The investigation of peer assessment in primary school cooperative learning groups with respect to gender

Irfan Yurdabakan*

9 Eylul University, Buca Education Faculty, Department of Educational Sciences, Izmir, 35160, Turkey

(Received 26 April 2009; final version received 3 July 2009)

There are studies especially at higher education level investigating the subsequent responses of students towards reciprocity, tacit agreement and assessment of peers, but research on the effect of gender on peer assessment is limited. The present study focuses on whether peer assessment used in cooperative learning groups varies with respect to gender and investigates the compatibility level of peer assessments with teacher grades. This study was conducted in a primary school fourth grade social sciences course with 46 participants, 28 female and 18 male, their ages ranging from 9 to 10. The study rendered different results of peer assessment, where male and female students scored their fellow and opposite sexes with respect to their contribution to group work and their learning levels. The compatibility between female and teacher scores was higher than male and teacher scores.

Keywords: constructivism; cooperative learning; primary school; peer assessment; gender

Introduction

In recent years, there has been an increase in the number of studies that investigate the peer assessment used in group work due to its compatibility with the social constructivist approach. Unfortunately, most of these studies are at higher education level (Birenbaum and Dochy 1996). However, in order for the learning activities that are in line with the constructivist theory to be successful, starting from the primary education level, the development of many skills such as peer assessment would be useful. While many researchers argue about the numerous benefits of peer assessment used in group work, some verbalise their concerns about peer assessment. These writers are concerned about the fact that factors such as reciprocity, tacit agreement, shared values, visual affection, beliefs, gender and racial differences may affect the success of peer assessment in a negative way. In the literature review, there are various studies to test these worries, but no research has been encountered that focuses on the effect of gender on peer assessment, especially at the level of primary education. That is why this study aims to find out whether peer assessment used in the primary education fourth grade group work varies with respect to gender, and to
investigate the relationships between the results of peer assessment and teacher evaluations.

**Background**

Constructivism started to gain importance with the second half of the twentieth century, thanks to the works of Piaget, Vygotsky, Ausubel, Bruner and Von Glasersfelt. Constructivism, which emerged as an epistemology of knowledge, provides the educators with a wide conceptual framework ranging from the organisation of learning environments to the construction of evaluation processes. Therefore, today, the evaluation/assessment process involves innovations that question the learning process, while the learning environments change from the traditional into more student-centred.

In the history of epistemology, the view of knowledge has changed from the static and passive into the more variable and active (Heylighen 1993). According to the objectivist theory, knowledge is invariable because the basic properties of objects can be known and are comparatively stable. From the perspective of the objectivists, learning can be defined as observable behaviours or the process of change in behaviours as a result of the individual’s reactions to the events around him/her. The objectivists believe in the existence of accurate knowledge about the world. For the learners, the aim is to profit from this knowledge, whereas the educator aims to transfer it. Yet constructivism based on the work of Piaget and Vygotsky (Bransford, Brown, and Cooking 2000) argues that knowledge and reality do not have any absolute value or at least it is impossible for us to know this reality. The individual constructs a reality on the basis of his/her experiences and interactions with the environment and then interprets it (Jonassen 1991).

Constructivism defines teachers as individuals who do not transfer their knowledge mechanically, but as facilitators who play an important role in the understanding and construction of knowledge as well as motivating and providing the students with the learning opportunities. For the students they are defined as people who make meanings (Mayer 1996). But, as opposed to Piaget, who claims that teaching is primarily an individual activity Vygotsky, the acclaimed founder of social constructivism, emphasises that education is a process of social change and learning occurs as a result of the interaction between the individual and the environment, so group work is important for the individual’s learning (Richardson 1997). The educators and writers who relate the constructivist theory to learning (Jonassen 1991; von Glasersfeld 1995; Mayer 1996; Gergen 1995) list the basic characteristics of constructivist learning and teaching processes as follows:

1. The student plays a central role in mediating and controlling learning.
2. Knowledge construction is emphasised, not reproduction.
3. Activities, opportunities, tools and environments are provided to encourage metacognition, self-analysis, self-regulation, self-reflection and self-awareness.
4. During the knowledge construction process, the learner’s previous knowledge constructions, beliefs and attitudes are taken into account.
5. Problem-solving, higher-order thinking skills and deep understanding are emphasised.
Collaborative and cooperative learning are favoured in order to expose the learner to alternative viewpoints. Assessment is authentic and interwoven with teaching.

These characteristics guide the composition of the learning processes to a great extent. Also, Collins (1991) lists the basic properties of student-centred learning conditions as below:

1. Instead of traditional class teaching, there is small group teaching.
2. Instead of direct teaching (lecture), there is guidance.
3. More time is spared for tasks and activities.
4. There is a shift towards cooperative learning.
5. Instead of summaries of evaluations, there is a move towards alternative assessment.

