THE MEANING OF PROOF IN MATHEMATICS EDUCATION

David A Reid

Acadia University

The issue of what mathematics education researchers mean by "proof" and "proving" has been the topic of three recent papers. The discussion in those papers are analysed in terms of a common terminology for identifying characteristics of the meanings of proof current in research.

> Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache (The meaning of a word is its use in the language). (Wittgenstein**,** 1963, §43.).

Perhaps it is a sign of the maturity of research into the teaching and learning of proof and proving that we are beginning to reflect on what it is we are researching, and whether, as a community, we are successful in communicating our work to each other. Three papers have been presented in the past seven years exploring the different meaning for "proof" and "proving" used in different research communities within mathematics education. In this paper I will continue this exploration and attempt to create a common terminology through which these meanings can be interpreted.

At PME 21 in Finland, Godino and Recio (1997) described some of the meanings proof has in the domains of research in mathematical foundations, mathematics, the sciences, and in classrooms.

At PME 25 in Utrecht I identified four usages of "proof" in mathematics education research and one from everyday life (Reid 2001). I classified these under the headings "the concept of proof", "proofs", "proving" and "probing".

At the Taipei International conference on "Mathematics: Understanding Proving and Proving to Understand" Nicholas Balacheff (2002/2004) described seven different perspectives on proof from which research has been done, which he characterised as being based on different researcher epistemologies.

In this paper I will revisit these three papers, as well as referring to the work of some other authors to illustration the different meanings of proof and proving. I will discuss the differing meanings of proof and proving in the research literature in terms of several different dimensions along which their meanings can differ. These dimensions include the concept of proof, the purpose of teaching proof, the kinds of reasoning proving is seen to involve, the needs that proving is seen to address and the relationship seen between proof and language. These dimensions are not explicitly mentioned in the three papers, but rather I have identified them through comparison of the distinctions made in them. Other dimensions, or a different set of dimensions, might also be useful for characterising the different meanings of proof and proving in use in the mathematics education research literature, but my reading of these three papers did not suggest them. For example, the word "rigour" is often used in discussions of proof, but it is only used in one of these three papers and so it did not provide a useful dimension for characterising the distinctions being made in all three.

The concept of proof

Godino and Recio point out that what counts as proof is different in formal logic, mainstream mathematics, science and schools. This is not really terribly surprising, as each of these is a different “domain of explanation”, a phrase Maturana (1988) coined to describe communities that have different criteria for accepting an argument as an explanation.

> … the observer accepts or rejects a reformulation … as an explanation according to whether or not it satisfies an implicit or explicit criterion of acceptability …. If the criterion of acceptability applies, the reformulation … is accepted and becomes an explanation, the emotion or mood of the observer shifts from doubt to contentment, and he or she stops asking over and over again the same question. As a result, each … criterion for accepting explanatory reformulations … defines a domain of explanations, …. (1988, p. 28)

The concept of proof that Godino and Recio claim operates in the domain of formal logic is believed by some researchers in mathematics education to operate in mainstream mathematics as well (as I pointed out in my 2001 paper). It is the belief that, in the words of Fischbein & Kedem (1982) “a formal proof of a mathematical statement confers on it the attribute of a priori universal validity.” This concept of proof is very old, and here I will call it the “traditional” concept of proof.

Another concept was ascribed to the mathematical domain of explanation by Godino and Recio. It has also been proposed by mathematicians, sociologists of mathematics, and philosophers of mathematics (e.g., Davis 1972, Lakatos 1976, Tymoczko 1986 , Crowe 1988). They point out that proofs are produced by fallible mathematicians and so cannot establish absolute truth. Instead they see proof as part of a quasi-empirical process, in which proofs operate to clarify and make the detection of errors easier (but never complete).

Whether a researcher has a traditional or quasi-empirical concept of proof can have a significant impact on the research done and the results reported. For example, Fischbein and Kedem (1982) concluded that high school students did not understand proof, because the students did not have the traditional concept of proof. Researchers operating from a quasi-empirical concept may have reached a different conclusion, or more likely, would have done the study differently.

