Engaging prospective teachers in peer assessment as both assessors and asessees: The case of geometrical proofs

Ilana Lavy
The max Stren Yezreel Valley College
Atara Shriki
Oranim- Academic College of Education

Abstract

One aspect of professional development of mathematics teachers relates to the development of assessment skills. The aim of this study is to examine the effects of engaging prospective mathematics teachers in peer assessment, both as assessors and asessees, on the development of their assessment skills in general and assessment of geometrical proofs in particular. The research was conducted within a Method course in which peer assessment activities were employed. Sixteen prospective mathematics teachers participated in the research and had to act both as assessors and asessees. Analysis of the research data reveals that during the various phases of the study the prospective teachers developed skills concerning the selection of categories and weights for the assessment of their peers' work. In the criteria set they selected for the peer assessment, they referred to meanings and roles of mathematical proof. In their reflections, the prospective mathematics teachers also referred to the effects of the peer assessment on their mathematical knowledge, asserting that by being exposed to different solution strategies and new problems they were able to widen their mathematical knowledge.

Keywords: peer assessment, professional development, prospective teachers, geometrical proof, "what-if-not?" strategy

1. Introduction

Many issues relating to various aspects of mathematical proofs have been investigated during the last two decades. Among these issues are the role of proof in mathematics classrooms (Hersh, 1993), students' difficulties in providing proofs (Moore, 1994), students' perceptions of proof (Harel & Sowder, 1998) and the construction of proofs (Weber, 2001; Weber & Alcock, 2004). However, the assessment of mathematical proofs has received little attention, especially from the students’ perspectives (Alcock & Weber, 2005; Mamona-Downs &
Both students and teachers have difficulties in evaluating proofs for correctness (Selden & Selden, 2003; Alcock & Weber, 2005).

There is no doubt that among the important skills that have to be developed during the learning of teaching practice are assessment skills since evaluation of students' work is one of the teacher’s responsibilities. Engagement in assessment activities in general and assessment of mathematical proofs in particular of prospective mathematics teachers (PMTs) may help them develop the required skills. The importance of teaching PMTs to properly evaluate students' work stems from the fact that the way teachers evaluate their students can either reinforce or undermine the learning process. Instructors are therefore obliged to consider the implemented evaluation approaches carefully and adjust them to their teaching and learning goals. Various assessment methods are customary in the field of education. One of the assessment methods is peer assessment, which is an arrangement for learners to consider and specify the level, value or quality of a product or performance of other equal-status learners (Topping, 2003). Peer assessment provides feedback that can reduce errors and have positive effects on learning when it is received thoughtfully and positively. The active engagement of learners in the assessment process is regarded as important, and peer assessment has been found to be an effective tool in self-improvement (Boud, 2000). Engagement in peer assessment may also raise the students' motivation for learning (Hanrahan & Isaacs, 2001).

Being aware of the importance of developing the PMTs' assessment skills and of the benefits of the peer assessment approach, we believe that engaging PMTs in peer assessment of geometrical proofs can support their professional development. Moreover, engagement in peer assessment, both as assessor and assessee, may expose students to both sides of the coin and hence may result in a deeper comprehension of the importance of both the mathematical constructs (geometrical proofs) and the process of their assessment.

In this paper we describe a study designed to examine the effects of experiencing peer assessment of mathematical proofs, both as assessors and asessees, on the development of assessment skills of PMTs and their mathematical knowledge. In implementing the peer assessment approach on PMTs we addressed the following questions:

(i) How do students choose, assign weights and justify categories for the assessment process?
(ii) How do they react to feedback received from their classmates about their own work?
(iii) How do they refer to the roles and meanings of proof in their assessment process?
(iv) In what ways does students’ exposure to their classmates’ work affect their own mathematical work?
2. Theoretical Background

Assessment methods can be utilized in a summative or a formative way. When an assessment method is used in a summative way, the learning outcomes and products are being assessed at the end of the learning process. Formative assessment, on the other hand, is intended to help students plan their own learning, identify their own strengths and weaknesses, target areas for remedial action and develop metacognitive and other personal and professional transferable skills (Boud, 2000). Assessment which specifically indicates strengths and weaknesses and provides frequent constructive individualized feedback leads to significant learning gains, as compared to traditional summative assessment (Black & Wiliam, 1998).

In what follows, we present a brief literature survey regarding roles and meanings of mathematical proofs, peer assessment, peer assessment in higher education, and assessment of mathematical proofs.

2.1 Roles and meanings of mathematical proof

Researchers have broadly discussed the meaning and roles of proof in the mathematics practice (Polya, 1954; Bell, 1976; de Villiers, 1990, 1991; Hanna, 2000). First of all the role of a proof is to verify/confirm a given statement. The proof also provides insight into why the statement is true (explanation). Proof may also provide the organization of various results into a deductive system of axioms, concepts and theorems (systematization). During the process of finding a proof, new outcomes may emerge. The proof also has an important role in communication of the mathematical knowledge by the transmission of it, and the process of looking for a proof is often an intellectual challenge.

According to the NCTM (2000) standards, reasoning is a practice of mind and students need opportunities to engage with reasoning and proof in many contexts, including courses in which technology plays a significant role (Thompson, 2009). Harel and Sowder (1998) defined a student’s proof scheme to be the processes student uses to become certain of the truth of a mathematical statement, and to convince others of this certain truth. Many educators and researchers have investigated students’ learning and abilities with proof concepts when using a dynamic geometry drawing tool (e.g., Chazan, 1993; Hadas, Hershkowitz & Schwarz, 2000; Laborde, 2000). Students who are accustomed to technology in all aspects of their lives may consider proofs done via technology as providing evidence to both ascertain, and persuade others of, the truth of a claim.

We believe that in order to be able to properly assess a mathematical proof, PMTs should be able to construct a suitable criteria set. Suitable criteria set is one that takes into account the roles and meanings of mathematical proof. Exposing PMTs to an assessment of mathematical
proof, in which they act both as assessors and assessees, may help them develop assessment skills needed for proper assessment of mathematical proof.

2.2 Peer Assessment

In educational context, assessment is defined as the process in which students' products, which are results of their knowledge and skills, are measured (Topping, 2003).

Peer assessment, is a process whereby students assess assignments of their classmates based on a given criteria list (McDowell & Mowl, 1996). Peer assessment activities can vary in a number of ways, operating in different curriculum areas or subjects. The product or output to be assessed can vary. It can be writing, portfolios, oral presentations, test performance, or other skilled behaviors. The peer assessment can be summative or formative. The assessors and the assessed may be individuals or pairs or groups. Feedback is essential to the development and execution of self-regulatory skills (Bangert-Drowns et al., 1991; Paris & Newman, 1990; Paris & Paris, 2001). Students of all ages react differently to feedback from peers and from adults (Cole, 1991).