According to this classification, with the help of the cooperative learning activities based on small group tasks and activities, constructivism emphasises the learning situations of the familiar approaches of assessment, in which the students are directed into learning.

Cooperative learning and peer assessment

Being one of the active learning methods, which meet the learning needs of the social constructivist approach, cooperative learning can be defined as the learning process in which the students work together in small groups in order to maximise their own and each other’s learning. In order for group work to turn into a cooperative learning process, conditions like group reward, positive interdependence, individual accountability, face-to-face interaction, social communication, assessment of group tasks and opportunity of equal success must be present (Johnson, Johnson, and Smith 1998).

Group reward serves to increase the efforts of group members working together and sharing the same responsibility. During the cooperative learning process, for the establishment of positive interdependence, it is necessary to have a collective task in order to have the students act together; it is necessary to have students work on a collective instrument in order to have them focus on a collective task. Through the learning process, the debates and feedback procedures among the group members compel face-to-face interaction which contributes to the learning of others. The members’ social interaction skills ease this process. The assessment of the group process provides the identification of the positive behaviours that increase group achievement or negative behaviours that decrease group achievement, thus rewarding those behaviours that increase group achievement. Opportunity of equal success means assessment of each member on his/her contribution to group work, ensuring opportunity of effort and success.

There is a strong belief that cooperation in group work with the aim of learning improves both learning and social skills (Slavin 1990). Many studies have shown the benefits of cooperative learning. For instance, cooperative learning helps students to develop interpersonal skills (Slavin 1987), like getting to know and trust team members, communicating effectively and clearly, providing support and challenging fellow team members, along with engaging in constructive conflict resolution
(Johnson and Johnson 1994). These interpersonal skills enable students to acquire a sense of social responsibility (Vermette 1988). In this regard, cooperative learning can be beneficial, especially in a multiracial society. Pate (1988) has found that people of different ethnic backgrounds working together on a task, problem or having the same goal, develop positive feelings as well as mutual respect for each other, which could serve to promote positive feelings and better understanding among the students of different ethnic origins in the long run. In addition to the interpersonal skills, cooperative learning helps to build higher-level cognitive skills and also has a positive impact on student achievement (Michaelsen 1992, as cited in Freeman 1995).

Cooperative learning, which composes an important part of classroom teaching including in the nursery school, depends on the fact that students learn not in an isolated environment but by cooperating with each other (Vygotsky 1978). According to this view commonly known as ‘social constructivism’, individual competence involves not only what a student can do without help but also what he can learn from the experience of cooperative learning groups. During the group work, the performance of each member increases the success of both the individual and the group (Webb 1997). For this reason, the educators emphasise the alternatives that would develop a strong relationship between the evaluation and the class practices (Webb 1997). These alternatives focus on learning from group cooperation as well as individual achievement, group productivity and evaluation of cooperative skills like participating, taking responsibility, receiving and giving feedback (Webb 1997).

Alternative methods of assessment have received much attention in the last decade and several forms of assessment have been used in higher education. As for Turkey, they have been introduced in primary education with the adaptation of the curriculum in primary education to the constructivist theory.

As in the case of Australia and the United States, leading experts of Europe argue that the age of testing has changed into the era of assessment in recent years. The age of testing can be defined as the measurement of the students’ knowledge of a decontextualised subject matter that doesn’t relate to their experiences and the calculation of a single total score for their products due to the dissociation of teaching and testing activities. The era of assessment, however, promotes the integration of assessment and instruction and regards the students as active individuals who share responsibility, reflect, collaborate and conduct a continuous dialogue with the teacher (Dochy, Segers, and Sluijsmans 1999).

The authors advocating alternative assessment approaches say that the assessment process is not only an instrument to give students their diplomas, but it is also a process that leads to the development of students, and it directs the students to better learning activities; and in order to reach the objectives, learning and assessment activities must be used together in learning environments (Boud 1995). Dochy and Moerkerke (1997) state that traditional testing approaches do not abide by the aims of life-long learning, careful thinking, being critical, assessing oneself, and problem-solving skills.

One of the alternative methods of assessment – peer assessment – is the process in which individuals in a group assess their peers. Peer assessment is not merely a grading procedure, but also a learning activity that develops skills (Falchikov 1995; Freeman 1995). Peer assessment, in which the learning and assessment activities coexist, can provide the essential conditions for cooperative learning: reducing the
effect of the ‘hitchhikers’, who fail to fulfill their responsibilities in team work, but manage to get the same score as their more successful team mates (Kaufman, Felder, and Fuller 1999). Peer assessment can contribute to the students’ active participation and reflection on the learning process by questioning the learning of their peers.

