12th International Congress on Mathematical Education
Topic Study Group 14
8 July – 15 July, 2012, COEX, Seoul, Korea

AN INVESTIGATION OF SENIOR MATHEMATICS AND TEACHING MATHEMATICS STUDENTS' PROOF EVALUATION PRACTICES

Yesim Imamoglu — Maltepe University — yesim.imamoglu@gmail.com

Aysenur Yontar Togrol — Bogazici University — yontar@boun.edu.tr

This study was conducted with 82 senior students from Mathematics and Teaching Mathematics Departments in a state supported university in Istanbul, aiming to investigate their proof evaluation practices. Proof Evaluation Exam (PEE) was developed for this purpose. PEE responses were categorized by examining whether the students think the provided argument proves the statement is true for all cases, true for some cases or does not prove the statement, and scored according to their accuracy of evaluating the given arguments. Seniors are generally aware of the necessity of generalizing their results and attempt to use procedures involving deductive reasoning. Nonetheless they still have difficulties in evaluating proofs. Significant differences are observed between departments. Implications for teaching and further studies are discussed.

Mathematical proof, proof evaluation, reasoning

INTRODUCTION

Proof plays an important role in mathematics education, and many studies have been conducted in recent years investigating teachers' and students' understanding of proof across all grades (e.g. Knuth 2002; Miyazaki 2000; Morris 2000, 2002; Healy and Hoyles, 2000; Weber, 2001; Hoyles and Küchemann, 2002; Stylianides and Stylianides, 2009). Most of these studies report that students have a poor understanding of proof and have difficulties in constructing their own proofs.

In order to examine students' understanding of proof, some studies have been conducted where the participants were asked to evaluate the validity of mathematical arguments; decide whether they can be accepted as proofs. Healy and Hoyles (2000) presented the participants of their study some mathematical conjectures and a set of arguments supporting them. Students indicated which argument was nearest to their own approach and which argument they believed would get the best mark from the teacher. They were also asked to assess the validity and explanatory power of the arguments. Finally students were given two conjectures to construct their own proofs. While most students used empirical arguments in their proofs, they recognized that it would not receive high marks from their teachers. They were aware that a valid proof must be general. Another finding is that students preferred arguments presented in words as choices of their own approaches and found them explanatory, whereas arguments containing algebra were less popular and found to be hard to understand. Students who used narrative form in their own constructions were more successful than students who attempted to use algebra. It is concluded that students held two conceptions of proof simultaneously: those about arguments they thought to would receive the best mark (containing algebra) and those about arguments they would use themselves (narrative from). In another study, Almeida (2000) found out that students publicly declare their agreement with the notions of formal mathematical proof; they appear to prefer or do not reject informal/visual methods of proving.

One of the main difficulties students have regarding proofs is distinguishing inductive arguments from deductive ones. Morris (2007) examined pre-service mathematics teachers' ability to distinguish logical deductive arguments from other forms of arguments. Participants were given a transcript of a classroom lesson where the students tried to prove a mathematical generalization, and were asked to evaluate the validity of the students' arguments. Students rarely used logical validity as a criterion for evaluating arguments, and exhibited a wide variety of conceptions about the relationships among mathematical proof, explaining why something is true in mathematics, and inductive arguments; and these conceptions affected their evaluations of students' arguments. Many of the participants were able to distinguish between student responses that did and did not explain why a generalization was true. However, they used their own knowledge to fill in holes in students' arguments which led to inappropriate evaluations of students' arguments and understanding.

Studies reported above indicate that students across all grades have difficulties in understanding and evaluating proofs. This study was conducted to investigate the situation in a state university in Istanbul, Turkey. The aim is to observe proof evaluation practices of students who are about graduate as candidates to teach mathematics to future generations. As prospective mathematics teachers and mathematicians; they are important figures who will shape high school and university students' mathematical conceptualizations in the future. Therefore, this study is an important step for understanding and comparing mathematicians' and prospective mathematics teachers' proof practices at the time of finishing their programs. Clarification of these proof practices will be helpful in developing instructional implications for teaching mathematics programs.