As to PMTs, providing effective feedback from peer assessment is a cognitively complex task requiring understanding of the goals of the assessed task and the criteria for success, and the ability to make judgments about the relationship of the product or performance to these. Formative peer assessment is likely to involve intelligent questioning, coupled with large self-disclosure and thereby assessment of understanding. Peer assessment could enable the identification and analysis of error and misconception. This could lead to the identification of knowledge gaps, and/or engineering their closure through explaining, simplification, clarification, summarizing and cognitive restructuring (Topping, 1998). Feedback which is corrective and/or confirmatory and/or suggestive could be more immediate, timely, and individualized. This might increase reflection and generalization to new situations, promoting self-assessment and greater meta-cognitive self-awareness. Cognitive and meta-cognitive benefits might accrue before, during or after the peer assessment. Peer assessment demands social and communication skills, negotiation and diplomacy (Riley, 1995), and can develop teamwork skills. Learning how to give and accept criticism, justify one's own position and reject suggestions are all useful transferable social and assertion skills (Marcoulides & Simkin, 1991).

Peer assessment involves students directly in learning, and might promote a sense of ownership, personal responsibility and motivation (Hanrahan & Isaacs, 2001). Giving positive feedback first might reduce anxiety among the assessees and improve acceptance of negative
feedback. Peer assessment might also increase variety and interest, self-confidence, and empathy with others for both assessors and assessees.

In the case of PMTs, peer assessment is the evaluation of PMTs by other PMTs from the same group. The goal of the peer assessment processes is to verify whether the work satisfies the accepted standards, as well as to provide constructive feedback which includes suggestions for improvements (Herndon, 2006). The use of peer assessment is based on the assumption that peers can recognize each other's errors quickly and easily, and that a larger and more diverse group of people might find more weaknesses and errors in a work. In their literature review, Falchikov and Goldfinch (2000) pointed to the increase in students' involvement in assessment across the spectrum of discipline areas. This approach, however, is rarely implemented in higher education (Zevenbergen, 2001).

Researchers also reported on disadvantages during the implementation of peer assessment (Falchikov & Goldfinch, 2000; McDowell, 1995; Topping, 1998; Zevenbergen, 2001). They reported that sometimes poor performers had difficulty in accepting peer feedback as accurate. Falchikov, (1995) found that especially in small socially cohesive group, students were not willing to take part in a process in which they had to assess their peers. In addition, in cases there are no clear criteria for the assessment process, problems that relate to validity and reliability can be raised, and accuracy and valuabily of the feedback is questionable. Thus, although peer assessment might yield added value, it should be used carefully.

2.3 Peer assessment in higher education

In England, Gatfield (1999) examined undergraduate students' satisfaction with peer assessment, and found that there was a high level of satisfaction with the process. Students tended to perceive peer assessment as an appropriate form of assessment; they regarded it as a fair method and believed that students should assess their peers. In higher education, student peer assessment both formative and summative, has many benefits in terms of empowering the learners: the process supports the development of autonomy and higher-order thinking skills; improves the quality of learning; develops the ability to evaluate and justify (Topping, 1996; 1998); provides opportunities for students to self-monitor, rehearse, practice and receive feedback (Falchikov & Goldfinch, 2000); increases the amounts of feedback that students receive regarding their work; and encourages students to take responsibility for their own learning (Race, 1998). By reviewing the work of peers, students in higher education gain insights into their own performance, as well as developing their ability to make judgments, which is an essential skill for professional life. At the same time, they learn how to refer to mistakes as an opportunity for learning rather than as a failure. Moreover, engagement in peer
assessment promotes the exchange of ideas and serves as a basis for guiding academic discourse (Berkencotter, 1995) and the interchange of ideas and methodologies resulting in a more refined product (Reese-Durham, 2005).

Peer assessment might expose problems that relate to validity and reliability, and there is no guarantee that the feedback provided will be accurate and valuable. A similar problem concerns peer grading (Falchikov & Goldfinch, 2000). It appears that the subjectivity associated with scoring students' work is influenced by students' different standards of what a numerical score should consist of (Conway, Kember, Sivan & Wu, 1993). Even when students are provided with clear criteria for evaluating and marking, there can still be a problem of subjectivity. For example, Zevenbergen (2001) examined peer assessment by mathematics prospective teachers of posters created by their peers and identified an observable trend among high and low achievers and the scores they gave to their peers' work: low achievers assigned higher than average marks and often provided very generic comments, whereas high achievers tended to assign lower than average marks and provided more insightful and critical comments.

2.4 Assessment of mathematical proofs

Among the various skills mathematics teachers have to develop during their teaching training are assessment skills in general and assessment of mathematical proofs in particular. Researchers found that teachers and students encountered difficulties in processes of proofs' evaluation for correctness (Alcock & Weber, 2005; Selden and Selden, 2003). More specifically, Knuth (2002) found that in-service high school teachers accepted geometrical proofs according to their format regardless their content. Meaning that, two column format proofs of claims and justifications were accepted as correct regardless their content.

Although assessment skills are fundamental to effective teaching, teachers rarely have opportunities to engage in assessment design (Webb, 2009). According to Aksu (2008), teachers have negative attitudes towards the idea of using alternative assessment techniques such as formative, peer or self-assessment because of their traditional thoughts and fixations. Aksu (ibid.) argues that one of the reasons for the above phenomenon originates in insufficient exposure of the teachers during their training period to these techniques. Many researchers support the idea of incorporating engagement in alternative assessment techniques, such as formative, peer and self-assessment activities, in professional
development programs of teachers (Lovemore & David, 2006; Sluijsmans & Prins, 2006; Aksu, 2008; Sato et al., 2008; McGatha et al., 2009).

In order to examine whether or not mathematics teachers are able to distinguish between proofs and empirical arguments, Stylianides & Stylianides (2009) engaged a group of teachers in a process termed ‘construction-evaluation’. In this process these teachers had to construct a mathematical proof to a given claim and then to evaluate their own construction. They found that only three to ten percent of the prospective teachers considered empirical arguments as a proper mathematical statement. These researchers attributed their findings to the nature of the tasks given to the study participants. Their results contradict results received in previous research (Goetting, 2005; Martin & Harel, 1989) in which more than 50% of the prospective teachers considered empirical arguments as a proper mathematical proof.

Most of the research literature concerning peer assessment of mathematical products refers to the assessor’s or the assessee’s point of view separately or to self-assessment (e.g. Stylianides & Stylianides, 2009). However, in this study we examine the effects of engagement in peer assessment as assessor and assessee at the same time on the PMTs’ ability to develop assessment skills in the context of geometrical proof. We conjecture that the PMTs’ engagement in peer assessment of geometrical proofs as assessors and receiving feedback on their own proofs provided by classmates as assessees, can help them both develop assessment skills and widen their mathematical knowledge.

3. The Study

In this section we present information about the study participants, a description of the Method course in which the study took place, and an outline of the various phases of the study. In addition, we present the methodology used for the data collection and analysis.