Many authors have mentioned the benefits of peer assessment used during group work. For example, peer assessment helps the students to get feedback from others besides their teachers. Also, peer assessment ‘stresses skills, encourages lesson participation, increases concentration on learning, provides feedback to the students, increases course attendance and teaches responsibility’ (Weaver and Cotrell 1986, 25). According to many other authors, peer assessment develops critical thinking (Searby and Ewers 1997; Stainer 1997), increases student learning (Michaelsen 1992, as cited in Freeman 1995) and encourages cooperative learning as opposed to competitive learning (Lejk and Wyvill 2001; Orsmond, Merry, and Reiling 1996).

According to the results of some qualitative studies investigating participants in cooperative learning groups, where peer assessment was utilised, the students stated that these kinds of studies increase active participation and said that peer assessment was an interesting experience for them (Lourdusamy and Divaharan 2000). Actually, in addition to the fact that cooperative learning where peer assessment was used was found to be rewarding and encouraging, it was also observed that students expected their group members to take the work more seriously and participate more, this way strengthening the sense of interdependence among group members (Purchase 2000). On the other hand, as Yueh and Alessi (1988) stated, one way of getting students to take part more actively in lessons is to reward their interests and participation in the lesson. Some studies done on this issue have shown that using peer assessment to reward student efforts was encouraging and motivating. Also as Conway et al. (1993), Goldfinch (1994) and Freeman (1995) stated, peer assessment was effective in checking those students who neglected their duties and added that the students who took part in assessment found it a good experience for them.

The studies carried out by the others, especially in higher education, and their results are as follows. In their studies, Dancer and Dancer (1992) showed that peer assessment had no validity as the students were prone to base their assessments on uniformity, race and friendship if they had no extensive training on peer assessment. In contrast, Topping (1998), in his study where he analysed 31 studies, found that 18 of those studies implemented in different areas had acceptably high levels of validity and reliability and only seven studies had unacceptably low levels of validity and reliability.

In their study, Oldfield and Macalpine (1995) searched the student skills in assessment. The results of the study revealed that there was a high correlation between the student and teacher grades for the individual essays and presentations. Fry (1990) conducted a study where teacher assessment and peer assessment were performed together. At the end of the study, he found high correlations between the teacher assessments and the students’ peer assessments.

On the other hand, Orsmond, Merry, and Reiling (1996) did a study where first students and later teachers graded the same answer sheets without seeing the grades of each other. The end result of the study showed that there was a significant and medium level (0.54) correlation between the grades of the students and teachers.

Boud (1995) investigated the students’ abilities to assess themselves and their peers. At the end of the study, the students found it quite useful to compose the assessment criteria, graded themselves higher than their peers and lower than their
teachers. Generally high congruity was observed between the assessments of peers and teachers.

The purpose of the study
As it is mentioned above, even though there is an increase in the benefits of using peer assessment during group work, same authors reveal their concerns over reciprocity, tacit agreement, shared values, visual affection and belief (Magin 2001; Edgerton and McKechnie 2002; Parsons and Drew 1996; Rafiq and Fullerton 1996), gender as well as racial differences, and they also fear that the other members may be offended by the assessment results or conflicts may arise among group members (Kaufman, Felder, and Fuller 1999). According to Abson (1994) and Edgerton and McKechnie (2002) these concerns may influence the validity and reliability of peer assessments.

In their study, where they analysed the studies of peer assessments, Dochy, Segers, and Sluijsmans (1999) classified the main findings of difficulties about peer assessment under four titles: validity, fairness, accuracy and effects. They related the factors such as reciprocity, tacit agreement, shared values, visual affection, belief, gender and racial differences to the validity of the peer assessment; compared to traditional assessments, they related student perceptions on the honesty of peer assessment to fairness; they related the harmony between the peer and teacher assessment to accuracy; and they related the benefits of peer assessment on individual and group achievement and attitudes of students on peer assessment to effects (Divaharan and Lourdusamy 2002).

As in the case of all the other alternative methods of assessment, there are reservations about the validity and reliability of peer assessment. Some writers (Cho and Schunn 2003; Dochy, Segers, and Sluijsmans 1999) emphasise that it is necessary to ensure whether peer assessment is valid or reliable and also to develop or revise the current practices.

In many studies, the ways of seeking evidence for the validity and reliability of peer assessment are varied (Dancer and Dancer 1992; Fry 1990; Boud 1995; Oldfield and Macalpine 1995; Orsmond, Merry, and Reiling 1996). But, Dochy, Segers, and Sluijsmans (1999) have pointed out the effects of factors like reciprocity, tacit agreement, shared values, visual affection, belief, gender and racial differences on validity, while they stated that the compatibility between the peers’ scores and the teachers’ scores was related to reliability.

