Senk, S. (1985). How well do students write geometry proofs? *The Mathematics Teacher, 78*(6), 448-456.

Stanic, G. M. A., & Kilpatrick, J. (1988). Historical perspectives on problem solving in mathematics curriculum. In R. I. Charles & E. A. Silver (Eds.), *Research agenda for mathematics education: The teaching and assessing of mathematical problem solving* (pp. 1-22). Reston, VA: National Council of Teachers of Mathematics.

Stewart, I. N., & Tall, D. O. (2014). *Foundations of mathematics, Second Edition*. Oxford: OUP. ISBN: 9780198531654

Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutiérrez, G. C., Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 315-351). Rotterdam: Sense Publishers.

Stylianides, G., Stylianides, A., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 237-266). Reston, VA: National Council of Teachers of Mathematics.

Tall, D. (2013). *How humans learn to think mathematically*. New York: Cambridge University Press. ISBN: 9781139565202.

Tall, D., Yevdokimov, O., Koichu, B., Whiteley, W., Kondratieva, M., & Cheng, Y.-H. (2012). The cognitive development of proof. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 13-49). New York, NY: Springer.

Thurston, W. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society, 30*(2), 161-177.

Thurston, W. (1995). On Proof and Progress in Mathematics. *For the Learning of Mathematics, 15*(1), 29-37.

Vergnaud, G. (1987). About constructivism. In J. C. Bergeron, N. Herscovics, & C. Kieran (Eds*.), Proceedings of the 11th PME International Conference*.

Voevodsky, V. (2014). The origins and motivations of univalent foundations. *The Institute Letter*, 8-9.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics, 56*(2), 209-234.

Weber, K., & Mejia-Ramos, J.P. (2009). An alternative framework to evaluate proof productions: A reply to Alcock and Inglis. *Journal of Mathematical Behavior, 28*(4), 212-216.

Wiedijk, F. (2008). Formal proof – getting started. *Notices of the American Mathematical Society, 55*(11), 1408-1414.

Williams, J. (2016). Alienation in mathematics education: critique and development of neo-Vygotskian perspectives. *Educational Studies in Mathematics, 92*(1), 59-73. doi:10.1007/s10649-015-9659-2

Yackel, E., & Cobb, P. (1994). The development of young children's understanding of mathematical argumentation. Paper presented at the annual meeting of the American Educational Research Association (AERA), New Orleans.

Yan, X. (2017). *Key ideas in proof in undergraduate mathematics classrooms*. (Unpublished doctoral dissertation.) University of Toronto, Toronto, Canada.

Zazkis, D., Weber, K., & Mejia-Ramos, P. (2016). Bridging the gap between visual arguments and verbal-symbolic proofs in calculus. *Educational Studies in Mathematics, 93(*2), 155-173.