3.1 The study participants

Sixteen PMTs from an academic college of education participated in the study. The PMTs group was heterogeneous: nine were regular PMTs, studying in their third year (out of four) toward a B.A. degree in mathematics education and computer sciences or physics for middle and high school. Four PMTs were in-service mathematics teachers who chose to complete a teaching certificate in mathematics. Of these four, two had 2 years of teaching experience and the other two had over 10 years of teaching experience. The remaining three PMTs were adults who were considering a second career as mathematics teachers.
3.2 The Method course
The method course in which the PMTs experienced peer assessment was a two-semester course and is the second mathematics method course the PMTs were required to take. In the first course the PMTs studied and discussed teaching methods and approaches for the practicing of various high school mathematics topics. In addition, they became acquainted with basic principles concerning the assessment process of an exam, such as how to set proper categories for an exam assessment, how to build an indicator for the evaluation of the exam's answers, and how to evaluate tasks given to school students. It is worth noting that this process was performed on a real exam of a student however, the student did not receive these assessments. In the second course, in which the research took place, the PMTs were mainly engaged in mathematical explorations through the process of problem posing and peer assessment activities. Since the process of problem posing yielded familiar and unfamiliar problems, they had to assess both solution of familiar (a problem they had already solved) and unfamiliar problems (problems posed by their classmates that were not familiar to them).

3.3 A ten-phase process
The study participants were engaged in a ten-phase process which lasted one semester. The phases were as follows:

1. The PMTs were asked to solve an identical given problem in which they had to construct geometrical proofs (Figure 1) and to submit their solutions by e-mail.
2. The PMTs' solutions were scanned and sent through the course forum mail for evaluation. Each PMT received two anonymous solutions for evaluation and scoring. The PMTs were asked to formulate a list of evaluation criteria, to assign each criterion a numerical weight, and to provide justification for each criterion and weight. In order to receive a unified form, the students had to complete a form that was prepared by us (Appendix 1). In addition, the PMTs had to specifically score the work according to each criterion and explain the underlying reasons for each score. These evaluations were e-mailed to us.
3. Each PMT was e-mailed the two anonymous evaluations and scoring of his or her work and was asked to reflect on them. In order to receive comprehensive and effective feedback, the students were asked to address eight questions (Appendix 2) concerning various aspects of the evaluation received from their peers and its contribution to their future work.
4. Using the "What If Not?" (WIN) strategy (Brown and Walter, 1993) and based on the initial problem, the PMTs had to start an inquiry process. The WIN strategy is based on the idea that modifying the attributes of a given problem can yield new and intriguing problems, which eventually may result in some interesting investigations. In this phase, the PMTs were
asked to produce a list of the initial problem's attributes, suggest alternatives for each of them, and choose one alternative or a combination thereof for further mathematical investigation. Again, the PMTs e-mailed us their tasks.

5-6. Phases 2 and 3 were repeated, based on the output of work produced in Phase 4.

7. The chosen new problem was solved, including a full written description of the conjectures, solutions, indecisions, and so on.

8-9. In order to enable the PMTs to experience a peer assessment process for unfamiliar problems, Phases 2 and 3 were repeated, based on the output of work produced in Phase 7.

10. Each PMT wrote a summative evaluation and final reflection on the entire process.

### 3.4 The initial problem

Midpoints of successive sides of the parallelogram ABCD form the polygon EFGH.

- a. What is the obtained figure of EFGH?
- b. What is the ratio between the area of EFGH and the area of ABCD?
- c. What is the ratio between the perimeter of EFGH and the perimeter of ABCD?

![Figure 1: The initial problem](image)

We chose the above problem for the following reasons: (a) solution of the problem necessitates proving rather than calculations; (b) the problem can be solved in various ways; (c) it serves as a rich source for problem posing activities; (d) it is a common problem taken from a school textbook.

### 3.5 Data collection and analysis methods

During the course, the PMTs were engaged in three peer assessment tasks (phases 2-3, 5-6 and 8-9). The first and the third tasks were similar in nature. In these tasks each student had to construct a criteria set, set weight to each criterion and assess two geometrical proofs done by classmates according to each criterion and explain the underlying reasons for each score. The second task was different - in this task the PMTs were asked to produce a list of the attributes
of the initial problem, suggest alternatives for each of them, and choose one or a combination of a few alternatives for further mathematical investigation.

Following analytic induction (Taylor & Bogdan, 1998) and content analysis (Neuendorf, 2002), the data collected from the first and the third tasks were analyzed according to the following focal points:

(i) Evaluation criteria: the nature of the criteria and its relation to the roles and meaning of geometrical proof.
(ii) Numerical scores and feedback: an examination of the scores assigned to each criterion by two different classmate assessors and the provided feedback.
(iii) Reflections on the whole process from both perspectives: assessors and assesses.
(iv) Effects of the peer assessment process on the PMTs’ mathematical knowledge.

4. Results and Discussion
As previously mentioned, during the whole process, the PMTs were engaged in three peer assessment tasks. Since the second peer assessment (phases 5-6) is different in nature from the others, the data analysis focuses only on the first and the third peer assessments tasks. In what follows, we present results and discussion of the data received from phases 2-3 (first task) and 8-9 (third task).

4.1 The peer assessment process
The process of the first and the third peer assessments comprised the following: (a) constructing evaluation criteria for the peer assessment of the geometrical proofs and setting a numerical weighting for each criterion; (b) determining the proper numerical scores according to chosen criteria; (c) inclusion of justifications for the scoring process; and (d) reflecting on the assessment process both as assessors and assesses.

Figure 2 illustrates assessment relations within a certain subgroup of PMTs (a certain PMT is marked by “Sk”). The notation A_{i,j} indicates that S_i assesses the work of S_j. It should be mentioned that the situation described in Figure 2 is only illustrative and can be different. Namely, S_1 can be assessed by S_3 and S_2.

In order to examine the development of assessment skills, we looked at the gap between the two assessments given by two different PMTs to the same task. A gap between assessments refers to: (i) the difference between the constructed criteria for the assessment of the same task (for example, the criteria included in A_{1,3} and A_{2,3}); (ii) the gap between the relative weights assigned to two identical criteria; (iii) the gap between the scores assigned to two
identical criteria; and (iv) the differences between the provided feedback. Moreover, the feedback were compared in order to check whether allegations raised by the assessed PMT about checking his work, were similar or not to the complaints of the PMTs that were assessed by him. We chose to use these indicators since reduction of these of gaps might imply on the development of PMTs’ assessment capabilities (Lavy & Yadin, 2010).

We also compared between the assessment done by each PMT (for example, A_{1,3}), and the assessments s/he received (in this case, A_{4,1} and A_{5,1}), as will be elaborated in section 4.2.

Moreover, in order to examine the appropriateness of the constructed criteria, we compared the PMTs’ criteria set with the one constructed by us. Each one of us constructed her own criteria set to each of the tasks and after discussing the differences between our criteria lists, we arrived at a criteria set that was acceptable by the two of us. Finally, the scores given by each PMT were examined in reference to their academic achievements in previous mathematics courses.