The purpose of teaching proof

Balacheff’s first example of a researcher’s epistemology is that of Fawcett (1938). Fawcett undertook a teaching experiment motivated by the belief that the study of proof in mathematics could develop his students' ability to think critically in other domains of explanation. In his introduction he makes it quite clear that the only justification he sees for teaching geometry in upper secondary school is the transfer of the method of proof learned in that context to other contexts. This can be contrasted with the purpose for teaching proof that can be inferred from the quotation Balacheff selects from the work of Healy and Hoyles (1998): "Proof … exemplifies the distinction between mathematics and the empirical sciences." This suggests a belief that the criterion of acceptability of explanation that defines mathematics as a domain of explanation cannot be transferred to other domains, as Fawcett claimed. Instead the motivation for teaching proof is a better understanding of the nature of mathematics itself, not better reasoning in other domains. This dimension of the meaning of proof can be labelled Critical Thinking versus Mathematics.

It is worth pointing out that this dimension is distinct from the dimension of the concept of proof. Fawcett is an example of the combination of the traditional concept of proof with the Critical Thinking purpose. Based on their emphasis on justification and the generality of a valid proof, I believe Healy and Hoyles also work from the traditional concept of proof, but as noted above their purpose for teaching proof is Mathematics. In some of my work (Reid 1995) I advocated a quasi-empirical concept of proof, but with a critical thinking purpose, suggesting that showing students the limits of mathematics would make them more critical of the use of numerical arguments in other domains.

Reasoning

Balacheff tries to contrast Fawcett's "epistemology" with that of Harel and Sowder (1998), an attempt that fails, I believe, because Fawcett is most definitely working from a traditional concept of proof, and Harel and Sowder explicitly state that the "proof schemes" they describe are not to be confused with mathematical proof at all. At least part of this confusion stems from their use of "prove" in the sense it has in everyday English, which is different from the mathematical use of the word, so perhaps Harel and Sowder could be said to be using the "everyday" concept of proof. One might expect that this confusion is lessened in Romance languages in which word like "démonstration" and "demostración" exist to refer specifically to "mathematical proof", but as the everyday words "preuve" and "*prueba*" are also used to refer "mathematical proof" confusion can still occur. (The possibility for confusion becomes almost absurd when comparing work written by native speakers of English and romace languages. For example, Sutherland, Olivero & Weeden (2004) describe the work of a teacher who stressed the difference between proof and demonstration in her teaching. Unfortunately, while one might think it reasonable to translate "proof" as "preuve/prueba" and "demonstration" as "démonstration/ demostración", the current usage of these words in the UK schools is such that "proof" should be translated as "démonstration/demostración" and "demonstration" as "preuve/prueba".

Harel and Sowder's work does, however, serve to illustrate another important dimension along which meanings of proof can differ: the kind of reasoning employed in proving. Their "proof schemes" include reference to authority and empirical reasoning as well as deductive reasoning, each proof scheme being, in fact, a specific kind of reasoning. Other researchers have suggested other kinds of reasoning (abductive, analogy) and sub-types (generic example) that will be familiar to most readers.

There is another meaning of proving in which it can involve several types of reasoning at the same time. Godino and Recio point out that in science proving involves a mixture of kinds of reasoning and the larger patterns of proving processes described in Knipping (2003) and Reid (2002) also involve other kinds of reasoning combined with deductive reasoning.

The majority of research on proof, however, uses "proving" in ways that assume it must involve deductive reasoning. Some, in fact, define proving in terms of the kind of reasoning involved and restrict it to deductive reasoning. For example, ten years ago I defined proving as "investigating using deductive reasoning" (Reid 1995 p. 7), and the NCTM Principles and Standards describes proofs as: "arguments consisting of logically rigorous deductions of conclusions from hypotheses"(NCTM 2000 p.56).

Thus, in the mathematics education research literature, one can observe three different points on the reasoning dimension of the meaning of proof: Several distinct kinds of reasoning are involved in proving; or several kinds of reasoning in combination are involved in proving; or deductive reasoning alone constitutes proving. This dimension is distinct from the dimensions of concept and purpose, but it seems reasonable to suppose that researcher with a traditional concept of proof would be more likely to see proof as purely deductive, while those with a quasi-empirical concept of proof might be more inclined to view proving as a combination of kinds of reasoning.

