Csaba Csíkos

The nature and development of proving ability

PhD theses

University of Szeged

doctoral program: „Development of cognitive competence"

Szeged, 2000

Addressing the problem

Proofs are used in mathematics, in philosophy, in jurisprudence, and in everyday life. The basic assumption of this investigation is that there exist a system of cognitive components that allows for judging and constructing proofs. If the process of proving of a given statement depends solely on having known the truth value and the proof itself, some things would be contradictory. We may realize that proofs constructed by the same individual but on different domains are quite similar to each other. It is also evident that we can judge other people's proofs even if we are unfamiliar with the topic. The system of cognitive components that may allow for judging and constructing proofs can be defined as proving ability. The aim of this study is to describe the nature of proving ability

Theoretical bases for the nature and development of proving ability

In order to characterize proving ability it is necessary to construct a pedagogical-psychological concept of 'proof'. Three different interpretations of proofs can be found in the dictionaries: philosophical, mathematical and jurisprudential.

It is about 2000 years since a hierarchical system of authoritarian, empirical and deductive proofs existed. In Thomas Aquinas' opinion - that is based on Boëthius' remark - authoritarian proofs are of little value from a scientific point of view. Deductive proofs are more valuable because they make it possible to go beyond human experience. A fundamental question about deductive proofs is the truth and validity of axioms on which deductive proofs are based. Descartes pointed out that the ways of mathematical thinking are not necessarily the same as the ways of publishing the results in the classic definition-theorem-proof sequence.

The problem of the vicious circle of theorems and axioms has been solved by Hegel: the axioms, in fact, are not starting points in constructing proofs, but they can often be the result of scientific development. In several cases the axioms can be proven in another system of scientific facts. Philosophers from Vienna made further steps towards an appropriate definition of the proving process. The core of their verification theory: The meaning of a sentence is the same as the process of stating the truth value of the sentence that is the mode of verification. Hans Hahn emphasizes the importance of logic applied in the proving process. His idea is that a statement can implicitly contain lots of other statements; the crucial role of logic is to make clear what have been stated. In the second half of the 20^{th} century, Popper and Lakatos made great contributions to the development of a philosophical concept of proof.

In mathematics, the development of the concept of proof spans over centuries. Till the 19^{th} century, mathematical proofs were of psychological nature: to make something clear or to demonstrate the truth of a theory by means of citing formerly proven statements and facts (Tarski, 1990). The important and characteristic elements of formal mathematical proofs (i. e., definitions, axioms, postulates, theorems) were consistently used for the first time in Euclid's 'Elements' (about 300 B. C.). In our century, mathematical proof theory has set as the aim to investigate the consistency and being free of contradiction of axiomatic systems - without using set theoretical tools. Proofs in mathematics are demonstrations in order to reveal the deducibility of a statement in a finite number of steps. Deductive-axiomatic proof theory has to cope with challenges from different branches of mathematics (constructivist mathematics, experimental mathematics, computer-associated mathematical proofs).

In jurisprudence, the basic principle is that proofs must be deductive in the scientific sense of deduction. Nevertheless, in jurisprudence, there is a great emphasis on inductive processes in the phase of gathering information.

From a psychological point of view, we emphasize the difference between the sequence of reasoning processes and the sequence of publishing the results of proving. Consequently, lessons from mathematical concept of proof may help in constructing a pedagogical-psychological concept of proof that can be applicable in non-mathematical domains as well. Therefore, proofs can be considered as the results of processes that aimed to demonstrate the truth value of a statement.

A pedagogical-psychological concept of proving ability must take into account the results of ability research. We have several choices (i. e., factor analytical concept of ability, abilities of experimental psychology), and we decided to embed proving ability in a multi-level, hierarchical concept of cognitive abilities. Three level models of human thinking that go beyond the factor analytical approach have appeared in the last 2-3 decades. In Sternberg's triarchic model (1988), the term 'white-collar components' refers to the meta-components of human thinking. These components plan, monitor and evaluate the functioning of lower-level ('blue-collar') components. The work of blue-collar or executive components can be characterized as algorithmic processes when solving problems.