This study aims to explore (1) whether the peer assessment used in cooperative learning environments changes according to gender at the fourth grade level in primary school, and (2) the level of compatibility between the peer assessments and teacher-made achievement test scores. The first purpose of the study can be related to validity as the effect of gender on peer assessment is studied, whereas the second purpose of the study can be related to reliability as the compatibility (accuracy) between the students’ assessments and teachers’ assessments is studied. On the other hand, most of the studies on peer assessment are at tertiary education level. This study can be significant because it deals with the subject of peer assessment at primary education level.

In this study, peer assessment in cooperative learning groups was performed according to the following two approaches (Webb 1997). In the first one, the peers evaluated other members with respect to cooperative skills like taking responsibility, performing tasks, contributing to others’ learning, participating in discussions, and
fulfilling one’s role. In the second one, the peers evaluated the others in terms of their level of learning which derived from the group cooperation and was in tune with the learning process. As for the teacher assessment, a teacher-made achievement test, which complied with the attainments of the cooperative learning process, was administered. The results of this study are limited to the data obtained according to these aims. In addition, the peer assessment results of males and females were compared by administering $t$-test statistics. On the other hand, the relationships between the students’ peer assessment scores in general, their peer assessment scores with respect to gender, and the teachers’ scores of assessment were calculated with Pearson Product Moment Correlation.

**Methodology and participants**

This study was conducted in a primary school fourth grade social sciences course, in a class of 46 (28 female, 18 male) pupils whose ages were 9–10. The school is situated in Izmir, a western province of Turkey. The school is a public school and because of its central location and experienced teaching staff, it is a highly popular one. That is also why the classrooms are quite crowded. But with the implementation of the new primary school curricula, there are efforts to gradually decrease the number of students in each class to 30. The course of social sciences is a three-hour course (120 minutes) in a week, where topics related to civic education, history, geography, sociology and anthropology are taught at the level of primary education. The present study is based on the results of those peer and teacher assessments performed at the end of a 12-week cooperative learning practice.

The programmes of primary education in Turkey have been redesigned according to the constructivist approach since 2005. According to the evaluation report of the Education Reform Initiative of Ministry of National Education (2005), these programmes place the students in the centre. Such skills as critical thinking, creative thinking, communication, problem-solving, research, decision-making, using information technologies, entrepreneurship, and giving importance to individual and social values are considered important in these programmes. Those activities that would enable students to take part in the research, inquisition, problem-solving and decision-making processes were suggested. Also, learning ‘by doing-thinking’ activities were emphasised and the use of cooperative learning strategies was highlighted. In these new programmes, evaluation and assessment were considered a way to give a decision not only on students’ final learnings, but also on their learning processes. New evaluation and assessment methods, including alternative assessment applications (self-, peer and portfolio assessment) and different ways of teachers assessing students were included in these programmes. Even though there are several problems considering the compatibility of the courses, course contents and applications to integrative approaches (Brundrett and Silcock 2002), still some kind of parallelism can be drawn with the skills integrated in these programmes with those in other countries (like Australia, England, Ireland, the United States, New Zealand, Spain, Finland, Ireland, Israel, Austria, Canada and Singapore). As a result of this change, the course of social sciences now involves student-centred, activity-based group work. In the study group, part of the social sciences course was taught with the cooperative learning method for 12 weeks. During the activities, some techniques like learning together, academic conflict and expert groups were utilised. The practice was carried out by the class teacher, who has an MA in
Curriculum Development and Instruction and is experienced in putting the cooperative learning method into practice. The class teacher also took the opportunity to experience peer assessment during this study.

The cooperative learning groups were organised by taking the students’ gender, achievement and social communication skills into account. Nine heterogeneous groups of four to six were formed with respect to achievement and social skills of the students on the basis of teacher observations and anecdotes about the students and the group members were assigned randomly to each group.

In the first session, the purpose of cooperative learning activities and the way the groups would work were explained; roles of group members were introduced and assigned. In the following sessions, by working cooperatively the students went on with their studies with the help of the study sheets prepared beforehand.

Also in the first session, peer assessment was introduced and general information was given about the purpose of peer assessment and how it would be conducted. The criterion related to peer assessment was discussed in the group and designated criteria were introduced to the students. In this session, the students were informed that at the end of cooperative learning activities there would be peer assessment, the purpose and criteria of which were introduced to students, and a test would be given compatible with the peer assessment criteria.

In the subsequent sessions, cooperative learning activities were carried out with the help of the study sheets that were prepared beforehand, while studies were implemented at the end of activities to informally improve the peer assessment skills of students.

In the twelfth session the students were seated separately and asked to assess their peers on the peer assessment forms, details of which will be explained later. The teacher-made test was administered at the end of this peer assessment process.