Needs

One of the most astonishing claims in the paper by Godino and Recio is their assertion that "mathematics students must [use proof to] convince themselves, and convince their teacher of the necessary and universal truth of theorems" (p. 2-318). It is not clear how a student can convince a teacher of something that the teacher either already believes (if the teacher has a traditional concept of proof) or cannot believe (if the teacher has a quasi-empirical concept of proof). And as to convincing themselves, Harel and Sowder's work suggests that students are unlikely to use mathematical proof to convince themselves of anything. This raises the question, when students engage in proving, what need are they trying to address?

Some researchers, notably de Villiers (1990) and Hanna (2000), have listed a range of functions for proof, or needs that proof fulfils for individuals, including verification, exploration, explanation, systematisation, communication and social acceptance. Most researchers in mathematics education take one of four approaches to considering what need motivates proving. Many, especially those who have the traditional concept of proof, assume implicitly or explicitly that the need of verification is the only one that motivates proving. Others (e.g., Thurston 1995, Mariotti 1997) describe the needs that motivate research mathematicians to prove and imply that the same needs *should* motivate students in schools to prove. Still others (e.g., Hanna 1989) describe the specific context of schools and suggest a need for proving appropriate to the school context, not necessarily a need that motivates research mathematicians. Finally, there are researchers (e.g., Reid 1995) who examine the mathematical activity of students and who identify the needs that in the existing school contexts do motivate students to prove.

The implications of differing beliefs about the needs that motivate proving for research and teaching are clear. Teachers who begin with an assumption that proving is about verification will create situations of doubt to motivate proving, while those who see explanation as motivating proving will create situations of wonder. Researchers who associate proof with verification (e.g., Harel & Sowder, 1998) will focus on the kinds of reasoning students use to verify, while those who see proving as motivated by many different needs will focus on other issues.

It is unlikely that all the possible combinations of the dimensions of the concept of proof and the need for proving can coexist. For example, a researcher who has a traditional concept of proof is likely to suggest that students' needs to verify or systematise could or should motivate them to prove. On the other hand, a researcher who has a quasi-empirical concept of proof might stress the needs to explore, to communicate, to gain social acceptance or to explain.

Proof and language

One of the distinctions Balacheff makes concerns the relationship between language and proof. He claims that for Duval (1991) proofs must rely only on syntactical elements, while others would claim that proofs rely on semantic or social elements for acceptance (Balacheff's examples are Pimm, 1987, and Burton and Morgan, 2000). Godino and Recio suggest a distinction that is useful for clarifying Balacheff's. They describe analytical and substantive arguments (from Krummheuer 1995, following Toulmin). Analytical arguments are based on specified, deductive rules of inference and explicit axioms; as such they are logical tautologies. Substantive arguments expand the meaning of what is argued. Godino and Recio use these categories to distinguish proofs in formal logic from other proofs, but Balacheff's distinction is between epistemologies of researchers in mathematics education.

It is evident that a researcher whose concept of proof is quasi-empirical is unlikely to mean by "proof" something analytical, however the substantive position seems compatible with both concepts.

The meaning of proof

Maturana (1988) emphasises that the criterion of acceptance used in a domain of explanation is often (necessarily) implicit. This may be part of the difficulty we have, as researchers in mathematics education, in being clear about what we are talking about when we refer to "proof" and "proving". That there are differences is apparent from the examples presented by Balacheff and in my 2001 paper, but it is only by considering many such examples that patterns of difference become evident. I speculate that it is possible to place each researcher in mathematics education somewhere in the space defined by the dimensions I have outlined above, but it is impossible to do this working only from existing publications, as we, as a community, do not make our positions clear. Perhaps the work of Balacheff, Godino and Recio, and my papers, will make it easier for us to communicate our positions, and thus provide a foundation for the kind of interaction needed to strengthen a community.

