



Proof construction and evaluation practices of prospective mathematics educators

Yeşim İmamoğlu^{1,*}, Ayşenur Yontar Toğrol²

^{1*} Department of Primary Education, Maltepe University, Istanbul, Turkey

² Department of Secondary School Science and Mathematics Education, Bogazici University, Istanbul, Turkey

For correspondence: yesim.imamoglu@gmail.com

Abstract:

This study was conducted with 93 freshmen and 82 senior prospective mathematicians and mathematics teachers in order to investigate how they construct and evaluate proofs and whether there are any significant differences in their proof construction (with respect to department and grade) and proof evaluation (with respect to department) performances. Instruments developed for this purpose are Proof Exam (PE) and Proof Evaluation Exam (PEE). While no significant differences were observed among freshmen with respect to department in PE scores; senior students' mean scores differ significantly both in PE and PEE. It has been observed that freshmen students mostly rely on inductive reasoning when they attempt to prove given mathematical statements. Even though seniors are mostly aware of the necessity of generalizing their results and attempting to use procedures involving deductive reasoning, they still have difficulties in constructing and evaluating proofs. Implications for teaching are discussed.

Keywords: Mathematical proof, proof evaluation, proof construction, proof methods, inductive and deductive reasoning

Introduction

Over the past several decades, research on mathematical proof has gained increasing attention in the area of mathematics education. Studies about proof focus on reading, understanding and validating proofs (e.g. Inglis and Alcock, 2012; Ko and Knuth, 2013; Selden and Selden, 2003; Weber, 2010), students' and educators' perceptions and attitudes towards proof (e.g. Almedia, 2000, 2003; Baştürk, 2010; Köğçe and Yıldız, 2011) and how they construct proofs (e.g. Harel and Sowder, 1998; Weber, 2005). There is also research emphasizing the role of proof in the classroom and how it should be taught (e.g. Hanna, 2000; Martinez et al., 2011).

Such emphasis is put on proof because it is an essential part of mathematics and therefore of academic mathematicians' daily practices. It is also important for mathematics educators because proof involves reasoning, conviction and communication and helps meaningful learning. Proofs can be used to show students that understanding and performing mathematics means more than just learning to execute certain procedures. According to Weber (2005), proving is "a complex mathematical activity with logical, conceptual, social and problem-solving dimensions." While this complex nature of proof makes it a valuable tool in learning mathematics, difficulties arise in classroom applications because, as many studies suggest (e.g. Healy and Hoyles, 2000; Hoyles and Küchemann, 2002; Knuth 2002; Miyazaki 2000; Morris 2000, 2002; Stylianides and Stylianides, 2009; Weber, 2001, 2010), students across all grades have a poor understanding of proof and have difficulties in constructing their own proofs.

While proof is obviously an important part of mathematics, some believe that its key role in the classroom is the promotion of mathematical understanding; that proof should be viewed primarily as explanatory tool and proofs that best help to explain should be valued most (Hanna 2000). Turkish mathematics curriculum (grades 5-8 and grades 9-12) also highlights active, meaningful learning

through processes related to proof. The students are expected to engage in mathematics actively, learn how to solve problems, share, explain and justify their solutions and ideas and find relations within mathematics as well as between mathematics and other subjects (MEB, 2011, 2013). In the mathematics program for grades 5-8 (MEB, 2013), it is stated that in order to improve students' reasoning skills, some of the indicators that should be considered include justifying the truth and validity of inferences, making reasonable generalizations and inferences (which are concerned with validation and construction of proofs respectively) and explaining and using mathematical patterns and relationships while analyzing a mathematical situation. In addition, among general aims of the mathematics program for grades 9-12 (MEB, 2011) it is stated that students should be able to make inductive and deductive inferences and choose appropriate methods during proving process, use mathematical terminology and language appropriately to explain and share their mathematical ideas.

Undoubtedly, mathematics educators have crucial roles in realization of these aims. Considering the importance of proof in mathematics education, emphasized both by current research and mathematics curriculum, and the difficulties that students face regarding proofs, this study aims to investigate mathematics and teaching mathematics majors' proof construction and evaluation practices. In addition, by involving freshmen students who are at the very beginning of their programs, the researchers aimed to shed light on the proof experiences students bring from high school.

Target population for the study consists of students from Mathematics Education (Primary and Secondary) and Mathematics Departments in a state supported university located in a metropolitan area. These students are freshmen and senior prospective mathematics teachers and mathematicians; which makes them important figures that will shape primary school, secondary school and university students' mathematical conceptualizations in the future. Therefore, this study is an important step for understanding and comparing prospective mathematicians' and prospective mathematics teachers' conceptions of proof at the time of starting the undergraduate program as high school graduates and finishing it. Clarification of these participants' conceptions will have instructional implications for teaching mathematics programs. In addition, mathematics teachers and instructors of freshmen mathematics courses will see some of the tendencies of proof patterns seen in high school graduates. With all these considered, the aim of this study is to;

- investigate whether there are significant differences in students' proof construction practices with respect to grade (freshmen and senior) and department (Mathematics, Secondary Education Teaching Mathematics and Primary Education Teaching Mathematics),
- investigate whether there are significant differences in senior students' proof evaluation practices with respect to department (Mathematics, Secondary Education Teaching Mathematics and Primary Education Teaching Mathematics),
- examine students' proof practices, when they are asked to prove mathematical statements,
- examine seniors' proof evaluation practices, when they are asked to evaluate freshmen students' mathematical arguments.