### 4.1.1 Selecting evaluation criteria

A total of 30 assessments (two by each of 14 PMTs and 1 by each of the other two PMTs) were obtained in the first and the third assessment tasks. Analysis of these assessments revealed that the PMTs used 12 different evaluation criteria in the first task and 8 criteria in the third one (Table 1).

Table 1 summarizes the list of the evaluation criteria provided by the PMTs in the first and the third task organized according to various aspects of proof (listed in column 1). The number of PMTs that selected the criterion (frequency) and the average weight (AW)
assigned to each criterion, and the standard deviations (SD) of the assigned weights are also shown.

Table 1: Evaluation criteria constructed by the PMTs in the first and the third tasks

<table>
<thead>
<tr>
<th>Aspects of proof</th>
<th>Evaluation criterion</th>
<th>First task</th>
<th>Third task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Frequency</td>
<td>Average weight (AW) and standard deviation (SD)</td>
</tr>
<tr>
<td>Structure</td>
<td>1. Customary proof structure</td>
<td>11 (68.75%)</td>
<td>AW=30 SD=10.44</td>
</tr>
<tr>
<td>Verification</td>
<td>2. Correct solution</td>
<td>14 (87.5%)</td>
<td>AW=36.32 SD=17.41</td>
</tr>
<tr>
<td></td>
<td>3. Calculations errors</td>
<td>1 (6.25%)</td>
<td>AW=10</td>
</tr>
<tr>
<td>Communication</td>
<td>4. Correct quotation of geometrical theorems</td>
<td>5 (31.25%)</td>
<td>AW=29 SD=11.93</td>
</tr>
<tr>
<td></td>
<td>5. Easy and understandable solution</td>
<td>5 (31.25%)</td>
<td>AW=12.2 SD=10.2</td>
</tr>
<tr>
<td></td>
<td>6. Adding a clear sketch of the problem</td>
<td>5 (31.25%)</td>
<td>AW=13 SD=6.7</td>
</tr>
<tr>
<td></td>
<td>7. Clear use of mathematical notations</td>
<td>2 (12.5%)</td>
<td>AW=10</td>
</tr>
<tr>
<td></td>
<td>8. Additional explanations to claims</td>
<td>2 (12.5%)</td>
<td>AW=27.5</td>
</tr>
<tr>
<td>Systematization</td>
<td>9. Reasonable organization and clarity</td>
<td>4 (25%)</td>
<td>AW=15 SD=7.07</td>
</tr>
<tr>
<td></td>
<td>10. Formulation of correct conclusions</td>
<td>1 (6.25%)</td>
<td>AW=25</td>
</tr>
<tr>
<td></td>
<td>11. Demonstration of connectedness knowledge between mathematical areas</td>
<td>2 (12.5%)</td>
<td>AW=27.5</td>
</tr>
<tr>
<td>Intellectual challenge</td>
<td>12. Providing the shortest possible proof</td>
<td>1 (6.25%)</td>
<td>AW=10</td>
</tr>
<tr>
<td>Systematization</td>
<td>13. Well-defined problem</td>
<td>9 (56.25%)</td>
<td>AW=17.7 SD=5.65,</td>
</tr>
<tr>
<td></td>
<td>14. Suggestions for elaboration/generalization</td>
<td>10 (62.5%)</td>
<td>Average=20 SD=10.27</td>
</tr>
<tr>
<td>Intellectual challenge</td>
<td>15. Interesting problem</td>
<td>1 (25%)</td>
<td>AW=20</td>
</tr>
</tbody>
</table>

Interpretation of Table 1 reveals that although in the first task the PMTs constructed 12 criteria referring to various aspects of proof, in the third task they constructed only five criteria for this matter. However, in the third task they constructed three additional criteria that refer to the nature of the examined problem (criteria 13-15 in Table 1).

Before referring to each criterion we can generally say that although the number of constructed criteria decreased from the first to the third task, an increase of the number of

1In the first task each PMT constructed 3-5 criteria and in the third task each PMT constructed 3-6 criteria
PMTs is observed in the third task with comparison to the first one concerning the construction of the following criteria (shaded rows in Table 1): 'easy and understandable solution' (criterion 5) and 'formulation of correct conclusions'(criterion 10). We attribute these differences to the different nature of the tasks. While in the first one the PMTs had to assess a proof of a problem familiar to them, in the third task, they had to assess a proof to a problem that was not familiar to them, as was reflected in similar words by five of the PMTs: 'while reading the proof to the problem suggested by my classmate I had difficulties in following his line of thought, hence, I included this criterion [5] in my criteria list'. As regards to criterion 10 one PMT wrote: 'the problem I had to assess was quite surprising and it was important to me to see whether my classmate was able to draw the problem's conclusions in a clear and correct manner'.

As evident from Table 1, 'correct solution' (criterion 2) is the most common criterion in both tasks. This might imply on the PMTs’ awareness of the significant role of verification in the context of mathematical proof.

The first criterion refers to the structure of proof. The relatively high number of PMTs who constructed this criterion, both in the first and the third tasks is in line with Martin and Harel (1989) and Knuth (2002) who found that many pre-service teachers accept the validity of a geometry proof mainly according to its structure. Namely, they would judge it valid if the proof was in the standard two-column format and invalid if it was in paragraph form, regardless of its mathematical content.

'Correct quotation of geometrical theorems' (criterion 4), 'Easy and understandable solution' (criterion 5), 'Adding a clear sketch of the problem' (criterion 6), 'Clear use of mathematical notations' (criterion 7) and 'Additional explanations to claims' (criterion 8)) refer to the communication role of proof. While criteria 4, 5 and 7 are phrased in a general manner, criteria 6 and 8 are more specific. Criterion 5 does not specify ways the PMT has to use in order to make the proof more understandable, however criteria 6 and 8 refer to specific attributes that might raise the understandability of the provided proof.

'Reasonable organization and clarity' (criterion 9), 'Formulation of correct conclusions' (criterion 10) and 'Demonstration of connectedness knowledge between mathematical areas' (criterion 11) refer to systematization. Seven PMTs recognized the importance of including criteria which examine the systematic structure of a mathematical proof in the first task while eleven of them did in the third task. This increase may point to the PMTs recognition of the significance of the systematization role of proof during the peer assessment process. This increase can be also attributed to the difference between the first and the third task. They
found it easier to follow and assess a proof, to a problem which is not familiar to them, that was written in a systematic manner than otherwise.

One PMT provided 'Providing the shortest possible proof'(criterion 12) which refers to intellectual challenge. It should be noted that the PMTs had trouble generating criteria that would reflect the consideration of non-measurable aspects, such as creativity and originality, and the clear majority of PMTs did not suggest such criteria at all.

The decrease in the number of the constructed criteria from the first to the third task might be explained by the different nature of the tasks, previously described.

The criteria set for the third task can be divided into two subgroups: the first refers to characteristics of the chosen problem and the second to the characteristics of the provided proof.