One reason we have difficulty in being clear is that the category "proof", like most categories, is not well defined, and probably cannot be. As Balacheff asserts, "almost all researchers will agree on a more or less formal definition of mathematical proof" (p. 1) but that definition, as Godino and Recio note, is useful only when working in the domains of logic and the foundations of mathematics. Most research mathematics and teacher of mathematics must contend with an informal and implicit definition, based on experiences they have had with prototypical examples of proofs.

The linguist Eleanor Rosch (Varela, Thomson & Rosch, 1991) has analysed the way categories are described in language, and has concluded that we do not have a definition in our heads for the words we use to classify our experiences. Instead we have what she calls "prototypical" examples. For example, if I ask you to think of a bird, you are likely to think of a common bird that lives where you live. You are unlikely to think of a penguin or an ostrich. For birds which resemble our prototype we recognise them as birds without making any recourse to a definition. For birds far from our prototype we rely on expert opinions to establish the borders of our categories. The experts tell us that penguins are birds, even though they do not fly and behave very much like seals, because they lay eggs. But on the other hand, they tell us that platypuses are not birds, even though they lay eggs and behave in many ways like ducks.

An experiment

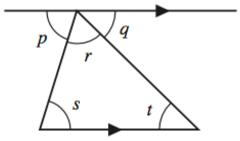
At CERME 4 the working group of proof brought together fifteen researchers with an interest in proof, from ten countries. I asked the fourteen others to form seven pairs, and to select the most exemplary proofs from a set of six I provided.

I have collected about fifty proofs, mostly from research papers on the topic (the full set is available on request). I distributed 40 of these (plus two duplicates) to the pairs. After choosing not more than three proofs as more exemplary than the others, they were to join another pair, and again select the most exemplary and discard the least exemplary. My plan was that either we would arrive at a consensus of a few very good examples of what a proof is (demonstrating that we share a common prototype for the category) or we would fail to reach a consensus (demonstrating that there are serious differences within the mathematics education community concerning what "proof" means). What actually occurred was a lot of interesting discussion within the pairs and groups of four, but the task of comparing and choosing exemplary proofs was so time consuming that the original plan had to be abandoned.

Nonetheless, some results can be reported. Three pairs (coded A, B, C) made two rounds of selections, and in the end they had selected three exemplary proofs from the 18 proofs given them (which included one duplicate). They selected #16, a visual "dot" proof that the sum of two even numbers is even, #19/44 a proof that the angle sum of a triangle is 180°, and #23, a "Behold!" proof of the Pythagorean theorem. #19/44 was the duplicate, a proof copied from one of the Hoyles-Healy-Kuchemann questionnaires (see Figure 1).

Proof that the sum of the interior angles of a triangle is 180 degrees

I drew a line parallel to the base of the triangle.



Statements	*Reasons*
$p = s$	Alternate angles between two parallel lines are equal
$q = t$	Alternate angles between two parallel lines are equal
$p + q + r = 180°$...	Angles on a straight line
$\therefore\ s + t + r = 180°$.	

Figure 1: A exemplary proof, #19/44, from Hoyles et al.

Prove: The sum of the first n positive integers is $n(n + 1)/2$.

Let $S(n) = 1 + 2 + 3 + \cdots + n$.

Then $S(n) = n + (n - 1) + (n - 2) + \cdots + 1$.

Taking the sum of these two rows,

$$2S(n) = (1 + n) + [2 + (n - 1)] + [3 + (n - 2)] + \cdots + (n + 1)$$
$$= (n + 1) + (n + 1) + (n + 1) + \cdots + (n + 1)$$
$$= n(n + 1).$$

Therefore, $S(n) = n(n + 1)/2$.

Figure 2: An exemplary proof, #10, cfrom Hanna (1989)

This proof was judged to be a better example of a proof than 13 others, via direct or indirect comparisons (the indirect comparisons occurred when it was judged to be better than a proof that had previously been preferred to others). Of course, as it wasn't compared to most of the proofs, or seen by most of the participants, it may not really be the best example. All three of the proofs chosen by this group use diagrams. In the second round they eliminated two other proofs, also using diagrams, of the Pythagorean theorem (#43) and the commutativity of multiplication (#30). In the first round the three groups A, B, C eliminated 12 proofs, 2 with diagrams and 10 without.