Method

Review of Some Basic Concepts. Before moving on to the details about collection and analysis of data, a quick reminder of the basic related terms (with definitions this paper assumes) are given below.

In the broadest sense, proof can be thought as establishing the truth of a certain claim. To prove a claim, mathematicians use definitions, already established truths and a series of logical rules to reach a conclusion. This process is called making an *inference*. In *deductive reasoning*, inference process leads from general to particular and premises provide necessary evidence for the truth of the conclusions. In *inductive reasoning*, inference process leads from particular to general, and premises provide probable, but not necessary evidence for conclusions (Morris, 2007; Overton, 1990).

An *argument*, in logic, is defined as "a sequence of sentences or propositions of which one (conclusion) is said to follow from others (premises), and the premises are said to provide evidence for the truth of the conclusion" (Overton, 1990). A deductive argument is *valid* when it is impossible to have true

premises and a false conclusion in the argument. Inductive arguments cannot be assigned as valid or invalid because the premises are only probable evidence for the conclusion. It must be pointed out here that validity refers to the process; and does not guarantee that the initial assumptions (premises) are true. If an argument is valid, it just simply means that if the premises are true then the conclusion has to be true as well.

Mathematicians use both inductive and deductive reasoning in their mathematical practices. Inductive processes usually occur prior to asserting a claim. Mathematicians consider many cases and examine patterns and relationships to come up with a generalization. Once that generalization is reached, its truth should be established using deductive processes.

One can use different methods to prove a mathematical statement. These can be categorized as *direct* and *indirect* methods. Direct methods assume the truth of the premises and set out to reach the conclusion (for example, *proof by cases* is a direct method) while indirect methods include assuming the conclusion is false and reaching the result that the premises has to be false as well (*proof by contrapositive*) or assuming the statement is false and reach a contradiction (*proof by contradiction*). To prove that a mathematical statement (where it is claimed that statement is true for all cases) is false, one can use a *counter-example* to show the statement is false for at least one case. Although *mathematical induction* method, where the truth of a claim is established in certain steps is sometimes considered separately, it can also be counted as a direct method.

With these in mind, for the context of this paper, constructing a proof means building a mathematical argument in order to establish the truth of a certain claim and evaluation of a proof refers to the process of determining whether a mathematical argument (or a justification of a mathematical claim) can be accepted as a valid proof.

Participants. There are two groups of participants; freshmen students who, in the time of data collection, had just graduated from various high schools and senior students who were about to graduate from the university to become mathematicians and mathematics teachers. Participation of freshmen students produced two types of information: High school graduates' competencies of related to proof practices were observed and development of these competencies throughout university was revealed via cross-sectional comparison of their competencies with seniors'.

Participants are from the following departmental programs: Department of Primary Education/Program of Teaching Mathematics (PRED), Department of Secondary School Science and Mathematics Education/Program of Teaching Mathematics (SCED) and Department of Mathematics (MATH). Primary Education graduates teach years 5 to 8 and Secondary Education graduates teaches years 9 to 12 (high school). For the characteristics of the sample see Table 1.

Table 1. Characteristics of the participants

Department	Freshmen		Seniors		Total
	Male	Female	Male	Female	
MATH	17	22	10	13	62
PRED	8	23	15	15	61
SCED	5	18	17	12	52
Total	30	63	42	40	175

At the university from which the participants are selected, all students from related programs in Faculty of Education enroll to mathematics courses given by the Mathematics Department, generally together with Mathematics students. Hence, their content knowledge is formed by the courses that they take from Mathematics Department. However, prospective secondary school teachers take more mathematics courses than prospective primary education students.

Instruments. Proof Exam (PE) was designed to collect information about participants' proof construction practices, types proof techniques they use, and how effectively they can use it. The students were asked to prove (or disprove) the given mathematical statements. Data were collected in paper-pencil format.

During the development process of PE; instruments used in the studies that investigate participants' mathematical reasoning and proof techniques were examined (Almeida, 2000, 2003; Miyazaki, 2000; Morris, 2002; Özer and Arıkan, 2002; Selden and Selden, 2003; Stylianides and Al-Murani, 2010; Stylianides et al., 2004, 2007; Recio and Godino, 2001; 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) were considered. 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. The content covered by the items is included in the high school curriculum (MEB, 2011) such as divisibility and properties of natural numbers and integers. All items can be proved in several ways using alternative proof methods. In addition, items of different levels of difficulty have been selected for the instrument in order to ensure a more accurate idea about participants' reasoning skills. To ensure content validity, opinions and suggestions of experts regarding the suitability of items for the target population (level of students) and aim of the study were taken into consideration. These experts were; a high school teacher, two instructors from Mathematics Department and an instructor from Teaching Mathematics Program.