As to the first subgroup: 'Well-defined problem'; 'Interesting problem' and 'suggestions for elaboration/generalization' (criteria 13-15) refer to the problem chosen by the PMTs. To more than half of the PMTs it was important that the chosen problem to be well defined since through this criterion the PMTs checked the mathematical validity of the problem. It was also important to them that the provided problem will include 'suggestions for elaboration/generalization'. However to only one PMT was it important whether the problem is interesting.

As to the second subgroup, most of the PMTs chose 'Correct solution' (criterion 2) which refers to the verification role of proof. Repetition of this criterion in both tasks 1 and 3, points on its perceived high importance. Concerning the communication role of proof in the third task the PMTs constructed 'Adding a clear sketch of the problem' (criterion 6) and 'Clear use of mathematical notations' (criterion 7). To the systematization role of proof, many PMTs provided 'Formulation of correct conclusions' (criterion 10). As stated by de Villers (1990) and Hanna (2000), communication of the mathematical knowledge is among the main roles of mathematical proof. To facilitate the communication of mathematical arguments in general and geometrical proofs in particular, it is important to add sketches and use correctly mathematical notations.

4.1.2 Determining the proper scores according to the chosen criteria

The peer scoring of the first task is presented in Figure 3. Each PMT received two scores from two different classmates. It can be clearly seen that in most cases there is a considerable gap between the scores each PMT received (average gap: 19.5 per cent, SD: 14.61 per cent). The high average gap might point to the fact that the PMTs had no previous experience in peer assessment and found it difficult to provide scores according to a set of evaluation
criteria they had to construct. The high variance (SD: 14.61 per cent) points also to significant differences between the average gap and the actual gaps between the scores some PMTs received from their classmates. These large gaps might also result from the significant differences of numerical weights the PMTs assigned to similar evaluation criteria. It should be noted that since there were two PMTs whose task, for technical reasons, was scored by only one PMT, they were not included in the data presented in Figure 3.

![Figure 3: Peer scoring of the first assessment task (phases 2-3)](image)

The following are some reflections regarding the scoring process according to the selected criteria and numerical weight:

Ruth: *It was a difficult decision – should I give points for the correct portions or deduct points from the total for the mistakes?*

Josef: *I could see from the work that the solver knew the solution. The final answer was correct. The method, however, was unclear. So I wondered whether I should give points for the solution or deduct points for the method...It was difficult to give a low score, because I felt that s/he knew...*

Both Ruth and Josef (pseudonyms are used throughout the paper) referred to the issue of the scoring technique. When a certain numerical weight is assigned to a criterion, usually it stands for the fulfilling of this criterion in the evaluated solution. But when the criterion is not completely fulfilled in the solution, the question of setting the proportional part of the numerical weight to this criterion is raised. Do we have to work upwards (give points to the
correct parts) or in reverse (reduce points from the total numerical weight assigned to this criterion)? According to Josef, the scoring process becomes even more difficult when the assessor ‘feels’ that the assessee has a certain understanding but fails to convey his ideas. Figure 4 presents the peer scoring of the third task.

![Peer scoring of the third task](image)

**Figure 4**: Peer scoring of the third assessment task (phases 8-9)

For the third task, the average gap between scores was 11 per cent and the standard deviation (SD) was 9.35 per cent. Comparing these values with those received in the first task we can observe a convergence tendency. The gap between the scores given by two PMTs to the same classmate was significantly reduced. This result may imply that the PMTs enhanced their assessment skills.

Examination of the scores given both in the first and the third task and the academic achievements in mathematical courses of the assessors shows that low achievers tended to give higher scores than high achievers. This finding is in line with Zevenbergen (2001), who found a similar phenomenon among mathematics PMTs who were engaged in peer assessment.

**4.2 The PMTs' reflections on the feedback**

Feedback is among the most critical influence on student learning (Hattie & Timperley, 2011). Hence, as part of the assessment process, the PMTs were asked to reflect on the feedback they provided to the scoring as assessors and on the feedback they received on their
own work as assesses. Content analysis (Neuendorf, 2002) of the PMTs’ reflections revealed that this feedback can be divided according to their nature and their mathematical content. This analysis revealed different types of feedback:
1. Feedback comprising of the following symbols: 'V' for a proper proof; '∀' for a partially correct proof, and 'X' for a faulty proof.
2. Feedback comprising of verbal explanations which can be characterized by their nature and content.

4.2.1 Feedback comprising of symbols
In the first task two PMTs provided feedback merely by using symbols. For example:

Yael: One of the works I had to assess was very good and I just marked the 'v' sign at the bottom of it.

When teachers assess their students’ works and the works are correct, many of them are accustomed to react by adding the 'v' symbol. Hence, since students are used to receive such comments they react similarly. Lortie (1975) termed this phenomenon as 'apprenticeship of observation' which describe a situation whereby pre-service arrive for their training courses having spent many hours as schoolchildren observing and evaluating their teachers in action. According to Lortie (ibid) this apprenticeship is mainly responsible for behaviors described above and many of the preconceptions that pre-service teachers hold about teaching. It should be mentioned that when the PMTs encountered a partially correct or a faulty proof, all of them added a verbal explanation to the assessment they provided.

4.2.2 Feedback comprising of verbal explanation
In this section we elaborate on feedback comprising of verbal explanations. This kind of feedback can be characterized by its nature and content.

4.2.2.1 Nature of feedback
Nature of these feedback refers to issues such as whether the explanation (i) was phrased in a positive/negative manner; (ii) was short and/or general or detailed and/or specific; (iii) included reinforcing comments in cases where the proof is correct; (iv) include alternative suggestions in cases where the provided solution was not correct or included guidelines that might help the assessee improve her/his solution. In what follows we present and discuss representative quotations from the PMTs' reflections according to the above categories from both perspectives: as assessors and as assesseses. In developing feedback typology, Tunstall & Gipps (1996) identified different types of feedback which part of them are similar to the ones
identified by us. The feedback types that are similar to the ones we identified are: verbal and non-verbal; distinctly positive or negative; feedback which was based on the use of explicit criteria.

**Phrased in a positive/negative manner**

From an assessee's perspective:

Dov: One of the assessments I received was quite harsh. I do not argue with the content of it but I think that if it was phrased in a positive rather than a negative manner – I could have accepted it more.

However from an assessor's perspective:

Dov: …Then it occurred to me that the explanations I provided in the peer assessments were not better. I now realize the importance of it.

From an assessee point of view, Dov refers to the articulation of the explanation (positive/negative) and shares his difficulties in accepting this kind of justification. He did not question the content of the justification but stressed that writing in such a harsh way does not help the assessee to accept and internalize the criticism he receives on his work. Then he recalls he acted similarly as assessor and becomes aware of the importance of the way one should justify his scoring.

**Be short and/or general or detailed and/or specific**

From an assessee's perspective:

Ella: ...Although the score was high, I was not satisfied, since the explanations were general and shallow and I had the impression that my work was not treated seriously…

From an assessor's perspective:

Ella: In the assessments I gave I did my best to justify each score carefully.

