

## COMPELLING BELIEF -- *Proof & Mathematics Education*

Hunter Ellinger ([hunter.ellinger.org](http://hunter.ellinger.org), [ellinger@io.com](mailto:ellinger@io.com))

**Proof**, *n.*, ... the cogency of evidence that compels acceptance by the mind of a truth or a fact, ... the process or an instance of establishing the validity of a statement, esp. by derivation from other statements in accordance with principles of reasoning, ... something that induces certainty or establishes validity ... --

**Prove**, *vb.*, .[archaic] to learn or find out by experience

-- *Webster's* (1987), pp. 942 & 948

The diversity of emphasis shown in the dictionary entries above is reflected in the scholarly discourse on the appropriate role(s) of proof in mathematics education. The differences of perspective are rooted in a variety of disagreements: about the nature and role of proof within mathematics, about the nature of learning, about the utility of deduction in supporting learning in other areas, about the legitimacy of informal methods of argument, and about which types of students should be expected to reach which levels of competence in deductive proof.

There are some areas of agreement: no one reviewed here defends the desirability of rote memorization or denies the utility of exploratory informal methods. There are also some surprising general omissions: for example, the legitimate bases for student skepticism of formal proof are seldom acknowledged, even by authors who have little use for formalism, and there are no references to cognitive-science research on how deduction is carried out by typical people.

The disputes in this area are also expressions of a fundamental and longstanding split within the mathematical community about the relative importance of insight (the ability to accurately perceive mathematical truths) and rigor (the ability to connect mathematical statements to each other by prescribed rules). While these goals are not mutually exclusive, very few mathematicians assign equal value to them -- one is seen as being justified largely by its services to the other. This difference in perspective shows through into the curriculum: Is the main reason that axiomatic geometry is taught for students to develop geometrical insight or to master the process of deductive proof? Note that this is not an issue of understanding versus performance, but rather a question of what the topic is for which understanding is sought.

Even when proof is the focus, there are different perspectives on its primary function. This is true even within mathematics research, with Lakatos' (1976) brilliant exposition *Proofs and Refutations* giving both an illustration and a history of the use of proof as a generative method of discovery and invention rather than the embalming fluid in which theorems are put to rest. And if respect for informal methods and exploratory cumulative development of not only proofs but of the theorems and definitions at which they are aimed is appropriate for research mathematics, a flexible and constructive orientation to proof in mathematics education seems even more sensible.

### So What is a Proof, Anyway?

There is general recognition in the papers reviewed that several different levels or kinds of proof are possible, and that all might be useful under certain circumstances. De Villiers (1999, pp. 11-12) and Senk (1982, p. 2) report on a 5-level schema for proof developed by the van Hiele: observation, disconnected reasoning, short deductive chains, axiomatic/logical proofs, and metalogical analysis. Senk (1982) reports on testing data that indicates moderately-good alignment between van-Hiele levels and performance, and the distinctive nature of the critical level 3 (short deductive chains) is supported by research in cognitive science (Rips, 1995, pp.

297-343). Note that this classification scheme suggests a hierarchy, with the implication that the instructional goal is to move people to higher levels; it is thus fully consistent with a formalist stance.

Most of the authors go further and distinguish different *uses* of proof as well as different levels. De Villiers (1999) has the most comprehensive list of uses: explanation, discovery, verification, challenge, systemization, and communication to others (pp. 3-5). He recommends that students be directed to explanation and discovery prior to verification and systemization. Other authors (Henderson, 1996; Hanna & Jahnke, 1993) de-emphasize or abandon the goal of formal logic and see production of a *convincing explanation* as the definitive element of proof. Hersh (1993), while defining proof as "a complete explanation" (p. 398), also approvingly quotes G. H. Hardy as describing the proof process as *pointing* to the truth at issue (pp. 390-391). In that view, what compels belief is perception of the mathematical reality pointed to, not the proof.