Even though there are some questions raised these days considering the teachers’ implementation of these programmes, practices like end-of-unit self-assessments and occasional peer and co-assessments are made use of. For these reasons, the students in the study group are quite experienced in completing the peer assessment forms and they had no difficulty at all.

Instrumentation

As the purpose of this study was to analyse the peer assessment results of primary school level students and identify their relationships with the teacher assessments, three different instruments were used to gather the necessary data. In the development of these instruments, the opinions of Webb (1997) were taken as basis. The first one was the Peer Observation Form (POF) where students assessed their peers’ participation in cooperative learning activities; the second one was the Peer Assessment Form (PAF) by means of which pupils assessed the other group members’ learning levels; and the third one was the teacher-made Achievement Test (AT) which covered the 12-week practice and was used to test the learners’ achievement levels.

The POF was prepared by considering the criterion that enabled students to assess cooperative skills like participation as a team member, taking and carrying out responsibility, giving and receiving feedback and contributing to others’ learning, together with the peers’ expectations of the other group members and the factors necessitated by cooperative group structure. The POF consisted of 13 items. During peer assessment, the peers marked one of the choices on each item designed as
'always = 3', ‘sometimes = 2’ and ‘rarely = 1’. According to the POF items, one student, for example, assessed one of his/her peer’s contribution to group work as ‘always’ if he/she thought that peer was constantly contributing to group work; or he/she would mark the same item for another peer as ‘rarely’ if he/she thought that particular peer’s contribution to group work was not enough. One member’s scores were calculated by taking the average of the points given by the other members. The followings are example items for the POF:

(1) Brings all his materials.
(2) Helps the group members.
(3) Carries out his responsibilities.
(4) Contributes to group work.

The PAF is a form that contains cognitive properties in line with the objectives of the cooperative learning sessions. The students assessed the group members’ performance according to the criteria in this form. The PAF consisted of 15 items. During the peer assessment, the peers marked one of the choices on each item designed as ‘good = 3’, ‘medium = 2’ and ‘poor = 1’. A student assessing one of his/her peers would mark ‘good’ if he/she thought his/her peer would give a good answer or a good definition; or he/she would mark ‘poor’ if he/she thought his/her peer would not be able to give a good answer or a good definition. One member’s scores were calculated by taking the average of the points given by the other members. The example items for the PAF are as follows:

(1) Knows the class and school rules.
(2) Gives the definition of ‘budget’.
(3) Knows the innovations provided by the Republic.
(4) Knows his basic needs.

The Achievement Test (AT) was comprised of 30 items that were prepared in concordance with the objectives of cooperative learning practices, and the criteria and items of the PAF. The items of this test were derived from a 60-item test that was tried before by those fourth grade students who had already mastered the same topics. The difficulty index of the selected items ranges from .32 to .75 and the discrimination index of the items ranges from .42 to .93. The KR-20 inner-consistency coefficient is .90. The following are some of the example items that were also in concordance with the PAF items mentioned above:

(1) Which of the following is one of the rules we should abide by while choosing the prefect of a class?
   (a) The one with the least votes is chosen
   (b) The chosen one thanks the class (√)
   (c) The ones chosen by the teacher are nominated
   (d) Those that are not chosen can be nominated again
(2) Which of the following is the definition of ‘budget’?
   (a) It is the money earned
   (b) It is a product being expensive
   (c) It is our income and expenditures (√)
   (d) It is the money spent
(3) Which of the following cannot be listed as one of the benefits of the republic to our country?
   (a) Adopting the new Latin alphabet
   (b) Giving the women the right to vote and to be elected
   (c) Adopting the Act of Surname
   (d) Passing the management rights from father to son (√)

(4) ‘Feeding and shelter’ are what kind of needs?
   (a) Social needs
   (b) Basic needs (√)
   (c) Consumption needs
   (d) Production needs

Results

According to the aims of the study, the results were handled in two ways. First, in cooperative teaching groups both the participation of members in group activities (POF) and the differences among assessments of peers of the others on their learning levels (PAF) were investigated according to their genders. Second, the relationship between peer assessment scores and teacher-made achievement test (AT) scores were calculated in terms of both participation in group activities (POF) and learning levels (PAF) and the accuracy levels of group members in their peer assessments were analysed.

The differences between POF and PAF scores according to gender

The average of the scores, their standard deviations and t-statistics of the findings obtained from cooperative learning group members’ POF and PAF assessments about the male and female students are given in Table 1. (Note that results of insignificant findings were excluded from the table; only the significant findings are presented.)

According to the table, there is a significant difference (d = .65, moderate practical importance (Cohen 1987)) among the girls’ POF scores they received from boys and girls. This result shows us that considering their contribution to group activities, girls were assessed differently by the boys and other girls. Considering the

Table 1. The differences between POF and PAF scores according to gender.