In order to develop the rubric, in addition to one of the researchers, two other experts (mathematics graduates/teaching assistants) coded the data from PE (they were given a randomly selected sample consisting about 1/3 of student responses). Results from these three coders were organized and final categorization for the rubric was established. For each item, scores between 0 and 3 were given according to the following criteria:

- Incoherent response, no basis for a valid proof construction, no attempt at generalization: 0 points
- Attempt at generalization; complete use of known formulas and information without any justification; correct idea with insufficient explanation; presenting a valid general argument that does not prove the given statement: 1 point
- Presenting a valid general argument but missing steps, needs more clarification or some justification; some use of mathematical language and symbols: 2 points
- Presenting a valid general argument with sufficient explanation and clarity; good use of mathematical language and symbols: 3 points

Second instrument, Proof Evaluation Exam (PEE), was developed in order to collect data about senior students' proof evaluation practices. PEE includes all mathematical statements of Proof Exam (PE) which need to be proved (or disproved). For each statement, alternative arguments were given as proofs. These arguments were chosen from freshmen responses to PE. 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) because student generated arguments are more authentic and they better represent the type of arguments the participants will have to make sense of as mathematics teachers/instructors in the future. For each alternative proof attempt 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".

In order to form the rubric for PEE, the instrument was initially administered to three experts who were working as academic staff 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 a senior academician from Mathematics Department.

Results

Data were collected from the freshmen students in the first week of fall semester, so that freshmen's responses would solely reflect their high school knowledge and experiences. Instruments were administered to all freshmen students during the first lecture session of an introductory mathematics course to which all three program students were registered. In the same year, PE and PEE were administered to senior students of the selected programs.

Parametric and nonparametric statistical techniques were used in order to investigate whether there are statistically significant differences in a) students' proof construction practices with respect to grade (freshmen and senior) and department, b) senior students' proof evaluation practices with respect to department.

In addition, responses to PE were examined to find out a) how students constructed proof by categorizing the methods they used and b) whether they were successfully using those methods.

PEE responses were also examined to see how seniors decide whether a given argument can be accepted as a proof.

Proof Exam. Means for the total PE scores with respect to grades and departments are given below in Table 2. Maximum score for PE is 12, while minimum score is 0.

Table 2. Means and standard deviations of PE total score

Department	Freshmen			Seniors		
	M	SD	N	M	SD	N
MATH	1.79	2.09	39	9.40	1.77	15
PRED	2.19	2.32	31	4.07	2.75	28
SCED	1.74	1.84	23	6.66	2.88	29

Shapiro - Wilk test was conducted to check normality and yielded significant results for PE scores of freshmen mathematics students ($W = 0.82, p = 0.00 < 0.05$), prospective primary school ($W = 0.84, p = 0.00 < 0.05$) and secondary school teachers ($W = 0.85, p = 0.00 < 0.05$). Significant results were also observed for senior prospective mathematicians ($W = 0.88, p = 0.048 < 0.05$) and freshmen students as a whole ($W = 0.86, p = 0.00 < 0.05$). Therefore, since normal distribution could not be assumed for most subgroups, nonparametric tests were carried out to see whether there are significant differences between total PE scores with respect to grade and department.

Kruskal-Wallis test conducted on PE scores, revealed no significant results among freshmen prospective mathematicians, primary and secondary school teachers.

However, significant results were observed among seniors: $\chi^2(2, N = 72) = 27.42, p = 0.00 < 0.05$. Mann-Whitney tests were conducted to make pair wise comparisons among seniors' departments. Tests yielded significant results between prospective primary and secondary school teachers ($U = 210, p = 0.02 < 0.05, r = 0.42$), in favor of prospective secondary school teachers; between prospective mathematicians and secondary school teachers ($U = 92.50, p = 0.02 < 0.05, r = 0.47$), and prospective

mathematicians and primary school teachers ($U = 23$, $p = 0.00 < 0.05$, $r = 0.73$), both in favor of prospective mathematicians.

PE scores also differ significantly among freshmen and seniors ($U = 964.50$, $p = 0.00 < 0.05$, $r = 0.62$) and they are in favor of seniors.

In summary, examining the scores of PE given in Table 2, it is observed that freshmen have an average score of 1.92, where maximum possible score is 12. No significant differences are observed between departments among freshmen. Seniors' average scores are 4.07, 6.66, and 9.40 for prospective primary and secondary school teachers, and mathematicians respectively. These results are significantly higher than freshmen, as can be expected, but when it is considered that the items in the instrument consist of high school level problems, one can expect them to be even higher. In case of seniors, as stated above, mean differences between all departments are significant.

Proof Evaluation Exam. Table 3 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 3. 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 ($p = 0.13 < 0.05$).

Looking at Table 3, it is seen that senior prospective primary school teachers have a mean score of 23.73 in total out of 54; which is, as mentioned above, significantly lower than the other two departments. Means for prospective secondary school teachers (35.56 out of 54) and prospective mathematicians (40.64 out of 54) do not differ significantly.

Proof Types and Proof Evaluation. In order to see types of proof participants use and how they evaluate proofs, their responses to PE and PEE were analyzed item by item. Due to space restrictions, detailed analyses of only two items are presented. Since the mathematical statements in both instruments are the same, results obtained from both instruments are given together for each statement. It must be kept in mind that PE was conducted to freshmen and senior students, while PEE was conducted to seniors only.

First item was as follows: "Prove that the statement is true: If the square of a natural number is even, then that number must be even".

When participants' scores for the first item of PE are examined in depth, it is seen that 42% of freshmen students did not receive any points and only 7.5% were given maximum points. Amount of

seniors who received minimum and maximum points are 13.9% and 34.3% respectively. Figures 1 and 2 show the distribution of scores for freshmen and senior students by departments.