That spirit is also evident in the commentary of Liu Hui (3rd. century) on the Chinese classic mathematical compendium *Jui Zhang Suan Shu* (Nine Chapters on the Mathematical Art) as reported in a fascinating article by Siu (1993). While non-axiomatic, the quoted problems show a full appreciation of proof (providing several different approaches in some cases to show "the consistency [of the theorem] with the body of accepted mathematical results", p. 352) without letting the proofs, which are often simply instructive diagrams, divert attention from the results. Liu carefully refrains from advancing mere conjectures, although he identifies and analyses some unsolved problems (pp. 353-354). Among the proofs he does include (pp. 347-349) is a delightful one that the area of a circle is "half the perimeter times half the width" (by showing that this is exactly true for any regular polygon with an even number of sides, thus including cases indistinguishable from a circle). The book's levelheaded approach is shown in a quote Siu provides (pp. 355-356) from Liu's preface to the commentary:

*Things are related to each other through logical reasons so that like the branches of a tree, diversified as they are, they nevertheless come out of a single trunk. If we elucidate by prose and illustrate by pictures, then we may be able to attain conciseness as well as comprehensiveness, clarity as well as rigor. Looking at a part we will understand the rest.*

This could be taken as premodern naivete, or alternatively as a voice of sanity from a more mature tradition which has dealt with formality and pedantry in enough contexts not to be intimidated by either.

### **What do proofs accomplish?**

A major element in the antiformalist rebellion is a set of arguments that formalism is (and will always be) incapable of delivering the certainty it claims. Henderson (1996) uses the self-proclaimed freedom from meaning of formal mathematics to assert that its results will never provide certainty to any actual question, since there is no way to ensure that the theory applies to a particular situation (p. 12). Hersh (1993, pp. 392-395) attacks even the correctness of formal proof, pointing out that the long chains of reasoning required for rigor give increased opportunities for undetected errors. Hanna & Jahnke (1993), whose comprehensive article seems to me to be consistently on target, develop Lakatos' theme, pointing out that the theorems themselves and the meanings (and thus the definitions) of the mathematical terms involved will evolve in the course of investigation (pp. 427-429). All these criticisms apply even to research work; they are so much the more cogent for teaching.

There is one accomplishment of proof to which all the authors would subscribe: *proofs connect statements to one another*. For strict formalists (Tall [1992] is the closest among these

authors, although he is willing to make extensive concessions in classroom practice to increase accessibility), the term "statements" means no more than that: math has nothing to do with meaning (p. 506). But most of the authors would, like Liu Hui above, expand this into an assertion that proofs show that many meaningful truths are related and mutually reinforcing. Hanna and Jahnke (1993, pp. 427-428) discuss how successful application of a proven theorem lends credibility to the axioms from which it was derived; there is thus a sense in which *theorems prove axioms*, since what we demand of axioms is that they entail a coherent set of interesting theorems.

Such a connection-oriented view of proof gives a basis for seeing short-chain deductions not rooted in axioms sets and exact logic rules as being proofs, rather than just extracts from or approximations to proofs. Using already-accepted reasoning processes to show how a questionable proposition is implied by ones that are already believed promotes understanding even if the already-known ideas have less than complete clarity or certainty.

### **Reports From the Field**

Separately from what proofs mean to mathematicians, educators must be concerned about how students interact with them. Chazan (1993) gives a clear and useful report on an extended set of clinical interviews with high-school geometry students about their views on the relative persuasiveness, validity, and acceptability of textbook and measurement-based proofs. He identifies, and provides student comments on, two common major deviations from the standard "a proof covers all cases" position: "evidence is proof" (pp. 360-362, 368-370), in which specific cases are taken as proving a proposition for all triangles, and "proof is just evidence" (pp. 362, 371-374), in which a deductive proof was not seen as guarding against all counterexamples. When pressed, even several of the minority of students who held the standard position expressed uncertainty that counterexamples were absolutely impossible (p. 377).

Hoyles (1997) reports some qualitative findings from an extensive United Kingdom survey of 15-year-old students on proof. She found that students usually described proof as being evidence to back up the result of an investigation, a viewpoint she ascribes to the process-oriented UK curriculum (p. 11). In multiple-choice selections among various categories of proofs ("empirical, enactive, narrative, visual, or formal [one correct, one incorrect]", p. 11), students and their teachers were asked to identify both their own favorites and the choices which they thought would result in the best grade. Only 21% of students made the same choice for both, with formal proofs seen by students and teachers as much more likely to get the best grades, while narrative was the preferred individual choice in algebra proofs (p. 12). The survey went on to ask student evaluations of the correctness, generality, and explanatory power of the proofs. Hoyles reports preliminary indications that individual preference was for proofs that were both general and explanatory, but that the explanatory element was not seen as needed for the best grades (p. 12). The results above pertained to the recognition of proofs. In proofs *generated* by students, Hoyles found a systematic use of "investigation" methods, in which the critical activity is the generation of data (whether needed or not), with no use of logical reasoning (p. 13). She sees this shift "from a *formal* to a *social* ritual" as being caused by the UK curriculum (p. 14).