<table>
<thead>
<tr>
<th></th>
<th>Mean score</th>
<th>SD</th>
<th>t-value</th>
<th>Cohen’s D</th>
</tr>
</thead>
<tbody>
<tr>
<td>From male to female (POF)</td>
<td>35.92</td>
<td>3.40</td>
<td>−2.34*</td>
<td>.65</td>
</tr>
<tr>
<td>From female to female (POF)</td>
<td>33.18</td>
<td>4.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From male to male (POF)</td>
<td>35.00</td>
<td>2.94</td>
<td>−2.83**</td>
<td>.98</td>
</tr>
<tr>
<td>From female to male (POF)</td>
<td>31.78</td>
<td>3.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From male to male (PAF)</td>
<td>39.19</td>
<td>2.23</td>
<td>−2.63**</td>
<td>.91</td>
</tr>
<tr>
<td>From female to male (PAF)</td>
<td>36.72</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From female to female (PAF)</td>
<td>38.39</td>
<td>2.87</td>
<td>−1.86*</td>
<td>.56</td>
</tr>
<tr>
<td>From female to female (PAF)</td>
<td>36.72</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.
averages, it can be seen that girls received higher scores than boys. The male participants found the female participants more active.

The differences among the POF and PAF average scores of boys given by the girls and the other boys were found to be statistically significant ($d = .98$ for POF and $d = .91$ for PAF, crucial practical importance). Boys were assessed differently by the girls and the other boys in the groups. The male participants scored the girls higher than their fellow members.

The next step was to compare the scores given by the boys to girls and to other boys and the scores given by the girls to boys and to other girls to see if the peer ratings differ according to gender from the perspective of POF and PAF.

No significant difference was found among the POF ($t = .89, p > .05$) and PAF ($t = .49, p > .05$) scores the boys gave to the girls and to the other boys. While scoring the boys and girls in their groups, the boys gave equal scores. That means the boys rated both the girls and the other boys equally.

While there was a significant difference ($d = .56$, moderate practical importance) among the PAF scores given by the girls to the boys and to the other girls in favour of the girls, no difference was found among the POF ($t = -1.05, p > .05$) scores. The girls gave higher PAF scores to the other girls than they did to the boys. The girls stated that the other girls have a higher learning level than the boys.

There is no significant difference among the POF ($t = 1.35, p > .05$) and PAF ($t = .95, p > .05$) scores the girls gave to the other girls and the boys gave to the other boys. In other words, girls and boys assessed their fellows equally both with respect to their participation in activities and their learning levels.

**The relationships among the POF, PAF and their AT scores for all members**

In order to find out whether these assessments were influenced by gender, and find out whether the girls or boys gave their fellows more accurate scores, the correlations among the POF, PAF and AT scores of girls and boys were analysed. The relationships among the POF, PAF and AT scores were analysed both for all members and for boys and girls separately and the correlation coefficients are given in Table 2.

According to the table, it can be seen that the correlations between POF, PAF and AT change between .92 and $-30$. A strong correlation of $r_{(46)} = .84, p < .01$ was found between the POF scores where all members assessed their peers for the participation in group activities and PAF scores where all members assessed their peers’ level of learning. This relationship shows that the members’ contribution to group activities (POF) was assessed parallel to their learning levels (PAF). The correlation between the POF and AT scores was found as $r_{(46)} = .17, p > .05$. This relationship is weak and not significant. On the other hand, a significant correlation of $r_{(46)} = .32, p < .01$ was found between AT scores and PAF scores where peers assessed the learning levels of others. This relationship is stronger than the one between POF and AT scores.

When we look at the scores girls received from the other members, we can see that there is a correlation of $r_{(28)} = .85, p < .01$ between the POF and PAF scores, $r_{(28)} = .37, p < .01$ between the POF and AT scores and $r_{(28)} = .55, p < .01$ between the PAF and AT scores. The correlation between the POF and PAF scores of the girls is significant and strong. Moreover, the relationship between the POF
scores given to girls and their AT scores is important. However, in this section a more interesting relationship is the one between the girls’ PAF and AT scores. This relationship is significant and above average level.

A correlation of $r_{(18)} = .79$, $p < .01$ was found between the POF and PAF scores the boys received from the other members, $r_{(18)} = -.23$, $p < .01$ was found between the POF and AT scores, and $r_{(18)} = -.19$, $p < .01$ was found between PAF and AT scores.

There is a correlation of $r_{(28)} = .71$, $p < .01$ between the POF and PAF scores the girls gave to the other girls, a correlation of $r_{(28)} = .37$, $p < .01$ between the POF scores and AT scores, and a correlation of $r_{(28)} = .40$, $p < .01$ between the PAF scores and AT scores. There is a nearly strong agreement between the POF and PAF scores the girls received from the other girls, and a medium level significant correlation between the PAF, POF and AT scores of those girls.