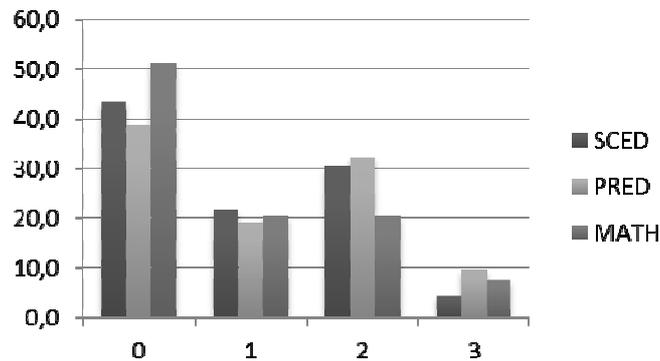


Figure 1. Percentage frequencies of freshmen scores for PE, item 1

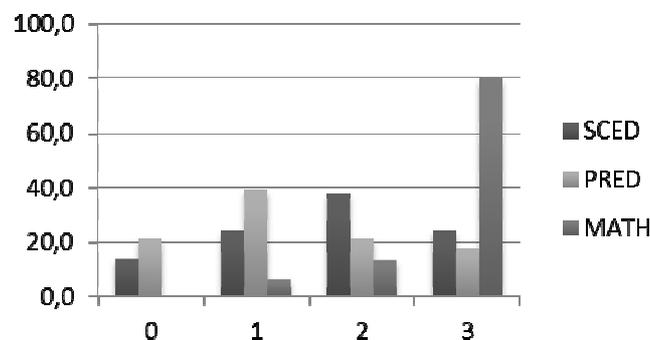


Figure 2. Percentage frequencies of senior scores for PE, item 1

Participants' responses to the first item of PE were also categorized, with respect to the proof methods they attempted to use, as follows:

- Proof 1A: If n was odd, then its square would be odd (proof by contrapositive).
- Proof 1B: If n is even then its square is even (this proves the converse of the given statement; not equivalent to the original statement).
- Proof 1C: Assume that n is odd but n^2 is even. If n is odd then n^2 will be odd (proof by contradiction).
- Proof 1D: Assume n^2 is even ...then n must be even (direct proof).
- Proof 1E: The square of an even number is even, the square of an odd number is odd. Hence, if the square of a number is even, then that number should be even (proof by cases).

Most attempted proof types by freshmen students for this item are direct proof (20.4%) and proof by cases (22.6%). 21.5% of freshmen students attempted to prove the converse of this statement: "if n is even then its square must be even". Even though it is a true proposition, it does not prove the given statement. More interestingly, senior prospective secondary school teachers (20.7%) and prospective primary school teachers (46.4%) also made the same mistake. No senior prospective mathematicians provided this type of response. Most attempted proof types for seniors with respect to departments are as follows: mathematicians used proof by contradiction (66.7%), primary school teachers attempted to prove the converse of the statement (46.4%), and secondary school teachers used proof by cases (31.0%).

For the same item in PEE, senior participants were asked to evaluate five alternative proof attempts which freshmen students provided. 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: $k = 2m$, $n^2 = 4m$, $\sqrt{(n^2)} = \sqrt{(4m)}$, $n = \sqrt{(2m)}$, 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 $q \rightarrow p$ instead of $p \rightarrow q$. 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 just a few 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 3.

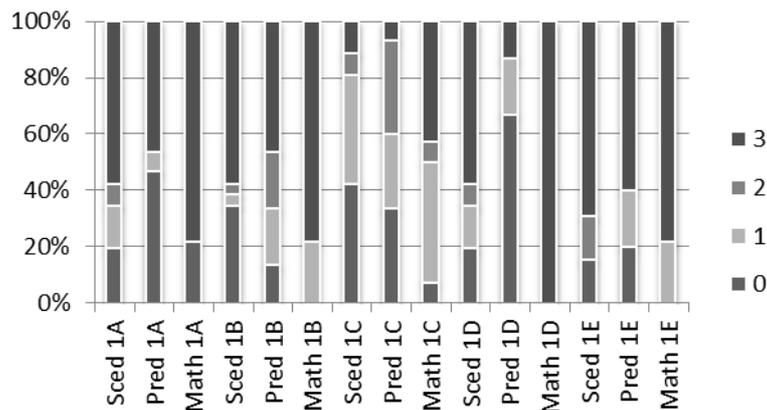


Figure 3. Percentage frequencies of PEE scores for item 1

It can be seen from Figure 3 that 1C was the hardest argument to evaluate for all groups, while 1E was the easiest. In all cases, more senior mathematics students received full points than the students from the other departments.

Second item was as follows: “Prove or disprove: The equality $1 + 3 + 5 + \dots + 2n-1 = n^2$ is true for all integers $n \geq 1$ ”.

The scores for the second item PE indicate that 78.5% of freshmen and 13.9% of seniors received minimum score. Maximum score was received by 5.4% of freshmen and 54.2% of seniors. Distributions of scores for freshmen and seniors are given in Figure 4 and Figure 5 respectively.