As an assessee, Ella refers to the generalized nature of the explanation which left her with the impression that the assessors did not check her work seriously, claiming that as an assessor she provided justifications to each part of the assessed solution.

Roy: The score I received in one of the assessment forms was not high but since it was logically justified down the smallest details, I felt I deserved the score I received.

Roy, on the other hand, reports on an experience in which a detailed explanation enabled him to accept the relatively low score he received. Such a detailed explanation may help the PMTs to learn how to refer to mistakes as an opportunity for learning rather than as a failure (Race, 1998).

Include alternative suggestions in cases where the provided solution was not correct or include guidelines that might help the assessee improve her/his solution

From an assessee's perspective:
Dina: In one of the assessments I received was written: ‘The solution is well organized; however, there are missing justifications to the first and the forth claims.’ The specific comment helped me to correct my proof.

From an assessor's perspective:

Dina: During the evaluation process I found several weaknesses: there were unjustified claims; there were jumps of logic in the proof which made the tracing of the proof's course very difficult. Giving scores without mentioning the underlying reasons does not help the student understand his faults.

In her reflection as assessee, Dina refers to the case in which the assessor comments do not specify guidelines that might help the assessee to correct his work, claiming she acted differently as an assessor. A certain frustration was expressed in cases where the PMTs felt they acted differently from their assessors.

**Include reinforcing/harsh comments**

From an assesse's perspective:

Yael: I was very proud to receive the comment: 'Excellent work, I enjoyed reading your work'.

Ruth: I felt humiliated when I received the following: 'Rehearse the relevant subject matter and complete your solution, elaborate on the various solution stages and add justifications'. It could be phrased more gently.

From an assessor's perspective:

Yael: One of the works I had to assess was very good and I just marked the 'v' sign at the bottom of it. After I received my work I realized that I should have acted differently.

Ruth: I read again and again the justifications I gave to avoid a situation in which the assessee could be offended.

In her reflection, both as assesse and assessor, Yael refers to another aspect of the nature of the justification which refers to the question whether or not to include reinforcing/harsh comments in the justification saying that she was proud to receive praise on her excellent solution and recalling she did not act similarly as an assessor. From the other side of the spectrum, Ruth complained about the harsh nature of the justification she received on her work since these comments conveyed an explicit criticism concerning her mastery of the subject matter.

All the above reflections refer to various aspects concerning the nature of the provided justifications. While some of the PMTs reported acting differently from the way they were treated as assesseees, others realize they acted similarly as assessors and became aware of their faults. As assesseees, when referring to the nature of the received justifications, the PMTs referred mainly to affective issues. For example, they raised aspects such as general and short justification that can leave the reader with the feeling that his work was not treated seriously. The lack of reinforcing comments or the absence of alternative suggestions to reach the
correct answer left them with feelings of frustration. Acting both as assessors and assessees confronted the PMTs with their behavior as assessors and with their expectations from their classmates while acting as assessors.

4.2.2.2 Content of feedback

Content of the feedback refer to aspects of roles and meanings of geometrical proof as follows: (i) correct solution; (ii) understandable solution; (iii) correct quotation of geometrical theorems; (iv) adding a clear sketch of the problem; (v) clear use of mathematical notations; (vi) additional explanations to claims; (vii) formulation of correct conclusions; (viii) demonstration of connectedness between mathematical areas. Review of the feedback as regards to their content revealed they were especially from the assessor perspective.

Correct solution

Ella: First of all I have to verify that the proof I have to assess is mathematically correct and clear.

Ella, as many of the other PMTs in both tasks, is aware to the fact that above all the work she has to assess must be correct. It is also in line with the research literature asserting that first of all the role of a proof is to verify/confirm a given statement (Polya, 1954; Bell, 1976; de Villiers, 1990, 1991; Hanna, 2000). This criterion is associated with the category of verification.

Understandable solution

Dina: I could not understand the course line of one of the proofs I had to assess, no matter how many times I read it. Eventually I turned to the class teacher to help me understand the proof.

In the first task, five PMTs included this criterion in their criteria set while none did in the third task. The possible explanation to this phenomenon is that since in the first task all the PMTs solved the same problem, when they had to assess a solution which was different from the one they did, they encounter difficulties in understanding it. In the third task each student focused in a different problem, so they tried to understand the provided solution without having in mind the prior solution for it. To facilitate the understanding of one’s provided proof, s/he has to write it in an understandable way. Hence this criterion can be classified under the communication category.

Correct quotation of geometrical theorems

Ruth: In case the quotation of geometrical theorem was incorrect, I directed the assessee to the proper place in the subject matter.
Five students in the first task included this criterion in their criteria set while none of them in the third task. This can be explained by the fact that the PMTs admitted they chose the criteria set after they read the classmates’ solutions and the absence of this criterion in the proofs of the third task suggests they found it unnecessary. This criterion can be attributed to the category of communication since in order to convince others that the provided proof is correct, one should establish his claims by adding correct quotation of the relevant geometrical theorems.

**Adding a clear sketch of the problem**

Dan: When a sketch was missing in one of the proofs, I wrote to the assessees that a clear sketch could have helped me to better understand her/his proof.

Approximately one-third of the PMTs both in the first and the third task included this criterion in their criteria set. This finding is surprising since during the PMTs' studies of geometry instructors emphasize the importance of adding a clear sketch to the proof, so we expected that more PMTs will include such a criterion in their criteria set. A possible explanation of this phenomenon is the PMTs' lack of experience in assessment of geometrical proofs, or that because they perceived the sketch to be an obvious part of the proof, they assumed it would be present in each geometrical proof. The understanding of one's proof by others can be facilitated when the provided proof includes a clear sketch of the problem. Hence this criterion can be ascribed to the category of communication.

**Clear use of mathematical notations**

Gila: It is very difficult to trace a proof course that is written only by mathematical notations. One should add explanations to make it more readable.

Only a few PMTs found this criterion important to include in their criteria set in both tasks. Since a geometrical proof can be written both in a formal way by using mathematical notations or in an informal way by using words, and both ways are usually acceptable by instructors, most of the PMTs did not recognize the need to include such a criterion. This criterion can be recognized as being part of the category of communication since the main purpose of the way the proof is phrased is to communicate with the mathematical community which in this case includes classmates and the class instructor.

**Additional explanations to claims**

Ronit: I enjoyed reading one of the proofs since each step of it was clearly justified both by relevant mathematical theorems and by additional explanations.
Only two of the PMTs included this criterion in the criteria set for the first task and none for the third task. Since the PMTs read the proofs they had to assess and did not find them in most cases difficult to understand, they did not find this criterion important to be included in their criteria set. Additional explanations to claims enhance the understanding of the provided proof by others and hence are part of the communication category.

**Formulation of correct conclusions**

Ella: The first two things that were important to me in the proof I had to assess were that the proof will be correct and the conclusions at every stage will be also correct and justified.