As to proving ability, we speak about - using in part Marr's (1982) terminology - hardware-, algorithmic and strategy-level. This triarchic system is similar to that of Nagy's (1998) who describes neural, implicit experimental, implicit conceptual and explicit decision levels as hierarchical strata of thinking: The strategy-level of proving ability contains Moshman's (1990) meta-logic and Johnson-Laird's meta-deduction (see Johnson-Laird and Byrne, 1991). The algorithmic level contains forms of reasoning described by classical experimental psychology, among them it is deductive reasoning that may have great importance when constructing or analyzing proofs. In this investigation we considered the hardware-level of proving ability as functioning normally in every subject.

Investigating the development of proving ability requires setting up an external system of criteria in order to compare individuals to each other and to the external criteria. It can be a hard task for researchers to find a valid system of criteria; nevertheless, cultural differences must be taken into account. If we borrow the authoritarian-empirical-deductive order of development from philosophy, we must be strong to reveal that deductive proofs are the most valuable not because of being the last in the order of appearance, but because of being the most difficult to achieve. Difficulties with deductive proofs can be in part derived from the late development of logical necessity in human reasoning.

Whatever kinds of abilities are studied, the basic principle of psychometrics should be respected: The development of an ability can be determined through the difficulties of problems solved successfully by the individual. Human performance on cognitive task is influenced by the context of the problem-solving situation. The problem of contextual effects can be summarized as follows: Two tasks that are identical or very similar to each other even in the content of the tasks may be of different difficulty because of the context. The contextual effects may cause significant differences in performance.

Fostering proving ability requires getting acquainted with the results of mathematics education. Consequences derived from training on proving skill or ability can be generalized to other domains as well. One important principle emphasizes the need for change in the traditional theorem-proof-definition order to a reversed 'proof-theorem-definition' sequence. Another basic principle is using metacognitive tools. One example may be Pólya's famous list of metacognitive questions.

Taking the view-point of educational assessment into account in this investigation we used Harel and Sowder's (1998) proof schemes that were developed for mathematical proofs but could be applicable to non-mathematical content, too. We used five main proof-schemes in this research: authoritarian, ritual, symbolic, empirical, and analytical (deductive) proofs.

Research questions

Research questions and hypotheses from the literature have to be re-formulated for the purposes of this study, since this research integrated several fields and paradigms from a new point of view. Research questions and the main hypotheses of the dissertation deal with basic theoretical questions on the nature and development of proving ability.

1) A pedagogical-psychological concept of proofs can be based on the philosophical, mathematical and jurisprudential concepts of proof. The pedagogical-psychological concept of proof must address the cognitive processes required for judging and constructing proofs. We must emphasize that from the beginning of our research we consistently avoided restricting the concept of proving ability to mathematical proving skills.

2) The hierarchical-multilevel model of proving ability must be based on the pedagogical-psychological concept of proofs and on the hierarchical three-strata models of cognitive abilities described in several monographs in the last two decades. From a methodological point of view, we may divide proving ability into two subsystems of cognitive components: 1) ability for judging proofs, and 2) ability for constructing proofs.

3) Research methods that address the problem of assessment and evaluation of proving ability must take the multi-level nature of proving ability into account. Different levels or subsystems of components require different evaluation methods to use. The hardware level of proving ability can be assessed by means of neuro-physiological and clinical psychological methods. Measuring the algorithmic-level components of proving ability may effectively result in using tests that have been originally developed for measuring deductive reasoning. Assessing the strategic level of proving ability may in itself be defined as assessing proving ability. Albeit three-stratum models of cognition do not allow for assessing algorithmic and strategic variables separately because there is a permanent flow of information between those levels. This notion is at the same time a critique of traditional methods of measuring deductive reasoning. We may also note that psychometrics has revealed the phenomenon of positive manifold, that is, achievement in tasks on domains that are thought to be very far from one another can be similar.

4) Describing the development of proving ability we are inevitably conflicted with the evolutionary approach of human reasoning. In that sense, development of proving ability can be described as increasing and

decreasing frequency-in-use of different proving schemes. Some schemes are more frequently used, some are not, depending on the process of enculturation and development of other reasoning processes.

5) The development of proving ability is determined by the development of its sub-components. For example, the development of logical necessity - which may play a crucial role in judging and constructing deductive proofs - is very slow till the age of 12.

6) Information given to students about the value of their proving schemes influence the development of proving ability can to a great extent. In accordance with the evolutionary approach, it can be hypothesized that the development of proving ability is a process mediated by internal selection (or the process of being selected) of proof schemes.