The research questions that have been posed in relation with the aims of the study listed below:

- How do senior Mathematics, Primary and Secondary Education Teaching Mathematics students decide what constitutes a mathematical proof, when they are asked to evaluate freshmen students' mathematical arguments?
- Are there any differences between senior students from Mathematics, Secondary Education and Primary Education Teaching Mathematics Programs, with respect to their proof evaluation practices?

METHOD

Sample

The participants consist of senior students from the following departments: Primary Education Teaching Mathematics-PRED), Secondary School Teaching Mathematics-SCED) and Mathematics - MATH. Numbers of participants in each group are given in Table 1.

Table 1: Sample characteristics

	MATH	PRED	SCED	Total
Female	13	15	12	40
Male	10	15	17	42
Total	23	30	29	82

Instrument and Data Collection

Proof Evaluation Exam (PEE) was developed in order to collect data about students' proof evaluation practices. PEE includes four mathematical statements which need to be proved (or disproved). For each statement, alternative arguments were given as proofs. These arguments were chosen from data collected for a larger study (İmamoglu, 2010), using an instrument called the *Proof Exam* (PE), where freshmen students from the same departments were asked to prove (or disprove) the same mathematical statements. They range from empirical-inductive to formal-deductive forms, similar to the selection process of Healy and Hoyles (2000). The approach taken during item development was to use student generated arguments similar to the study of Selden and Selden (2003) rather than partly or all expert generated ones, as used in the studies of Stylianides et al. (2004) and Healy and Hoyles (2000); because student generated arguments are more authentic, they better represent the type of arguments the participants will have to make sense of as mathematics teachers/instructors in the future.

To develop the PE, initially, 13 mathematical statements were produced / selected, examining related studies (Almeida, 2000, 2003; Miyazaki, 2000; Morris, 2002; Selden and Selden, 2003; Stylianides et al., 2004, 2007; Healy and Hoyles, 2000) and typical examples that can be found in books about methods of mathematical proof (Cupillari, 2001; D'Angelo and West, 2000; Solow, 2005). Content knowledge required for the items were aimed to be kept at minimum, so that the participants' reasoning would not be obstructed by the lack of knowledge in a certain mathematical subject. Content covered by the items is included in the high school curriculum (MEB, 2005). All types of proof methods mentioned in the curriculum for grade 9, subject of logic (MEB, 2005) are covered by the items. No other proof method is needed, though may be used by the participant. Attention was also paid to include items that can be proved in several ways using alternative proof methods. These items were examined by experts to determine which are most suitable for the target population (level of students) and aim of the study. Adjustments were made accordingly.

For each alternative argument (proof) in PEE, participants were asked to choose one of the following: “A. The proof shows the statement is true for in some cases”, “B. The proof shows the statement is always true”, “C. The proof is false”, “D. I have no opinion”. They were also asked to explain their choice.

Data were collected separately for each department. PEE was administered to participants in paper-pencil form during lecture hours. The lecturer of the course and the researcher were present during data collection.

DATA ANALYSIS AND RESULTS

In order to form the rubric, first the instrument was given to three experts who were working in mathematics department. Their responses to the questions were used for the development of the rubric and the following criteria were used in scoring:

- Wrong choice (A or C) without any explanation or incorrect explanation: 0 points
- Wrong choice but reasonable explanation or correctly indicates a mistake or a missing step: 1 or 2 points
- Correct choice without any explanation: 1 point (for A and C), 3 point (for B, if the given response is a full proof)
- Correct choice but insufficient or irrelevant explanation: 1 or 2 points
- Correct choice with sufficient explanation: 3 points

Final version of the rubric was also examined and approved by an expert (associate professor from mathematics department).

Due to space restrictions, detailed analysis of only two items (item 1 and item 2) are given below.

The first item of PEE was: “If the square of a natural number is even, then that number must be even”. The participants were asked to evaluate five alternative proof attempts for this statement.