While only one of the PMTs found this criterion important to be included in his criteria set in the first task, 11 PMTs did in the third one. As was stated before, the two tasks differ in their nature. In the first task they had to assess a problem they had already solved and its correct conclusion was obvious for them. However, in the third task each PMT chose a different problem to solve, hence the assessors were not familiar with the problems they had to assess; they found this criterion to be crucial since through its use they could assess the validity and the mathematical correctness of the unfamiliar problem. This criterion can be ascribed to the category of systematization since one of the required attributes of a structured proof is the providing of correct conclusions.

**Demonstration of connectedness between mathematical areas**

Dan: One of the proofs I checked was so original, it included ideas connecting algebra and geometry. It made me wonder about how humans think. It was easy to give a high score to such a proof.

Only two PMTs included this criterion which can be attributed to the systematization role of proof in their criteria set. In the PMTs' reflections on the whole process of the assessment most of them admitted that they began by reading the proofs and then constructed assessment criteria to match the proofs. Retrospectively, however, they believe that the order should have been reversed: "The proof should have been judged according to presumed standards and not vice versa". The construction of the criteria set after reading the proof, can explain why Dan and another PMT included the criterion 'demonstration of connectedness between mathematical areas' in their criteria set.

To conclude, all the PMTs' reflections referring to the content of the feedback were uttered from the assessor perspective. As can be observed in Table 1, the PMTs referred to various aspects of geometrical proof such as verification, communication and systematization. As was
stated by the PMTs, many of the criteria were added to the criteria set after they read the proofs to be assessed.

4.3 The effect of the peer assessment on the PMTs' mathematical work

During the reflection on peer assessment as assessors after the first task, many of the PMTs referred to the effect of the peer assessment on their own mathematical work from two different perspectives. First, observing a different solution from the one they had provided in the first task, made them realize there are many different ways to solve the problem. Second, observing a solution which is more elegant and effective from the one they had provided, made them rethink the solution they had provided and look for ways they could improve it. In the third task, the assessment process enabled the students to be exposed to variety of problems and solutions which enabled them to expand their mathematical knowledge.

Usually the assessment process is done by the teacher whose mathematical knowledge is way beyond that of her students and as a result the assessment process is often done offhand (Hansen, 1991; McMillan & Forsyth, 1991). In our study the situation is different - the PMTs act at the same time both as assessors (taking the teacher’s role) and as students (being assessed by classmates). When taking the teacher’s role in this situation, they were exposed to different solution strategies which enabled them to view new mathematical ideas they did not think of. This situation was described by Reese-Durham (2005), who asserted that in reviewing the work of peers, students gain insight into their own work, as well as developing their ability to make judgments, which is an essential skill in professional life. As was reflected by 13 of the PMTs in similar words:

Ronit: When I read one of the proofs I had to assess I was impressed by its simplicity which made it very easy to follow, and during the whole reading I thought why I couldn’t think the same way.

This exposure was intensified in the third task in which each PMT had to assess two new problems that were totally different from the one s/he chose to focus on, a process that enabled her/him to expand her/his mathematical knowledge. As was reflected by seven of the PMTs in similar words:

Ruth: When I read the problems of my classmates in the third task it made me think about the variety of the human thinking and the richness of the mathematics. I wondered how students in my class with similar knowledge to mine can come with so interesting ideas which are so different from the ones I chose.

Six of the PMTs referred to the following issue:
Dan: After the third task, I realized that it was a different kind of learning. Instead of learning geometry in the traditional way, I learnt it through my engagement in the assessment process which was more intensive and more interesting for me – it was different.

In order to assess their classmates’ solutions, six of the PMTs said that they needed to rehearse the relevant concepts to be able to judge the correctness of the assessed solutions. By rehearsing with the relevant learning materials they experienced a different way of learning.

In the PMTs’ reflections, an additional issue that was raised was the PMTs’ doubts regarding their ability to assess the mathematical correctness of solutions. Three of them referred to this issue in similar words:

Gila: In the third task I had to assess new problems that were not familiar to me and I was not sure whether the provided solutions were mathematically correct so I turned to the course instructor for help.

In her reflection Gila referred to the problematic situation in which the assessor does not possess the requisite mathematical knowledge to judge the correctness of the solution. In cases like this the PMTs chose to be helped by the course instructor. This problem can occur when school students are engaged in peer assessment activities, however since in our case the peer assessment was done by PMTs that were on the verge of becoming mathematics teachers, we believed that the occurrence of such cases will be infrequent.

4.4 Assessment of the first task vs. the third task – PMTs' experiences

After the third task we asked the PMTs to reflect on the whole process and to refer to the assessments they had to provide and receive. Their reflections refer to three main issues: the construction of the criteria set; their experience as assessors and their experience as assessee.

4.4.1 Construction of the criteria set

More than two thirds of the PMTs reported on encountering difficulties in constructing criteria set in the first task ascribing their difficulties to their lack of prior experience. As was reflected in similar words:

Ruth: I had difficulties deciding whether the categories should relate to the methods, techniques, line of thought, or perhaps the final solution…. How to weight them? Which is more important? Which deserves a higher weight?

From the above quote it appears that the PMTs were uncertain as to the nature of the criteria set - what should they assess? The PMTs also had hesitations concerning the weight they should assign to each category, assuming that the greater the criterion's weight is, the greater
is its importance to the process of solving the geometrical problem. The PMTs argued that the weight itself "bears a certain message"; namely, "if a teacher wishes to educate his students to work in a certain way, s/he should weight specific categories in accordance".

Nevertheless, in both tasks the first and the third the PMTs refer to roles and meanings of proof while constructing their criteria set. In the first task the PMTs constructed 12 criteria referring to roles and meanings of mathematical proof while in the third task there were only nine criteria. As was previously mentioned, in both tasks, the PMTs reported that the construction of the criteria set was done after they had read the proofs they had to assess. Hence this could explain why part of the criteria of the first task did not appear in the list of the third task (e.g. criterion 3). Observation of all the provided proofs in the third task revealed that they did not include numerical values; hence, the PMTs did not find the criterion referring to calculation errors to be necessary. As to criteria 8, 9 and 11 we wondered why the PMTs did not include them as well in their criteria set for the assessment of the third task. A possible explanation is that these criteria were used only by PMTs who found reference in the proofs they had to assess in the first task but not in the third one.

After the third task, however, there were no references in the PMTs' reflections regarding the encountering of difficulties while constructing the criteria set. The fact that the PMTs did not specify any difficulties regarding the construction of the criteria set for the third task may imply on the developing of the construction of criteria set skills.

As we mentioned earlier, we also constructed our criteria set for the assessment of the first and the third tasks and compared the PMTs criteria set of each task with the one constructed by us respectively. These comparisons revealed that both lists were very similar. However, since not all the PMTs constructed the same criteria set (see the frequency column in Table 1), we believe that there is more to be done in the process of the PMTs' developing their abilities to construct criteria set according to the roles and meanings of proof for the assessment of geometrical proofs.