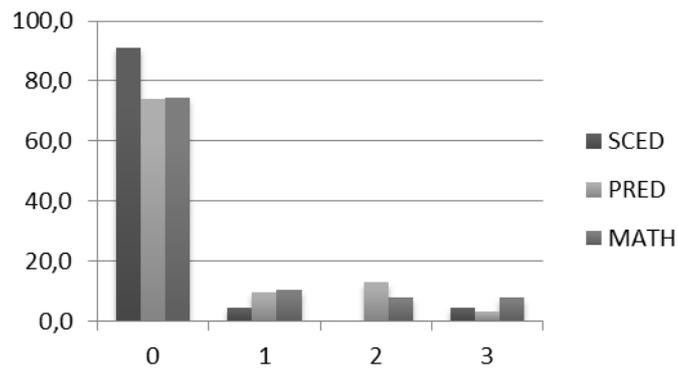


Figure 4. Percentage frequencies of freshmen scores for PE, item 2

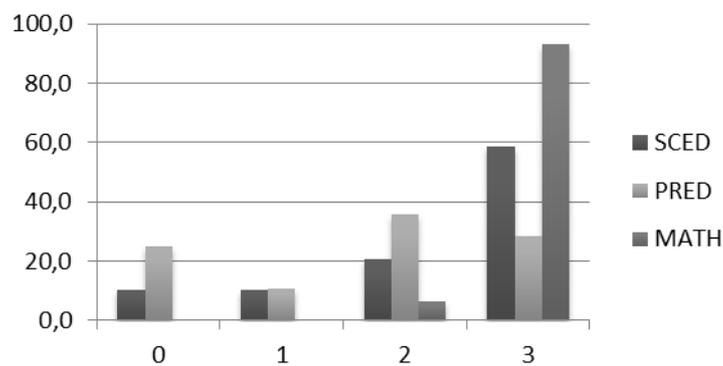


Figure 5. Percentage frequencies of senior scores for PE, item 2

Participants' responses to item 2 were categorized with respect to proof types as follows:

Proof 2_A: By formula: $\text{Sum} = \text{number of terms} \times (\text{last term} + \text{first term}) / 2$.

Proof 2_B: Using Gauss' method (writing the same sum in reverse and adding up the terms).

Proof 2_C: By induction.

Proof 2_D: Using the equality $1+2+3+\dots+n = n(n+1)/2$.

Proof 2_E: By giving numerical examples.

Majority (67.7%) of freshmen either did not attempt this item or failed to provide a coherent response. Among the rest, most commonly observed (9.7%) response was to use a known general formula (without justification) which verifies that the statement is true.

This statement is one of the common examples used explaining proof by mathematical induction. While 66.7% of seniors used mathematical induction (MATH 80%, SCED 65.5% and PRED 60.7%), only 4.3% of freshmen attempted to prove the statement with this method.

In PEE, five alternative proof attempts were given for this statement. 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% 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% of the participants pointed out the missing step but failed to give the correct choice. Only 12.7% 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 of these sums is equal to $2n$, and if we add them all up we get $2n^2$, which is twice the sum we are looking for. Figure 6 shows percentage frequency distributions of the scores for proofs 2A through 2E.

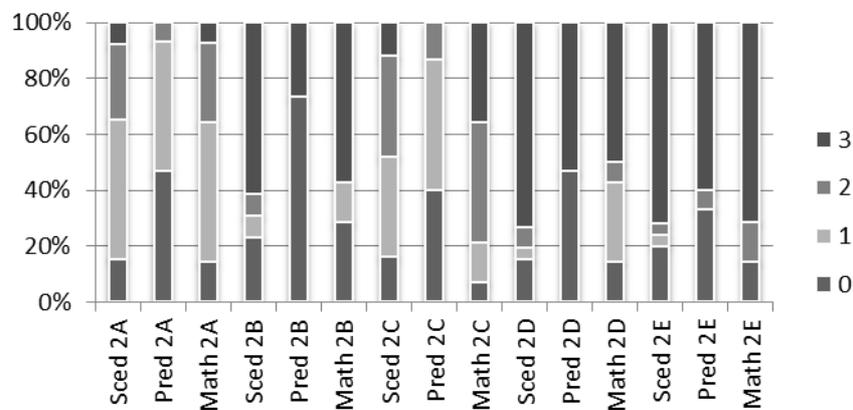


Figure 6. Percentage frequencies of PEE scores of item 2

Looking at Figure 6, one can see that more than 60% of students from all departments received full points evaluating argument 2E, while less than 10% received full points from evaluating option 2A.

Discussion

Results of the study indicate that there are no significant differences between departments among freshmen regarding their proof construction abilities. Considering that the instruments were conducted to freshmen at the very beginning of their first semester in the university, which means their responses reflect their high school knowledge and experiences, it can be assumed that students have more or less the same exposure in high school regarding proof. Findings also indicate that significant differences were observed in seniors' proof construction and proof evaluation practices. This suggests the differences occur as a result of their university education.

It is seen from the findings related to senior students that prospective mathematicians have the highest scores in Proof Exam and Proof Evaluation Exam 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. As mentioned before, prospective teachers' content knowledge is formed by the courses that they take from Mathematics Department. While it can be argued how much high level content knowledge is required for prospective teachers (especially at the primary school level), more exposure to university level mathematics may increase students' ability to construct and evaluate proofs.

Type of proof methods freshmen and seniors attempted to use were categorized using data collected by the Proof Exam. The first item in the Proof Exam, “if n^2 is even then n is even”, can be proven by using various methods, and the responses of both freshmen and seniors reflect that. This item is one of the common examples of proof by contradiction or contra-positive, and these indirect proof methods were mostly used by mathematics seniors and some of the seniors from Teaching Mathematics Programs. With the exception of mathematics seniors, most participants preferred direct approaches such as direct proof and proof by cases. According to Antonini and Mariotti (2008), studies regarding indirect proof report that “students’ difficulties with indirect proof seem to greater than those related with direct proof”, and assuming that what needs to be proved is false may be mentally demanding and false hypotheses and contradictions make it harder to follow the deductive steps of the proof.