Pair G selected three proofs that have an algebraic element (#10, see Figure 2, #13 & #24, of the sum of the first *n* integers, the sum of two even numbers, and the Pythagorean theorem). The other formats available were an action proof (from Tall 1995), a narrative proof from Euclid (Book IX, Prop 20, the infinitude of primes) and an empirical argument. Their selection seems to have been influenced by the use of algebra, at least more than by the reputation of the author (Euclid) or the accessibility (action) of the proofs.

Pair E selected three narrative proofs in the domain of number theory (#8, #27, & #28, of the sum of the first *n* integers, the irrationality of $\sqrt{5}$ and the infinitude of primes). #8 also includes a visual generic example. They rejected a textbook geometry proof of an insignificant theorem, a proof from Euclid (Book I, Prop 1) and a non-standard proof of the angle sum of a triangle (based on tessellations, from Hoyles et al.). It may be that the narrative element in the proofs they chose is coincidence, but their preference for number theory over geometry seems clear.

When a similar experiment was done at Universität Oldenburg with only ten proofs, #8 and #10 (which are included in the choices of groups G and E) were selected as highly exemplary.

Note that the nine proofs chosen by these five pairs are proofs of only six theorems, all of them well known. This is partly a result of the set of proofs they had to choose from, as half the proofs given were proofs of these six theorems, but the fact that these and only these were selected suggests that the mathematical significance or familiarity of the theorem proven may make a proof more exemplary.

Pair D combined with either Pair E or Pair G in the second round (I have no method of determining which pair they joined), at which point the two proofs they had selected in the first round were eliminated. One of their selections, #11, is similar in structure to #8 and #10, but uses a numerical generic example. The other, #31, is a diagram based proof of an algebraic identity. Neither of their choices include algebra or a narrative element, and this might have been a factor in their elimination when Pair D joined another pair.

Pair F spent a great deal of time discussing the proofs they were given, and by the end of the allotted time they had eliminated one, a formal proof from a logic text.

What does all this mean? There seems to be some common ground about what a proof is, at least in the context of this working group. But there seems also to be some differences. There is more work to be done before we have a clear idea of what those differences are and what significance they have for the research we do. But if we can acknowledge that there is an issue here, and discuss the characteristics of proof, we may be able to come to, if not agreement, then at least agreement on how we differ.

References

Balacheff N. (2002/2004) The researcher epistemology: a deadlock from educational research on proof. Fou Lai Lin (ed.) 2002 *International Conference on Mathematics - "Understanding proving and proving to understand"*. Taipei: NSC and NTNU (pp. 23-44). Reprinted in *Les cahiers du laboratoire Leibniz*, no 109, August 2004, Online at http://www-leibniz.imag.fr/NEWLEIBNIZ/LesCahiers/Cahier109/ResumCahier109.html

Burton L., Morgan C. (2000) Mathematicians writing. *Journal for Research in Mathematics Education* 31(4) 429-452.

Crowe, M. (1988). Ten misconceptions about mathematics and its history. In W. Aspray & P. Kitcher (Eds.) *History and philosophy of modern mathematics, Minnesota Studies in the philosophy of science, Volume XI,* (pp. 260-277). Minneapolis: University of Minnesota Press.

Davis, P. (1972). Fidelity in mathematical discourse: Is one and one really two? *American Mathematical Monthly, 79*, 252-263. (Reprinted in Tymoczko, 1986).

de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras,* n. 24, pp. 17-24.

Duval R. (1991) Structure du raisonnement déductif et apprentissage de la démonstration. *Educational Studies in Mathematics 22*(3) 233-263.

Fawcett, H. P. (1938). *The nature of proof: a description and evaluation of certain procedures used in a senior high school to develop an understanding of the nature of proof.* (NCTM yearbook 1938). New York: Teachers' College, Columbia University.

Fischbein, E. & Kedem, I. (1982). Proof and certitude in the development of mathematical thinking. In A. Vermandel (Ed.), *Proceedings of the Sixth Annual Conference of the International Group for the Psychology of Mathematics Education,* (pp. 128-131). Antwerp.

