European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available on-line at: www.oapub.org/edu

doi: 10.5281/zenodo.1249959 Volume 4 | Issue 7 | 2018

SELF, PEER AND TEACHER ASSESSMENTS: WHAT IS THE LEVEL OF RELATIONSHIP BETWEEN THEM?

Fatma Sasmaz Oren[i]
Manisa Celal Bayar University,
Faculty of Education,
Department of Science Education,
Turkey

Abstract:
In this research, self, peer and teacher assessment applications were carried out in a science education course included in the teacher education programme. The purpose of this study was to determine the level of relationship between self, peer and teacher assessment. Another aim of the study was to analyze whether there was reciprocity bias in these assessments. In the research, the pre-service science teachers (203 participants in total) assessed themselves and their peers in terms of presentation skills in higher education. The research is a quantitative research that employs descriptive and inferential statistical methods. Self-assessment and peer assessment scores showed moderately high correlations with teacher scores but both were higher than teacher scores. The analyses of reciprocity bias level demonstrated that the scores received or assigned by peers were almost unaffected by bias. On the basis of all these results, it could be argued that self-assessment and peer assessment applications can be used to evaluate presentation skills in teacher education programmes or different tasks in other areas in higher education.

Keywords: peer assessment; pre-service science teacher education; reciprocity bias; self-assessment

1. Introduction

The use of self and peer assessment practices increased in different areas of higher education in recent years. With the increased attention to learner-centred curricula, the topic of self-assessment and peer assessment has become of particular interest in assessment (Birjandi & Tamjid, 2012). For Patton (2012), student-centred pedagogies have been a growing acceptance of innovative forms of assessment, particularly peer and self-assessment in the field of higher education. Current curriculum reform in Turkey which organized on the basis of constructivist approach emphasizes different

[i] Correspondence: email fsasmaz@gmail.com

process-based assessment methods two of which are self and peer assessment. With social constructivism, different modes of peer assessment have become increasingly popular in higher education (Raes, Vanderhoven & Schellenswith, 2015). In self-assessment, an individual assesses himself/herself according to certain criteria. Peer assessment, on the other hand, involves students' evaluation of some work of their fellow students such as research papers, projects, reports, posters, and homework etc. (MoNE 2006, 26). Peer assessment is a method of motivating students, involving students discussing, marking and providing feedback on other students' work (Sitthiworachart & Joy 2008). At the same time peer assessment enables students to judge their own work by recognising the strengths and weaknesses of their peers' efforts (Crane & Winterbottom, 2008). Topping, Smith, Swanson and Elliot (2000, 150) maintains that peer assessment 'can be defined as an arrangement for peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others of similar status'. In a different way, self-assessment is usually referred to as the learner's self-grading of a piece of work (Mok et al. 2006). Reinholz's (2015) research which examined how peer assessment supports self assessment state that self assessment involves an individual comparing his or her performance to a desired goal to adjust and improve his or her practice. For Kearney (2013), self-assessment has the capacity to promote autonomous learning, which is a valuable aspect of sustainable learning. Self assessment and peer assessment are key factors in authentic assessment because they provide students with the opportunity to reflect objectively on their own accomplishment and learning (Sadeghi & Khonbi, 2015). Consequently self assessment and peer assessment share common features, as they both involve students judging the quality of student work and regarded as two assessment strategies with the potential to support learning (Poon, McNaught, Lam & Kwan, 2009) and these can be considered as learning tools, because they are part of a learning process where different skills are developed (Lindblom-Ylänne, Pıhlajamäkı and Kotkas, 2006).