There have been a number of hypotheses formulated during the process of data analysis:

- In which grade is there a big change in students' opinion about the value of authoritarian proof?
- Why do so many students overvalue the symbolic proofs?
- How does the familiarity of the content influence the proofs scheme constructed in that domain?
- How can we construct an index of proving ability on the basis of Likert scale results?
- How can the nature of correlations among proving ability and several background variables be characterized?

Methods

Within a larger investigation called 'Development of mathematical abilities' six tests were administered to 2572 students, in 3 counties in Hungary, between April and May, 1998. The sample consisted of children of the 5^{th}, 7^{th}, 9^{th}, and 11^{th} grades (with ages ranging from 11 to 17 years). There were two additional questionnaires assessing personal data, school marks, and mathematics and physics academic self-concept. The tests were developed for this study, and two of them were previously piloted.

The second pool of data (called the 'large scale' investigation) was taken in May, 1999. Two tests of proving ability had been developed for this study. The first one consisted of mathematical statements each with its proofs of 5 types: authoritarian, ritual, symbolic, empirical and deductive. Students were required to judge in a five-point Likert scale the value of proofs. The second test of proving ability consisted of both closed and open ended questions from various content domains, using students' answer types that had been observed a year before.

A questionnaire was administered to the mathematics teachers of the schools that had been involved in the project in both the pilot and the large scale investigation. The questionnaires measured teachers' judgments about students' most frequently observed proof types (pilot study) and the five proof schemes: authoritarian, ritual, symbolic, empirical, and analytical (large scale study).

From a methodological point of view, using non-parametric tests, multidimensional scaling and path-analysis might be subjects of special interest, because these statistical methods are rarely used in Hungarian educational research.

Results

Results of the pilot study

A dichotomous categorization system has been developed for each task, by which both hierarchical and non-hierarchical evaluations of students' proofs can be performed. The nominal categories of this dichotomous system can serve as a basis for an hierarchical evaluation of proof types: An ordinal scale measure can be developed based on Harel and Sowder's proof-categorization system. Three hierarchically ordered stages can be identified in their model: 1) externally-based proofs, 2) empirical proofs, and 3) analytic proofs. The categories of our dichotomous system were transformed into ordinal scale categories on the basis of agreement among experts.

Three expert raters independently recoded the nominal categories using an ordinal scale derived from Harel and Sowder's taxonomy. Kendall's coefficient of concordance ($W=.909$, $p<.001$) indicates a high level of agreement. In each case it was possible to recode the nominal categories into ordinal scale in 3:0 or 2:1 rate. The results suggest that there is a tendency to construct higher-order proofs even in non-mathematical domains as a function of school grades. Spearman-correlation coefficients suggested that there were significant relationships between proof types of different domains.

Mathematics teachers' judgments can be characterized as giving too much appreciation to symbolic proofs whereas undervaluing empirical proofs. A source of students' misconceptions about the nature of mathematical proofs can be traced back to their teachers' judging proof schemes.

Maths teachers' undervaluing empirical proofs may result in students' mathematical academic self-concept of the lowest level in case of constructing empirical proofs. (Józsa and Csíkos, 1999). It has also been revealed that students' judging proof types can to a great extent be traced back to students' ideas about their math teachers' judgments.

Results of the large scale study

The reliability coefficients of students' proving ability tests were satisfactory. Also Kaiser-Meyer-Olkin-indices were computed in order to characterize an aspect of item-consistency.

Students' judging of different proof types can be summarized as follows:

a) Authoritarian proofs: This type of proofs is widely accepted among 5^{th} and 7^{th} graders but are strongly rejected in the sample of 9^{th} and 11^{th} graders.

b) Ritual proofs: There can be revealed significant differences between both age-groups, and - within the 9^{th} and 11^{th} graders' groups - between grammar school and vocational secondary school students.

c) Symbolic proofs. In every age-group symbolic proofs obtain higher grades than authoritarian and ritual proofs. This may be due to the fact that in a mathematical context the appearance of mathematical symbols in itself may increase the value of a proof.

d) Empirical proofs. There were quite large differences between empirical proofs of different domains. This can be traced back to the difference in the statements from the aspect of containing or not containing universal quantifier or to the difference according to Balacheff's (1988) naiv empiricism and crucial experiment.

e) Deductive proofs. In every domain and also in every age-group deductive proofs yield the best scores. This means that students even before 5^{th} grade are taught to value a certain type of proof. This 'certain' is in our culture the type of deductive proofs.