It can be seen that there was a correlation of $r_{(18)} = .79$, $p < .01$ between the POF scores which the girls gave to the boys in the groups for their levels of attendance to group activities and PAF scores where the girls assessed the boys’ learning levels, whereas the correlation between the POF scores and AT scores was $r_{(18)} = .40$, $p < .01$, and the correlation between the PAF scores and AT scores was $r_{(18)} = .25$, $p < .05$. While the girls assessed the boys’ levels of participation in group activities (POF) and their learning levels (PAF) similarly, when they are associated with AT scores, it can be deduced that the girls assessed the boys’ levels of participation in group activities more accurately than they did their learning levels.

The correlation between the POF and PAF scores the boys gave to the other boys is $r_{(18)} = .46$, $p < .01$, the correlation between the POF scores and AT scores is $r_{(18)} = -.30$, $p < .01$, and the correlation between the PAF scores and AT scores is

Table 2. The relationships among the POF, PAF and their AT scores for all members.

<table>
<thead>
<tr>
<th>From all members to other members</th>
<th>PAF</th>
<th>AT</th>
</tr>
</thead>
<tbody>
<tr>
<td>POF</td>
<td>.84**</td>
<td>.17</td>
</tr>
<tr>
<td>PAF</td>
<td>.32**</td>
<td></td>
</tr>
<tr>
<td>From the other members to female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.85**</td>
<td>.37**</td>
</tr>
<tr>
<td>PAF</td>
<td>.55**</td>
<td></td>
</tr>
<tr>
<td>From the other members to male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.79**</td>
<td>-.23**</td>
</tr>
<tr>
<td>PAF</td>
<td>-.19*</td>
<td></td>
</tr>
<tr>
<td>From female to female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.71**</td>
<td>.37**</td>
</tr>
<tr>
<td>PAF</td>
<td>.40**</td>
<td></td>
</tr>
<tr>
<td>From female to male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.79**</td>
<td>.40**</td>
</tr>
<tr>
<td>PAF</td>
<td>.25*</td>
<td></td>
</tr>
<tr>
<td>From male to male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.46**</td>
<td>-.30**</td>
</tr>
<tr>
<td>PAF</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>From male to female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POF</td>
<td>.92**</td>
<td>.28*</td>
</tr>
<tr>
<td>PAF</td>
<td>.28*</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.
As can be seen, the correlations between the assessment results of the boys given to the other boys in the groups are different from the ones found until now. While the correlations between the POF scores given by the girls to the boys and the other girls in the groups and their AT scores vary from .37 to .40, the correlation between the POF scores where the boys assessed the other boys and their AT scores is -.30 (p < .01).

The correlation between the POF and PAF scores the boys gave to the girls is $r_{(28)} = .92$, $p < .01$, the correlation between the POF scores and AT scores is $r_{(28)} = .28$, $p < .05$, and the correlation between the PAF scores and AT scores is $r_{(28)} = .28$, $p < .05$. The boys’ assessment of the girls’ participation in group activities (POF) and their learning levels (PAF) was highly consistent. The correlations among the POF and PAF scores the boys gave to the girls and AT scores are significant, but they are under medium level. As mentioned above, it can be said that the boys assessed the girls accurately, while they failed to do so with their assessments of their fellows.

Discussion

Alternative assessment approaches have been one of the appealing approaches in the last 20 years (Birenbaum and Dochy 1996). During this period many ideas were put forth about its basic function and studies started to focus especially on the function of assessment procedure in improving learning activities. In literature there are strong proofs for the usage of alternative assessment approaches as learning tools (Arter 1997; Dochy and McDowell 1997; Dochy, Segers, and Sluijsmans 1999). Nevertheless, the developments in the alternative assessment approaches are mostly at higher education level. However, the development of the skills required for the use of assessment as a learning tool must start at the primary school level. Research investigating the effects of gender on peer assessment done by the learners at primary school level can serve to obtain important findings.

The first aim of this study was to analyse the cooperative learning group members’ assessments of their peers in terms of their participation in group activities (POF) and their learning levels (PAF) according to gender differences. According to analysis results, we can see that while assessing the participation in group activities, the boys gave girls higher scores than the girls did. In the assessment of participation in group activities and learning levels, the scores given by the boys to the other boys are higher than the scores given by the girls to the boys, but there is no significant difference between the scores given by the boys to the girls and to other boys. That means that while the boys assessed the other boys and girls in the groups equally, the scores given by the girls to the other girls and boys were lower than the scores given by the boys. On the other hand, while there is no significant difference between the scores given by the girls to the other girls and the boys for their participation in group activities, their scores for the learning levels were found to be significant in favour of the girls. In other words, girls think that the other girls are more successful than the boys.