4.4.2 The PMTs' experience as assessors

The PMTs referred in their reflections to the feedback they provided as assessors. The PMTs who reported that they changed their feedback from the first to the third task were those who had received feedback on the first task that was significantly different from the feedback they had provided as assessors. As was reflected in similar words by five PMTs:

   Ella: After I read the feedback on my work, which was detailed and included hints for improvement, I realized that the feedback I provided was not effective.
Hence in the third task I invested many efforts to provide a detailed and a better feedback.

In cases where the situation was opposite, namely, feedback provided as assessor was significantly better than feedback received as assessee, the PMTs reaction was expressions of frustration. They expected feedback similar in nature to what they had provided and were disappointed. Experiencing disappointment made them recognize the importance of the feedback as was reflected by four PMTs in similar words:

Ruth: In the assessments I gave I did my best to justify each score carefully and although my work did not receive a similar attention, I still believe that assessment of students should be done seriously since they can learn a lot from the feedback they receive.

From the PMTs' reflections after the third task as assessee, we can learn that complaints about the nature of the received feedback became rare. In this case, we can infer that the PMTs become aware to the consequences of giving a poor or negative feedback in the assessment process.

4.4.3 The PMTs' experience as assessee

Receiving detailed and useful feedback or harsh or minimal feedback led to two kinds of reactions from the assesses. As to the first kind, PMTs who received such a feedback praised it specifying its merits. Three PMTs said in similar words that "such a detailed feedback helped me to understand why the score I received was low and to accept it. Moreover, it gave me the feeling that my work was treated seriously". On the other hand, harsh or minimal feedback was perceived by the PMTs as negligence. The PMTs' reactions to the harsh feedback they received as assessee were expressions of frustration and disappointment. They interpreted harsh feedback to be a result of their assessor's lack of tolerance towards their weaknesses and difficulties as was reflected by one PMT "the intolerant expressions in one of the feedback I received left me with the feeling that the student who had to assess my work had no empathy with my difficulties".

Looking at the reflections after the third task reveals that there were no indications regarding the receiving of harsh feedback. We believe that the PMTs exposure to both sides of the coin—acting at the same time as assessors and assesses confronted them with strengths and weaknesses of the assessment process and as a result enabled them to develop their sensitivity to various aspects of the assessment process.

Nevertheless, review of the PMTs' assessment of the first task reveals that most of the feedback was phrased in a polite manner. In the PMTs reflections on the peer assessment process most of them referred to the problematic situation in which they have to assess and be
assessed by their classmates. Having this situation in mind, made them choose their feedback carefully in a way that, on the one hand, will not be offensively phrased but, on the other hand, will enable the correction of the assessed solution. This is in line with Falchikov (1995) who found that students have difficulties in assessing their friends especially in small socially cohesive groups.

Examination of the PMTs' assessments of the third task shows an improvement in the explanations provided. This improvement can be observed by the decrease of the gap between the criteria set provided by the PMTs and the one constructed by us. There were no complains about receiving harsh feedback and most of the PMTs reported the feedback they received included references to the correct parts of the proof, and only then suggestion of alternatives to correct the faulty parts of it.

5. Concluding Remarks

From the above results and discussion it can be concluded that:

First, comparison between the first and the third assessment tasks revealed that the PMTs developed their abilities to select a criteria set that addresses the roles and meaning of mathematical proof and assigned a numerical weight to each criterion according to its importance. A decrease in the number of the criteria and in the standard deviation of the assigned numerical weights was observed between the first and the third assessment tasks. In addition, a decreasing tendency was observed in the gap between the scores each work received from two different PMTs. Also, between the first and the third tasks, a decreasing tendency was observed in the gap between the criteria set constructed by the PMTs and the one constructed by us.

Second, the PMTs' reflections on the assessment process show that their engagement in peer assessment helped them develop a sense of the roles and meanings of geometrical proofs and to be sensitive to the feedback they provide as assessors. This development may indicate an improvement in their assessment skills that are essential to their professional development as future mathematics teachers (Lovemore & David, 2006; Sluijsmans & Prins, 2006; McGatha et al., 2009).

Third, in their reflections the PMTs refer to the feedback they provided as assessors and received as assessees. From assessor’s perspectives, the PMTs referred to the categories according to which they assessed their classmates work. Most of these categories refer to roles and meanings of geometrical proof. However, from assessees’ perspectives, the PMTs referred mainly to the nature of the feedback. Namely, the writing style of the feedback, the inclusion of harsh/encouraging comments and so on. From the PMTs' reflections on the whole
process we may say that by exposing the PMTs to peer assessment in which they had to provide and receive feedback simultaneously, helped them develop their sensitivity to various aspects of assessment.

Fourth, the richness of the expressed reflections indicates the powerful impact the engagement in peer assessment, both as assessors and assessee, had on the PMTs.

Fifth, the effect of the peer assessment on the PMTs’ mathematical work came to fruition in them realizing that there are many different ways to solve the problem and in rethinking of the solution they had provided, looking for ways they could improve it. They came to the first conclusion after observing different solutions to the problem in the first task which was different from the one they had provided. They came to the second conclusion after observing solutions which were more elegant and effective than the one they had provided. In the third task, the assessment process exposed students to a variety of problems and solutions which enabled them to expand their mathematical knowledge. By reviewing the works of their peers they gained insights into their own performance (Berkencotter, 1995).

Finally, in order to substantiate the findings of this study, this research should be repeated with a larger group of PMTs.

References


### Appendix 1:

#### Work assessment form – phase 1

<table>
<thead>
<tr>
<th>Assessor name:___________</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assesee no.:_________</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion</th>
<th>The maximal weight assigned to criterion and explanation</th>
<th>My scoring for this criterion</th>
<th>Explanation for my scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

General comments to the assessee:
_______________________________________________________________________________________
_______________________________________________________________________________________
_______________________________________________________________________________________

### Appendix 2:

#### Asseesee name:___________ | Date:___________

<table>
<thead>
<tr>
<th>Phase 2: Peer assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part A: My attitudes to the assessments I received</td>
</tr>
<tr>
<td>1. Did the assessments you received provide you with constructive feedback which can assist you in improving your work?</td>
</tr>
<tr>
<td>2. Did the evaluations include offending comments? If yes- describe their nature.</td>
</tr>
<tr>
<td>3. Did the evaluations you received enable you to see the problem and its solution in a different way?</td>
</tr>
<tr>
<td>4. What parts of the evaluations caused you to revise your work and why?</td>
</tr>
<tr>
<td>5. What parts of the evaluations did not help you to revise your work and why?</td>
</tr>
<tr>
<td>6. Did you find criteria in the evaluations you received that did not occur to you? If yes- describe the criteria you decided to 'adopt' and why.</td>
</tr>
</tbody>
</table>

#### Part B: The assessments I gave to my classmates

| 1. Did the engagement in peer assessment cause you to alter your solution? Why? |
| 2. Describe your doubts and hesitations while engaging in peer assessment. |