The difference between age-groups can be characterized as follows:

– Elementary school pupils score to a greater extent medium or high scores to the external proof types.

– Differences between 9^{th} and 11^{th} graders can often be traced back to differences in the content.

– Students of vocational secondary schools are in the middle between elementary and grammar-school students when judging proof types.

Connections between content-familiarity and proof types in case of open-ended tasks has become very clear. If the content is familiar, many students feel courage to construct a deductive proof, in spite of the fact that a correct deductive proof would require knowledge about facts and rules that are not yet known.

The index of proving ability developed from Likert scale results has proven to be a reliable and effective measure of the development of proving ability. The curve of the development is similar to the upper branch of normal ogiva. This observation supports the hypothesis that proving ability can be considered to be an ability in a psychometric sense of the term 'ability'.

Students' judges of proof types is a subject of slow changes. From this aspect, it is similar to those of scientific misconceptions. Math teachers are, however, optimistic in the question of changes, and they are also in accordance with each other in the question about age-limits of certain processes in the development of proving ability.

Summary

The pedagogical-psychological concept of proofs and proving ability developed in this investigation is based on proof-concepts developed by philosophy, mathematics and jurisprudence. Our concepts are general from an aspect of not restricting the content to be proven to mathematical theorems, and the concepts are specific from an aspect of formulating hypotheses that are relevant and interesting for pedagogy.

Being in accordance with results from ability research, proving ability is a three-level, hierarchical system of cognitive components that allow for judging and constructing proofs of a given statement.

Several tests have been developed for assessing proving ability, and they proved to be valid and reliable.

Concerning the development of proving ability we emphasize two features: 1) the system of correlations among test items may be an indicator of the development, and 2) using a one-dimensional score to measure proving ability, the inflexion point of the developmental curve is around age 13.

Mathematics can play an important role in fostering the development of proving ability. However, as the results suggest, maths teachers' judgments do not necessarily correspond to the hierarchical stages of Harel

and Sowder's model. This means that from math teachers' view-point there is a positive bias towards symbolic proofs whereas there is a tendency to undervalue empirical proofs.

Practical considerations about the results of the present investigation may involve emphasizing the importance of 'exploring the territory' (Edwards, 1997) before proving a statement. Statistical tendencies revealed by cross-sectional comparisons will give advice to text-book and curriculum writers.

REFERENCE

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

Edwards, L. D. (1997): Exploring the territory before proof: Students' generalizations in a computer microworld for transformation geometry. *International Journal of Computer for Mathematical Learning* 2., 3. sz. 187-215.

Harel, G. és Sowder, L. (1998): Students' proof schemes: Research from exploratory studies. In: Dubinsky, E., Schoenfeld, A.és Kaput, J. (Eds.): *Research Issues in Collegiate Mathematics Education* Vol. 7.,American Mathematical Society, Washington, D. C., 234-283.

Johnson-Laird, P. N. és Byrne, R. M. J. (1991): *Deduction*. Lawrence Erlbaum Associates, Hillsdale, New Jersey - Hove and London.

Józsa Krisztián and Csíkos Csaba (1999): *The relationships between mathematics self-concept and cognitive abilities required for mathematics achievement*. Paper presented at the 8th European Conference for Research on Learning and Instruction, Gothenburg, Sweden, 24-28, August.

Marr, D. (1982): *Vision*. Freeman, New York.

Moshman, D. (1990): The development of metalogical understanding. In: Overton, W. F. (Ed.): *Reasoning, necessity, and logic: Developmental perspectives*. Lawrence Erlbaum Associates, Hillsdale, New Jersey - Hove and London, 205-225.

Nagy József (1998): A kognitív képességek rendszere és fejlődése. *Iskolakultúra*, 10. sz., 3-21. [in Hungarian]

Tarski, A. (1990): *Bizonyítás és igazság*. Gondolat, Budapest. [in Hungarian]

Proof categories in Harel and Sowder's model

(source: Harel, G. and Sowder, L. (1998): Students' proof schemes: Research from exploratory studies. In: Dubinsky, E., Schoenfeld, A.és Kaput, J. (Eds.): Research Issues in Collegiate Mathematics Education Vol. 7., American Mathematical Society, 234-283.)