Another important observation regarding this item is that 21.5% of freshmen, 20.7% of senior prospective secondary school mathematics teachers and 46.7% of senior prospective primary school mathematics teachers proved the converse of this statement: “if n is even then its square must be even”. While this statement is also true, it is not logically equivalent to the original statement. No senior mathematics student provided this type of response. Inability to distinguish between a statement and its converse indicates a poor understanding of logical implication. Such difficulties were also reported in the study of Hoyles and Küchemann (2002). They report that most of the participants in their study stated that a conditional statement and its converse were equivalent and did not check whether this claim was true using the truth values of the statements. Since in the current study both the statement and its converse is true, making a distinction between the statement and its converse may be even more difficult.

Even though freshmen’s proof scores are low, they produced a more variety of proof approaches (successfully or not) than seniors, which was especially apparent in their responses to the item “prove or disprove: the equality $1 + 2 + \dots + 2n + 1 = n^2$ is true for all integers $n \geq 1$ ”. This equality is one of the classic examples of proof by mathematical induction, and expectedly, majority of seniors attempted to use this method. Freshmen however, attempted other methods which could be considered as more creative. Mingus and Grassl (1999) found a similar result, while they were examining middle and high school students’ responses to the item “show that there are just as many even numbers as there are odd numbers”. Authors report it was the middle school students who constructed the most creative arguments, not the high school students who were more familiar with formal proof. Even though freshmen were seemingly not too familiar with the proof method most suitable for this item, they were able to produce various arguments probably because central exams include type of questions that require computing similar sums. In addition, among the students who attempted mathematical induction, a common mistake observed was to omit the basis step of induction, which was also observed by Stylianides, et al. (2007). While mathematical induction was the most commonly attempted proof method by seniors from all departments, only 4.3% of freshmen attempted to prove the statement with this method, even though high school mathematics program includes proof by mathematical induction (mathematical induction method is a deductive process and should not be confused with inductive reasoning).

To summarize, participants’ responses to Proof Exam reveal that inductive methods are usually preferred by freshmen students. Seniors attempt to generalize their arguments, but they (mostly prospective primary and secondary school teachers) have difficulty in distinguishing the difference between a proving statement and its converse, prefer direct proof approaches even though indirect approaches could have been conveniently applied, and omit the basis step in the mathematical induction method.

In addition to freshmen and senior students’ proof construction practices, proof evaluation practices of senior students were also examined in this study. For the item “if n^2 is even then n is even”, seniors were asked to evaluate five alternative attempts. Proof 1A was an example of proof by cases, with minor calculation mistake, which did not affect the generality of the result. Proof 1B was constructed using the contradiction method, even though it could have been expressed better. An attempt at direct proof was given in proof 1C, with a missing justification. As the percentages given in results section indicate, this argument was harder to evaluate because no obvious mistakes stood out.

The proof of converse statement “if n is even then n^2 is even” was given in proof 1D. As mentioned before, this is a valid argument but does not prove the given statement. Another interpretation emerged is that the proof is incomplete, and the case where n is odd should also have been examined, then it would be proof by cases. Both interpretations were given full points. Percentages of students who received maximum points are 100%, 57.69% and 13.34% for mathematics, secondary and primary school teaching mathematics programs respectively. It is also worth noting here that 64.29% of prospective primary school teachers thought the argument proves the given statement for all cases. As mentioned above, majority of prospective primary school teachers provided this type of proof for the corresponding item in the Proof Exam. Finally in proof 1E, truth of the statement is verified for only a couple of values of n . Since there is no generalization, this cannot be accepted as a valid proof. 18.3% of seniors thought the argument showed the statement is true for all cases. While majority of students correctly detected that this cannot be accepted as a proof, one would expect that the percentages of correct responses would have been higher, since this argument is the most apparent example in the Proof Evaluation Exam where the statement is not proven for all cases.

When mean scores for each argument related to this item are examined, it can be said that prospective mathematicians were best at correctly distinguishing between a statement and its converse, while prospective primary school teachers had the most difficulty with it. Prospective secondary and primary teachers were best in recognizing that giving a finite number of numerical examples cannot be accepted as a valid proof (where the domain of discourse is infinite). Prospective secondary school teachers and mathematicians had the most difficulty with proof 1C, where there was a crucial step needed to be justified.

There were also five arguments in the Proof Evaluation Exam for the item “prove or disprove: the equality $1 + 2 + \dots + 2n + 1 = n^2$ is true for all integers $n \geq 1$ ”. 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. 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. Similar situation is observed in the argument presented in proof 2D, which 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 in order to find the sum of even numbers, the formula $n(n+1)/2$ is used again without proof.

Students are apparently familiar with certain formulas that are used to calculate various sums. In the case of 2A, a formula is directly used to find the required sum where in cases 2B and 2D, some reasoning is observed but still a formula to find another sum is used. Students might have been convinced by these arguments because formulas are true and used appropriately. Evaluation of this kind of proofs can be tricky when there is uncertainty to what extent previously known facts can be used without proof.