Proof 1A was the first argument to evaluate: “If n is odd, $n = 2k + 1$, then $n^2 = (2k+1)^2 = 4k^2 + 4k + 1$ is odd (even + even + odd). If n is even, $n = 2k$, then $n^2 = (2k)^2 = 4k^2 + 4k$ even (even + even). Since n^2 is even, n must also be even.” This argument correctly proves the statement (proof by cases); there is one calculation mistake which does not affect the generality of the argument. Therefore, the correct choice here is B. While most students indicated the correct choice (63.6 %), some students concluded that the proof is false (23.6 %) or only shows the statement is true for some cases (12.7 %) because of the calculation mistake or claiming that the argument shows the converse of the statement. As a result, 60 % of participants received maximum score.

Second argument to evaluate was proof 1B: “Assume n is odd. $(2k+1)^2 = 2m$, $4k^2 + 4k + 1 = 2m$. Left hand side is odd, right hand side is even. Contradiction. This means n must be even.” This is an attempt a proof by contradiction. While the argument proves the statement, the wording can be a bit confusing, it could have been clearer. Again the correct choice is B and 61.8 % of students correctly identified it.

Third alternative proof attempt was proof 1C: "$n^2 = n \cdot n = 2k$. Here k must be even because $2k$ is a whole square: $k = 2m, n^2 = 4m$, $\sqrt{n^2} = \sqrt{4m}$, $n = 2\sqrt{m}$, hence n is even." There are missing steps in this argument; the premise "k must be even because $2k$ is a whole square" should be justified because it is the essence of the proof. It would also explain why $\sqrt{m}$ must be a whole square. 18.2 % of the students pointed out this missing step (choice A of B) and received full points.

Next argument was proof 1D: "Assume $n = 2k$. Then $n^2 = 4k^2$, which is even." This argument proves the converse of the statement. The mistake here is proving the truth of the implication q → p instead of p → q. These two propositions are not equivalent. Therefore the correct choice is C. Another correct interpretation observed in responses is that the proof is incomplete; the case where n is odd should also be checked (with choice A). Then it would be valid proof (proof by cases). Both responses received full points.

Last proof attempt for the first item was proof 1E: "Even = {2, 4, 6, 8 ...}. If $n^2 = 4$ then $n = 2$, $n^2 = 16$ then $n = 4$, if $n^2 = 36$ then $n = 6$ .... $n^2 = 114$ then $n = 12$." Here, the truth of the statement is verified for only a couple of values of n. Therefore the correct choice is A. Since there is no generalization, this cannot be accepted as a valid proof. Students who stated that giving examples is not a proof (choice C) also received full points (69.1 %). The percentage frequency distributions of evaluation scores for proofs 1A through 1E are given in Figure 1.