The second aim of the study was to analyse the accuracy levels both of the members’ participation in group activities (POF) and their learning levels (PAF) according to gender. Some of the studies on the accuracy of peer assessment (Oldfield and Macalpine 1995; Orsmond, Merry, and Reiling 1996) rest upon investigating the relationships between peer assessments and teacher assessments.
The same path was followed in this study, although the teacher assessment was limited to the achievement test (AT) which was prepared by the teacher and matches the attainments of the course. The results obtained and the interpretations made should be handled within the frames of these limitations.

The relationships show us that there are high and significant correlations between the students’ participation in group activities (POF) and their learning levels (PAF). The highest and the lowest of those correlations are among the scores given by the boys to the other members. While the boys assessed the girls’ participation in group activities and their learning levels so consistently, they assessed the other boys’ participation in group activities and their learning levels less consistently. If we compare these results with the correlations between the scores given by the girls to the other girls and boys, we can see that the boys were not accurate especially when they assessed their fellows. The negative correlation between the POF scores given by the boys to the other boys and their AT scores, and the insignificant low correlation between the PAF scores and AT scores in a way supports this result. In other words, it can be said that the boys were biased in their assessments of their fellows.

Moreover, when the relationships among the POF, PAF and AT scores were analysed, it can be said that the girls gave more accurate scores to the boys and to the other girls in groups. When all the members’ AT scores are taken as the criterion, we can see that the girls assessed the participation in group activities and learning levels more accurately than the boys did. The negative correlations among the POF and PAF scores given by the other members to the boys and their AT scores can have their source in the boys’ biased ratings of their fellows.

When the relationships between the POF and AT scores are analysed, we can see that the highest correlation is among the scores given by the girls to the boys; the lowest and negative correlation is among the scores given by the boys to the other boys. These results reveal that the girls made more accurate assessments compared to the boys.

When the relationships between the PAF and AT scores are analysed, the highest positive correlation can be found among the scores given by all the group members to the girls, the negative and insignificant low correlation can be found among the scores given by all the group members to the boys. These results show that the other members assessed the girls more accurately than they assessed the boys.

Looking back at the study results, it can be said that the boys’ scores were not differentiating enough and they need to be trained seriously in peer assessment. Dancer and Dancer (1992) stated that peers would tend to base their assessments on uniformity, race and friendship if they weren’t given extensive training on peer assessment. In one of their studies where they tried to find the students’ competency in making assessments, Oldfield and Macalpine (1995) compared the assessment results of peers and teachers, and found high correlations between the student scores and teacher scores. In our study, the same results can be true for girls.

Dochy, Segers, and Sluijsmans (1999) expressed that peer assessment can be valuable as a formative assessment method and hence as a part of the learning process. However, they warned that friendship-biased marking which can result in high scores, and biased grading which ends up in no difference in a group can be observed in peer assessment. In our study, the reasons for the boys’ inaccurate assessments of their fellows can be their friendly and biased assessments.
On the other hand, in a study done by Orsmond, Merry, and Reiling (1996), a correlation of .54 was found between the assessment results of peers and teachers. In our study, the correlation between the PAF scores of girls given by the other members and their AT scores was found $r(28) = .55, p < .01$. The accuracy levels of the assessments of the girls are similar to the findings of Orsmond, Merry, and Reiling (1996).

**Conclusion**

This study investigated the effects of gender on the peer assessments of those students at primary school level grade four. At the end of this study, it was found that considering their contribution to group work (POF) the boys scored the girls and other boys higher and closer than the girls scored the boys and other girls. The scores given by the girls to the boys and other girls are much lower. Taking their learning levels into account (PAF), both girls and boys scored their fellows higher than the opposite gender. On the other hand, when the relationships among the POF, PAF, and AT scores are analysed, it was found that the girls scored the boys and the other girls more accurately. While the boys were more accurate in scoring the girls, they were not that accurate in scoring their fellows. This may be due to the biased scoring of the boys when they assessed their fellows in group studies. The results showed that the girls were more successful in peer assessment than the boys. This also supports the result that the boys need more preparation (training) on peer assessment.

However, it has some limitations. For example, the effect of gender on peer assessment was studied by taking the teacher-made test as a criterion. Although a teacher-made test reflects the learning in a lesson, it might not be seen enough for teacher assessment. In future studies the teacher assessment part could be performed by the teacher assessing the groups’ skills in working together and their group work or by administering an instrument other than a test. On the other hand, as in the present study, peer assessments could be done by taking the peer contributions to group work or peer learning levels resulting from group activities, or those assessments could rest on presentations of group members and/or considering the products of the groups.

**References**