Senk (1982) reports on a multi-school U.S. survey of high-school geometry students that correlated proof-writing achievement with van-Hiele level and achievement on standard content. Students took tests both at entry and exit to geometry classes (p. 3-4), and the testing and scoring methods reflected several consistency safeguards and showed excellent inter-rater consistency

(pp. 4-6). The study results indicate that most students (about 70%) can do simple geometry proofs requiring only a single additional step (p. 7), indicating that they have mastered at least the basic concept and syntax of proof. Successful performance on longer proofs is much more limited. Even at the end of a year of instruction, only about half the students can do even simple multi-step proofs (pp. 7-8). There was substantial variation in performance among the successful half, indicating that mastery of deductive proof requires more than a single insight (p. 8). The study also found that proof-writing achievement showed strong school-dependent effects and minimal sex-related differences (p. 8). Van Hiele levels correlated reasonably well with proof-writing abilities (pp. 9-10).

### Assessment of the Cited Literature

Despite the substantial variation in perspective among these authors, this review found little disagreement about the nature of the different kinds of reasoning or about the types of difficulties encountered by students in dealing with deductive proofs. There was also substantial agreement about the advantages of using informal methods to introduce students to the subject. The primary area of difference is in whether nonaxiomatic methods are valuable in themselves or mainly as a stepping-stone to formalism. This is really a policy (or philosophical, or religious) question rather than a scientific one, although I think that the antiformalists have much the better of the argument, with Hanna & Jahnke (1993) providing a particularly comprehensive and balanced view of how proof grows out of meaningful application (esp. pp. 428 & 432-434).

I do feel that the reported student responses would have benefited from a closer analysis from perspectives more willing to examine the logical weaknesses of the curriculum. Chazan (1993) does an excellent job of eliciting the relevant student thinking, but stops short of a critique of the math-class assumptions. I think that such a critique would find that many student "misconceptions" about proof merely reflect sensible caution in the unbounded extrapolation of ideas or sensible assessment of the generality of the particular cases used as examples. Take for example the theorem Chazan investigates: *the line joining the midpoints of two sides of a triangle is parallel to the third side*. The truth of this theorem is based on the flatness of the surface in which the figure is imbedded—it would not be true on a sphere, for example. As with all assertions of parallelness, it requires a reckless assertion that the lines will not intersect even in their infinite straight-line extension (note that Euclid himself put off using parallels as long as he could). To be completely confident of the parallel postulate (and thus of the proof), a student must have either successfully emptied her mind of the meaning of "parallel" (a budding formalist!) or, more likely, be directing her confidence not to the truth of the theorem but to what is being expected by the teacher. (Substantial confidence is reasonable, since the world does seem to be quite flat.) And of course confidence can spring from sources other than correctness: I notice that all the students who were absolutely steadfast in their certainty about the proof were male (pp. 376-378); it would have been instructive to try some false proofs (cf. Hoyles, 1997, pp. 11-12) on this group.

The "overgeneralization" that sees the theorem as proved after being told that it had been found to be true in four dissimilar illustrated cases could also be more sensible than it first seems. A triangle is a simple object, and it is almost impossible to contrive a straightforward property that is true for several distinctly-different triangles but not true of all triangles. While these students clearly were not using the "argument that explains" meaning of proof, they may have used simplicity-based axioms of their own that have more meaning to them than the parallel postulate and its derivatives. These personal axioms might even be more general, since they could more easily extend to non-flat surfaces. The students who were confident in this "proof by

example" did not assert certainty; they may, as Chazan suggests (p. 384), have been skeptical about the absolute dependability of any method.