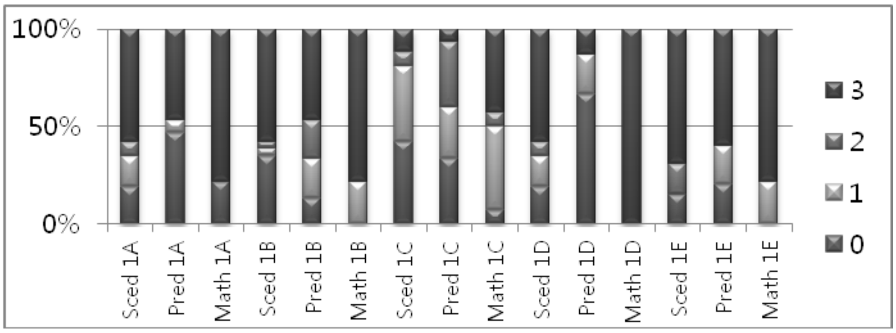

Figure 1. Percentage frequencies of scores for item 1

For the second item in PEE; "$1 + 3 + 5 + ... + 2n-1 = n^2$ holds for all integers $n \geq 1$", five alternative proof attempts were given. In proof 2A, a general formula to find sums is correctly used to verify the statement is true. However, no explanation about why this formula is true or why it can be used in this particular case is given. Students who stated that the proof would be valid if the formula was also proved received full points. In proof 2B, first the sum from 1 to $2n$-1 is calculated, and then sum of even numbers in this range is subtracted from the total to find the sum of odd numbers. This shows the statement is true for all cases, however, the fact that sum of integers from 1 to n is calculated by the formula $n (n +1) / 2$ is used without proof. 50.1 per cent of the participants gave this explanation. Proof 2C is an attempt at proof by mathematical induction. The missing step is the induction basis: Truth of the statement for all cases would be shown if it was also checked that the equality holds for $n = 1$. But since it is missing, it cannot be proved that the statement is true for any n. Hence the correct choice is C. 30.1 per cent of the participants pointed out the missing step but failed to give the correct choice. Only 12.7 per cent of the participants concluded that the missing step would make the proof invalid. One reason for this can be that usually checking that the smallest number satisfies the condition is trivial but showing that if the statement is true for n, then it would also be true for $n + 1$ is the challenging part of the proof. The argument presented in proof 2D, shows that if 1 is subtracted from each even number from 2 to $2n$, the resulting numbers give the terms of the desired sum. But, again it should be noted that in order to find the sum of even numbers, the formula $n (n + 1) / 2$, which gives the sum of integers from 1 to n is used without proof. Proof 2E is a valid proof which does not use any previously known formulas or facts. Here the terms of the sum are written in reverse order and the first term is added to the last, second term is added to the second one from the last etc. Each these sums are equal to $2n$, and if we add them all up we get $2n^2$, which is twice the sum we are looking for. Figure 2 shows percentage frequency distributions of the scores for proofs 2A through 2E.

The third item of PEE; "Given any three consecutive integers, one of them must be divisible by three" had four alternative proofs. Figure 3 below shows the percentage frequency distributions for proofs 3A through 3D for item 3.

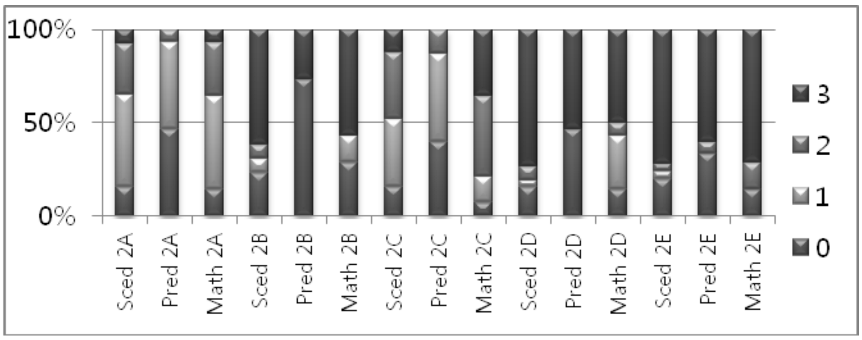

Figure 2. Percentage frequencies of scores of item 2

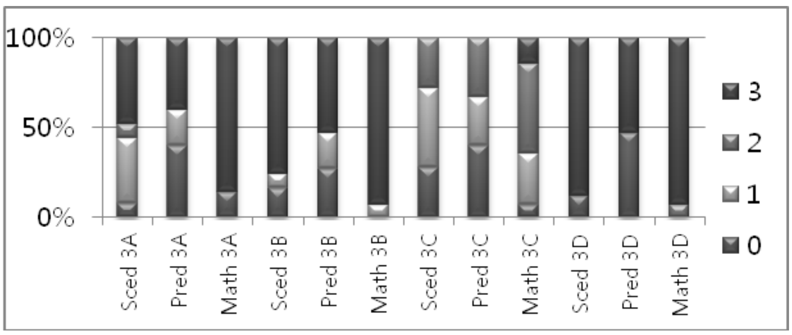

Figure 3. Percentage frequencies of scores for item 3

Last item of PEE is: "Suppose there are *n* people at a party ($n \geq 2$). Prove that there are at least two people who have the same number of friends in this party." Four alternative proof attempts are given to the participants to evaluate. Figure 4 shows the percentage frequency distribution of scores for item 4.