Godino J. D., Recio A. M. (1997). Meaning of proofs in mathematics education In Pekhonen, Erkki (Ed.), *Proceedings of the Twenty-first Conference of the International Group for the Psychology of Mathematics Education.* (Vol. 2 pp. 313-320). Lahti, Finland.

Hanna, G. (1989). Proofs that prove and proofs that explain. In *Proceedings of the Thirteenth International Conference on the Psychology of Mathematics Education,* (pp. 45-51). Paris.as

Hanna G. (2000) : Proof, Explanation and Exploration: An Overview. *Educational Studies in Mathematics*, *44*(1&2), Keith Jones, Ángel Gutiérrez Maria Alessandra Mariotti (eds.) 5-23.

Harel G., Sowder L (1998) Students' proof schemes: Results from exploratory studies. In: Schonfeld A., Kaput J., and E. Dubinsky E. (eds.) *Research in collegiate mathematics education III*. (Issues in Mathematics Education, Volume 7, pp. 234-282). American Mathematical Society.

Healy L. & Hoyles C. (1998) *Justifying and proving in school mathematics. Summary of the results from a survey of the proof conceptions of students* *in the UK*. Research Report. Mathematical Sciences, Institute of Education, University of London.

Hoyles C., Healy L. & Kuchemann, D. (1995-2003) Justifying and Proving in School Mathematics Project and Longitudinal Proof Project. Description and list of publications online at: http://www.ioe.ac.uk/proof/index.html.

Knipping, C. (2003). *Beweisprozesse in der Unterrichtspraxis – Vergleichende Analysen von Mathematikunterricht in Deutschland und Frankreich*. Hildesheim: Franzbecker.

Krummheuer G. (1995) The ethnography of argumentation. In: Cobb. P., Bauersfeld H. (eds.) *The emergence of mathematical meaning: interaction in the classroom culture*. Hillsdale, NJ: Erlbaum.

Lakatos, I. (1976). *Proofs and Refutations*. Princeton: Princeton University Press.

Mariotti, M.A., (1997) Justifying and Proving in Geometry: the mediation of a microworld, Revised and extended version of the version published in: Hejny M., Novotna J. (eds.) *Proceedings of the European Conference on Mathematical Education* (pp.21-26). Prague: Prometheus Publishing House. Online at http://www.lettredelapreuve.it/resumes/mariotti/mariotti97a/mariotti97a.html.

Maturana, H. (1988). Reality: The search for objectivity or the quest for a compelling argument. *The Irish Journal of Psychology, 9*(1) 25-82.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston VA: Author.

Pimm D. (1987) *Speaking mathematically*. *Communication in the mathematics classroom*. London: Routledge and Kegan Paul.

Reid, D. (2002) Conjectures and Refutations in Grade 5 Mathematics. *Journal for Research in Mathematics Education, 33*(1) 5-29

Reid, D. (2001) Proof, Proofs, Proving and Probing: Research Related to Proof. Paper based on a Short Oral Presentation at the Twentieth-Fifth Annual Conference of the International Group for the Psychology of Mathematics Education, Utrecht, Netherlands. Online at http://ace.acadiau.ca/~dreid/publications/proof/proof.htm

Reid, D. (1995). *The need to prove*. Unpublished doctoral dissertation, University of Alberta, Department of Secondary Education.

Sutherland, R., Olivero, F. & Weeden, M. (2004) Orchestrating mathematical proof through the use of digital tools. In Marit Johnsen Høines &Anne Berit Fuglestad (Ed.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 4 pp. 265–272). Bergen, Norway.

Tall D. (1995) Cognitive Development, Representations & Proof. *Justifying and Proving in School Mathematics,* Institute of Education, London, (pp. 27-38).

Thurston, W. (1995). On proof and progress in mathematics. For the Learning of Mathematics, 15(1), 29-37.

Tymoczko, T. (Ed.), (1986). *New directions in the philosophy of mathematics: An anthology*. Boston: Birkhäuser.
Varela, F., Thompson, E., & Rosch, E. (1991), *The Embodied Mind*. Cambridge, MA: MIT Press.
Wittgenstein, L. (1963). *Philosophical Investigations*. New York: The Macmillan Company.