Some final examples of the how student misconceptions may be based in things other than student naivete or inappropriate transfer are derived from Tall (1992), whose article addresses student concepts of functions, limits, and infinity as well as proof. What I found striking about several of his examples of immature thought is that they were not "natural" or "simple" concepts at all, but instead seemed to be the product of clumsy prior mathematics instruction. The existence of limits (pp. 501-504), for example, is an inevitable (and easy to demonstrate) consequence of any bounded infinite set. There is no need to introduce such oxymora as "varying numbers". The cognitive complications arise in part because of the introduction via *sequences* rather than the simpler concept of *sets*, and because of the unnecessary contortions in the definitions to rule out the possibility that a set can have multiple limits. The discussion of infinity (pp. 504-506) is also a compendium of ideas which are incompatible with each other but share the characteristic of replacing concepts constructable from natural thinking patterns by ones which hijack the names of those patterns to label elements of formal rules (sometimes modifying them with a qualifier such as "nonstandard", which adds to the confusion while maintaining deniability). To be fair, this is no doubt an effective strategy for driving meaning out of mathematics, if that is the goal, as is implied by the approving reference (p. 506) to Hilbert's quote: "One must be able to say at all times—instead of points, straight lines, and planes—tables, chairs, and beer mugs." Tall does give a reasonable and insightful exposition (pp. 507-509) of approaches to proof which are effective prior to reaching the formalist level, showing a clear awareness of the dangers of ignoring student understanding, but this "stops short" of professional standards as he sees them (p. 507).

### **Implications for Curriculum and Instruction**

Ironically, the success of formalism in monopolizing mathematical respectability (despite any intuition and informal tests researchers use behind closed doors) has removed much of the societal motivation for including formal proof in the mathematics curriculum. Geometry was of greater educational use when its proofs were good examples of the application of normal thought methods to discerning truths and connections between meaningful mental objects that were closely related to perceived patterns. What was missing in rigor was gained in relevance, and in the potential for generalization to other domains. The ornate axiomizations of formal logic tell students that they need to abandon their own thinking methods and to apply a set of non-obvious authority-supplied rules without being distracted by the meaning of either the rules or the objects to which they apply them. This is a lesson that will be "more honour'd in the breach than the observance" (Shakespeare, ~1600).

Such breaches of dishonourable custom would be abetted by curricula based on careful mutual alignment of the descriptions of natural cognitive abilities and professional-grade math concepts. Work in this area will have to come from those of us who see this as a task from which both education and mathematics will benefit, of course: it will seem nonsense to formalists, and may well turn out to be beyond our capabilities. But knowledge of both novice and expert cognition is still too sketchy (and is changing too fast) to rule out possible successes.

An example of a new front that may provide reinforcements to this crusade is information and computer-programming theory. Much of the knowledge (both theory and technique) that is gained in trying to educate computer-based systems may be relevant to cognition and instruction for humans. Of particular promise are the current efforts to recast single-serial-processor

concepts and processes into massively-parallel-network ones. It will be no surprise if this effort produces breakthrough illuminations on both the abilities and the confusions of the components and structures of human society. Educators should make sure that they take advantage of any opportunities such efforts create.

### References

- 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.
- de Villiers, M. (1999). *Rethinking proof with the Geometer's Sketchpad*. Emeryville, CA: Key Curriculum Press.
- Hanna, G. & Jahnke, H. N. (1993). Proof and application. *Educational Studies in Mathematics*, 24(4), 421-38.
- Henderson, D. (1996, June). Alive Mathematical Reasoning. Plenary address to the annual conference of the Canadian Mathematics Education Study Group, Halifax, Nova Scotia. (ERIC Document Reproduction Service No. ED 425 054)
- Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics*, 24(4), 389-399.
- Hoyles, C. (1997). The Curricular Shaping of Students' Approaches to Proof. *For the Learning of Mathematics*, 17(1), 7-16.
- Lakatos, I. (1976). *Proofs and Refutations*. Cambridge, UK: Cambridge Univ. Press.
- Rips, L. (1995). Deduction and Cognition. In E. Smith & D. Osherson (Eds.), *An Invitation to Cognitive Science, 2nd Ed.*, 297-343. Cambridge, MA: MIT Press.
- Senk, S. (1982, March). Achievement in Writing Geometry Proofs. Presented at the American Educational Research Association annual meeting, New York, NY. (ERIC Document Reproduction Service No. ED 218 091)
- Shakespeare, William (~1600). *Hamlet*.
- Sui, M. (1993). Proof and pedagogy in ancient China: Examples from Lui Hui's commentary on Jui Zhang Suan Shu. *Educational Studies in Mathematics*, 24(4), 345-357.
- Tall, D. (1992). The transition to advanced mathematical thinking: Functions, limits, infinity and proof. In D. Grouws (Ed.), *The Handbook of Research on Mathematics Teaching and Learning*, 495-511. New York: Macmillan.

*Q: How many legs does a dog have, if you call its tail a leg?*

**A: Four. Calling it a leg doesn't make it a leg.**

— Folk wisdom on the limits of formal definition