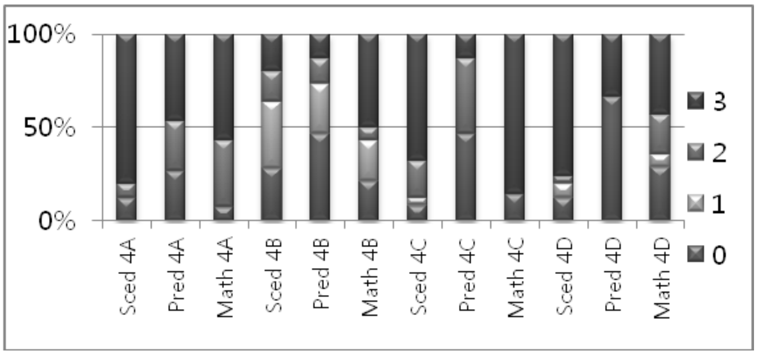

Figure 4. Percentage frequency of scores for item 4

Table 2 shows the means of total scores for each item and total PEE score. Maximum possible scores for item 1 and item 2 are 15, for item 3 and item 4 are 12. Total maximum possible score is 54.

Table 2. Mean scores and standard deviations for each PEE item

Scores		Item1 (15)	Item 2 (15)	Item 3 (12)	Item 4 (12)	Total Score (54)
SCED	Mean	9.19	9.64	7.96	8.76	35.56
	Std.Dev.	2.73	2.72	2.56	2.85	6.84
PRED	Mean	7.20	5.67	5.73	5.13	23.73
	Std.Dev.	3.36	3.52	3.35	3.27	8.55
MATH	Mean	12.36	9.57	10.07	8.64	40.64
	Std.Dev.	2.34	3.78	1.77	3.50	8.12

Shapiro - Wilk test did not reveal any significant results for any sub group, therefore normal distributions can be assumed and parametric tests are carried out.

One way ANOVA was performed for seniors' total PEE score. Results show that there are significant differences between mean scores with respect to department: $F(2, 51) = 19.11$,

$p = 0.00 < 0.05$. Post hoc analysis revealed that prospective primary school teachers have significantly lower mean score than prospective secondary school teachers ($p = 0.00 < 0.05$) and mathematicians ($p = 0.00 < 0.05$). Mean score difference between prospective mathematicians and secondary school teachers is not significant.

CONCLUSION

The results of data analysis of *Proof Evaluation Exam* reveal that most students were successful at differentiating between inductive and deductive arguments and stated that giving specific examples cannot be accepted as proof. They also were good at indicating the arguments that did not check the truth of the statement for all cases. Nonetheless, proof evaluation scores of seniors showed significant differences between primary and secondary education students, and primary education and mathematics students. Results suggest that students were better at accurately evaluating arguments that prove the statement is true for all cases; or arguments that clearly do not prove the statement, like an incorrect counter-example (the third item, for example had the following argument: -1, 0, 1 are three consecutive integers none of which are divisible by three) or giving numerical examples instead of a general proof. They do have difficulties in evaluation when there is not an obvious mistake in the argument, but some steps are missing or a crucial piece of information is given without justification (like using formulas without proof, in item 2).

It is seen from the findings that prospective mathematicians have the highest scores and prospective primary school teachers' scores are the lowest in most cases. One explanation for this situation is that prospective primary school teachers do not take as many math courses as prospective secondary school teachers and prospective mathematicians. Students in education departments enroll to mathematics courses given by the Mathematics Department, together with Mathematics students. Hence, their content knowledge is formed by the courses that they take from Mathematics Department. Comparing Primary and Secondary Teaching Mathematics Programs, it is seen that the additional courses prospective secondary school teachers take are Discrete Mathematics, Linear Algebra and Introduction to Complex Analysis. Secondary School Teaching Mathematics Program has additional six elective mathematics courses. Prospective primary school teachers also have elective course options in their program, where they may choose mathematics courses, but it is not compulsory. While one can argue that it is not necessary for primary education students to take all these courses because they will not need the content knowledge, it can also be claimed that taking more mathematics courses would help them have a more clear understanding of the mathematical process.