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 . 30.1% of the participants pointed out the missing step but failed to give the correct choice. Only 12.7% of the participants concluded that the missing step would make the proof invalid.

When this result is compared with the corresponding responses of the same item in the Proof Exam, it is seen that in PE majority of senior students preferred induction and much higher percentage of them received maximum points. This result indicates that while most students did not make this mistake in their own proofs, they do not consider omitting the basis step of induction as a major mistake. 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.

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. Most students in all departments did not have difficulty in evaluating this argument, probably because it is easy to follow, convincing and uses no assumptions.

When mean scores for item 2 of Proof Evaluation Exam (which were given in Table 3) are examined, it is seen that students from all departments had difficulties evaluating the argument where the result is obtained by using a formula, which should not have been used here without proof. Prospective primary school teachers and prospective mathematicians were most successful in evaluating the proof where the result is obtained simply by adding the terms of the required sum in reverse order, and prospective secondary school teachers were most successful in evaluating the argument where the terms of the sum is obtained by subtracting one from each even number from 2 to $2n$.

To sum up, results of data analysis of Proof Evaluation Exam reveal that most seniors 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, 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.

Conclusion

The role and importance of proof in mathematics education has been discussed in recent studies (e.g. Almeida, 2003; Mariotti, 2006; Brown et al., 2008; Hanna and Barbeau, 2008; Martinez, et al., 2011). Jahnke (2007) points out that findings of studies suggest “many school and university students and even teachers of mathematics have only superficial ideas on the nature of proof”. To investigate the situation in a particular setting, this study was conducted with the aim of examining freshmen and senior students’ proof construction and proof evaluation practices. The sample consisted of students coming from Mathematics and Primary and Secondary Education Teaching Mathematics Programs. Instruments developed for the study were administered to freshmen students at the very beginning of their programs; so their responses reflect their high school knowledge and experiences.

According to the results of this study, freshmen students in all three departments have difficulties in constructing proofs and their scores do not significantly differ with respect to department. One possibility for high school graduates’ poor performance may be that even though curriculum includes proof, it is probably not emphasized much in class because of the pressure of succeeding in test based central exams or because teachers themselves may have difficulties about proof, which may cause them to avoid the subject.

Most of freshmen and some senior students use inductive methods while proving. Inductive reasoning is similar to everyday reasoning where people make predictions and decisions based on observational evidence (their experiences). As stated in Weber’s (2010) study, students in middle school and high school, university students and even some mathematics teachers are convinced by such arguments. Martin and Harel (1989) reported that inductive and deductive proof schemes exist simultaneously in the student. Healy and Hoyles’ (2000) study revealed that while students chose deductive arguments as the ones their teacher would give the best mark, they chose inductive ones as the arguments which they would adopt as their own approach. This suggests even when the students know the proofs need a formal deductive approach; it does not come naturally to them.

Mathematicians use inductive form of thinking to discover the relationships and patterns within mathematical objects. After that they pose conjectures and try to prove or disprove these conjectures using deductive methods. Both approaches can be considered valuable from an educational perspective, but emphasizing the distinction between inductive and deductive reasoning would help students use them consciously so that their mathematical thinking skills can be improved. Current high school mathematics curriculum (MEB 2011) points out that, in the traditional teacher-centered approach, mathematics lessons usually follow the order Definition, Theorem, Proof, Application and Evaluation; which does not offer opportunities for students to use high level mathematical skills. Promoting a student centered approach; the program suggests that the order should be as follows: Problem, Discovery, Hypothesizing, Justification, Generalization, Formation of relationships, Inference. Further research can be carried out to investigate the situation high schools and primary schools to see how this recent change of approach is reflected in classroom settings and in which ways it affects practices related to proof. Whether the teachers are equipped and ready to apply the strategies suggested by the curriculum in mathematics classes remains to be seen. Attitudes and practices of mathematics teachers, textbooks and teaching materials can be examined in that regard.

In order for these changes to be effective, teacher education programs in Turkey should be revised, too. As the results of this study show, even though significant improvement is observed between freshmen and seniors as a result of university education, senior teacher candidates still have difficulties related to proof. If we consider that the items of the instruments used in this study are based on high school curriculum (proof methods and related content are present in the program), one would expect both freshmen and seniors to be more successful. As Ko and Knuth (2013) suggests, examining teacher candidates' strategies for validating proofs provides insight into their mathematical thinking and reasoning. If teacher candidates are expected to follow a student centered approach and achieve meaningful learning in mathematics classes in future, they should learn the nature of mathematics at university in a similar environment so that they can understand firsthand what student centered approach and active involvement in mathematical process means. Specifically, bridge courses can be added in mathematics education programs that focus on incorporating methods used in theoretical mathematics courses to middle and high school mathematics. Learning environments involving proof activities can be created in these classes, where students are given the opportunity to form their own conjectures and the truth of these conjectures are discussed and determined by the class. This way, deeper understanding of the subject can be achieved and risk of students seeing proof as a topic to be learned instead of as a process that is in the essence of mathematics can be decreased.

Note

Results related to analysis of data collected by one of the instruments used in this study, Proof Evaluation Exam, were partially presented at the 12. International Congress on Mathematical Education (ICME 12), Seoul, KOREA under the title of "An Investigation of Senior Mathematics and Teaching Mathematics Students' Proof Evaluation Practices".