As prospective mathematics teachers, the courses that these students take from mathematics department are aimed to cover their content knowledge and understanding the mathematical process. Then there is the question of how to teach proof, which a matter of pedagogical content knowledge, hence the responsibility of education departments. So it is essential to design the bridge courses in these departments in a way that facilitates creating learning environments that involves proof activities. Mathematicians' proof processes involve inductive reasoning; intuition, trial and error lead to a conjecture and then formal proof practices requiring deductive methods follows. In addition, since one function of proof is convincing oneself and others, including the conjecture development in the proof process in classrooms, allowing the students to form their own conjectures and urging them to convince others that their conjecture is true may result in a deeper understanding of the subject and decrease the risk of students seeing proof as a topic to be learned, instead of as a process that is in the essence of mathematics.

Mathematical statements used in the study were chosen so that they could be proven using high school knowledge and experience; these students come across more complex mathematical tasks involving proof in their university courses. Future studies can focus on students' difficulties regarding proof used in specific mathematical topics and designing appropriate proof activities. In addition, these statements in the instruments were typical examples used, when the notion of proof is introduced in mathematics courses. So the participants were familiar with these problems. In further studies, presenting students with proof problems that they are not familiar with can give an idea whether they can transform their knowledge into unfamiliar situations. The instrument developed for this study can be used in future studies involving both high school and university students' conceptions regarding proof. Teachers in high school and instructors of introductory courses in university can also use them in classrooms as assessment tools. A toolkit for mathematics school teachers and university instructors can be developed; which can help them evaluate arguments students have generated, and can also be used to help students evaluate their own proofs.

References

Almeida, D. (2000). A survey of mathematics undergraduates' interaction with proof: Some implications for mathematics education. International Journal in Mathematics Education, Science and Technology, 31(6), 869–890.

Almeida, D. (2003). Engendering proof attitudes: can the genesis of mathematical knowledge teach us anything? International Journal of Mathematical Education in Science and Technology, 34(4), 479-488.

Cupillari, A. (2001). The nuts and bolts of proofs (Second ed.). San Diego, CA: Academic Press.

D'Angelo, J. P. and D. B. West (2000). Mathematical thinking: Problem solving and proofs (Second ed.). Upper Saddle River, NJ: Prentice Hall

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

Hoyles, C. and D. Kuchemann (2002). Students' understandings of logical implication. Educational Studies in Mathematics, 51, 193-223.

Imamoglu, Y. (2010). An investigation of freshmen and senior mathematics and teaching mathematics students' conceptions and practices regarding proof (Doctoral dissertation). Bogazici University, Istanbul, Turkey.

Knuth, E. J. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. Journal of Mathematics Teacher Education, 5, 61-88.

MEB. (2005). Orta öğretim matematik (9, 10, 11 ve 12. sınıflar) dersi öğretim program (Turkish high school mathematics curriculum; grades 9-12). Retrieved May 16, 2008, from http://ogm.meb.gov.tr/.

Miyazaki, M. (2000). Levels of proof in lower secondary school mathematics. Educational Studies in Mathematics, 41, 47-68.

Morris, A. K. (2000). Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument. Developmental Psychology, 36(6), 741-758.

Morris, A. K. (2002). Mathematical reasoning: Adults' ability to make the inductive-deductive distinction. Cognition and Instruction, 20(1), 79-118.

Morris, A. K. (2007). Factors affecting pre-service teachers' evaluations of the validity of students' mathematical arguments in classroom contexts. Cognition and Instruction, 25(4), 479-522.

Selden, A. and J. Selden (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? Journal for Research in Mathematics Education, 34(1), 4-36.

Solow, D. (2005). How to read and do proofs (fourth ed.). Hoboken, NJ: John Wiley and Sons, Inc.

Stylianides, A. and G. Stylianides (2009). Proof constructions and evaluations. Educational Studies in Mathematics, 72, 237-253.

Stylianides, A., G. Stylianides and G. N. Philippou (2004). Undergraduate students' understanding of the contraposition equivalence rule in symbolic and verbal contexts. Educational Studies in Mathematics, 55, 133–162.

Stylianides, A., G. Stylianides and G. N. Philippou (2007). Preservice teachers' knowledge of proof by mathematical induction. Journal of Mathematics and Teacher Education, 10, 145-166.

Weber, K. (2001). Student difficulty in constructing proofs. The need for strategic knowledge. Educational Studies in Mathematics, 48(1), 101-119.