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.
- Antonini, S. and Mariotti, M. A., (2008). Indirect proof: What is specific to this way of proving? *ZDM-The International Journal of Mathematics Education*, **40**(3), 401-412.
- Baştürk, S., (2010). First-year secondary school mathematics students' conceptions mathematical proofs and proving. *Educational Studies*, **36**(3), 283-298.
- Brown, J., Stillman, G., Schwarz, B. and Kaiser, G., (2008). The case of mathematical proof in lower secondary school: Knowledge and competencies of pre-service teachers. In M. Goos, R. Brown and K. Makar (Eds.), *Navigating currents and charting directions*, Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, Brisbane (Vol. 1, pp. 85-91) Adelaide: MERGA.
- Cupillari, A., (2001). *The nuts and bolts of proofs (Second ed.)*. San Diego, CA: Academic Press.
- D'Angelo, J. P. and West, D. B., (2000). *Mathematical thinking: Problem solving and proofs (Second ed.)*. Upper Saddle River, NJ: Prentice Hall.

- Hanna, G., (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, **44**(1-2), 5-23.
- Hanna, G. and Barbeau, E., (2008). Proofs as bearers of mathematical knowledge. *ZDM-The International Journal of Mathematics Education*, **40**(3), 345-353.
- Harel, G. and Sowder, L., (1998). Students' proof schemes: Results from exploratory studies. *CBMS Issues in Mathematics Education*, **7**, 234-283.
- Healy, L. and Hoyles, C., (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, **31**(4), 396-428.
- Hoyles, C. and Kuchemann, D., (2002). Students' understandings of logical implication. *Educational Studies in Mathematics*, **51**(3), 193-223.
- Inglis, M. and Alcock, L., (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, **43**(4), 358-390.
- Jahnke, H. N., (2007). Proofs and hypotheses. *ZDM-The International Journal of Mathematics Education*, **39**(1-2), 79-86.
- Knuth, E. J., (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, **5**(1), 61-88.
- Ko, Y. and Knuth, E.J., (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *Journal of Mathematical Behavior*, **32**(1), 20-35.
- Köğçe, D. and Yıldız, C., (2011). A comparison of freshman and senior mathematics student teachers' views of proof concept. *Procedia Social and Behavioral Sciences*, **15**, 1266-1270.
- Mariotti, A. (2006). Proof and proving in mathematics education. In A. Gutiérrez and P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 173-204). Rotterdam: Sense Publishers.
- Martin, W. G. and Harel, G., (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education*, **20**(1), 41-51.
- Martinez, M. V., Brizuela, B. M. and Superfine, A. C., (2011). Integrating algebra and proof in high school mathematics: An exploratory study. *The Journal of Mathematical Behavior*, **30**(1), 30-47.
- M.E.B., (2011) "Ortaöğretim matematik (9, 10, 11 ve 12. sınıflar) dersi öğretim programı (secondary school mathematics program-grades 9-12)" <http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72>, (accessed June 2014)
- M.E.B., (2013) "Ortaokul matematik dersi (5, 6, 7 ve 8. sınıflar) öğretim programı (primary school mathematics program-grades 5-8)" <http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72>, (accessed June 2014)
- Mingus, T. Y. and Grassl, R. M., (1999). Preservice teacher beliefs about proofs. *School Science and Mathematics*, **99**(8), 438-444.
- Miyazaki, M., (2000). Levels of proof in lower secondary school mathematics. *Educational Studies in Mathematics*, **41**(1), 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.
- Overton, W. F. (Ed.), (1990). *Reasoning, necessity, and logic: Developmental perspectives*. Hillsdale, N.J.: L. Erlbaum Associates.
- Özer, Ö. and Arıkan, A., (2002). Lise matematik derslerinde öğrencilerin ispat yapabilme düzeyleri. Paper presented at the Congress of V. National Science and Mathematics Education 16-18 September, Ankara.
- Recio, A.M. and Godino, J. D., (2001). Institutional and personal meanings of proof. *Educational Studies in Mathematics*, **48**, 83-99.
- Selden, A. and Selden, J., (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.
- Stylianides, A., Stylianides, G. and Philippou, G. N., (2004). Undergraduate students' understanding of the contraposition equivalence rule in symbolic and verbal contexts. *Educational Studies in Mathematics*, **55**(1-3), 133-162.
- Stylianides, A., Stylianides, G. and Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics and Teacher Education*, **10**(3), 145-166.
- Stylianides, A. and Stylianides, G., (2009). Proof constructions and evaluations. *Educational Studies in Mathematics*, **72**(2), 237-253.
- Stylianides, A. and Al-Murani, T., (2010). Can a proof and counter-example coexist? Students' conceptions about the relationship between proof and refutations. *Research in Mathematics Education*, **12**(1), 21-36.
- Solow, D., (2005). *How to read and do proofs (fourth ed.)*. Hoboken, NJ: John Wiley and Sons, Inc.
- Weber, K., (2001). Student difficulty in constructing proofs. The need for strategic knowledge. *Educational Studies in Mathematics*, **48**(1), 101-119.
- Weber, K., (2005). Problem solving, proving and learning: The relationship between problem solving processes and learning opportunities in the activity of proof construction. *Journal of Mathematical Behavior*, **24**(3-4), 351-360.
- Weber, K., (2010). Mathematics majors' perception of conviction, validity and proof. *Mathematical Thinking and Learning*, **12**(4), 306-336.