1.1. Self and Peer Assessment Practices in Teacher Education Programmes

Recently, much work has been done on peer assessment at higher education level in different areas. Examples of such studies include those carried out in teacher education programmes (Li & Gao, 2015; Liu & Li, 2014; Patton & Marty-Snyder, 2014; Sluijsmans, Brand-Gruwell, & van Merriënboer, 2002; Sluijsmans et al. 2004; Tsai et al. 2001; Woolhouse 1999), computer science (Davies 2000; Lin, Liu, & Yuan, 2001; Purchase 2000; Venables & Summit 2003), business (Brooks & Ammons, 2003; Carvalho, 2013; Gatfield 1999; Hassan, Fox & Hannah, 2014; Lejk & Wyvill, 2001; Weaver & Esposto, 2012), engineering (Kim, 2014; Liow, 2008), medical education (Norcini 2003; Schönrock-Adema et al., 2007; Speyer et al., 2011), psychology (Smith, Cooper, & Lancaster, 2002; Topping et al., 2000; Walker, 2001) and history programmes (van den Berg, Admiraal, & Pilot, 2006). Also at higher education level, there are some studies dealing with self and peer assessment alike in various fields such as physical therapist programme (Miller 1999), biology programme (Orsmond, Merry & Reiling, 2000; Yucel

et al., 2014), mathematics programme (Kearney, Perkins & Kennedy-Clark, 2015), management education (Willcoxson, 2006), and information system education (Tu & Lu 2005).

Studies on self and peer assessment in teacher education programmes are relatively few when compared to some other fields. Although peer assessment has been frequently employed in areas such as engineering, business and health, its use in teacher education is new (Bagci-Kilic & Cakan, 2007). In a similar view, Sluijsmans et al. (2004) argue that training student teachers in assessment skills is an ill-defined area, and therefore, teachers are unfamiliar with ways to involve students in the assessment process through peer assessment. Additionally for Zundert et al. (2012), the great majority of published studies on peer assessment are case studies or studies using a pre-experimental design. For Kearney (2013) the attempts and trials to implement new innovative forms of assessment (authentic self and peer assessment) into a pre-service teacher education course support contentions that traditional assessment practices are not meeting the needs of new century learners. Thus, arguably, the self and peer assessment applications performed in the present research will be a significant educational experience for pre-service teachers.

Kim's (2009) experimental study on peer assessment is another such research in teacher education. It reveals how certain variables are influenced by the assessee's role in peer assessment. In this study, the author maintains that peer assessment offers students the opportunity to reflect upon the learning process, as well as the assessment process. Tsai et al. (2001) applied peer assessment with pre-service science teachers and preliminary peer assessment experience was carried out in a science education course. In this research, the researchers collectively used peer assessment, network and developed a system in designing science activities. Lopez-Real and Chan (1999) applied peer assessment in evaluating group projects and performed this application in a primary mathematics education course at a university.

Another relevant research in teacher education at higher education level is a self-assessment study carried out by Mok et al. (2006). The study employed the metacognitive approach, also called the know-want-learn method, in self-assessment. In this piece of research using five different teacher education modules, self-assessment was associated with learning and the students were asked to evaluate their own learning processes. The results demonstrate that self-assessment contributes to metacognition and know-want-learn method is a valuable self assessment tool for teachers in higher education. Another research in teacher education is research by Sluijsmans, Brand-Gruwell and van Merriënboer (2002) on peer assessment. In their study, the authors examined, using an experiment-control group design, the impact of peer assessment upon performance and perceptions. In this study, the students in the experimental group outperformed the control group students with regard to the quality of assessment skill and the end products of the course.

Al-Barakat and Al-Hassan's (2009) qualitative research with student teachers is another research in teacher education program. In the study, the researchers used semi-structured interviews to take student teachers' opinions concerning the benefits of using

peer assessment as a learning tool in the process of improving teacher education programs. The research's results demonstrated that one of the advantages of peer assessment was the development of student teachers' instructional competencies. The other was that peer assessment made an important contribution to enabling student teachers to form a set of criteria for sound judgment on classroom performance. Moreover, the research also showed that peer assessment has certain benefits such as improving student teachers' ability to assess their peers' performance, fostering their self-confidence in assessing classroom performances, and forming positive attitudes towards peer assessment as a learning tool. In another research, Nurov (2000) addressed self-assessment in foreign language achievement and examined the relationship between self-assessment results, students' achievement and teachers' estimations. The results revealed only weak correlations between the students' self-assessment, achievement test and teachers' assessment.

1.2. 'Reciprocity Bias' in Self and Peer Assessment Practices

For Magin (2001a), "*perhaps the most worrisome criticism of peer assessment methods, and the most difficult to counter, is that of lack of fairness due to peer raters being influenced by relationships with others …"* (p. 54). In the small group behaviour and interactionist theory in the literature, validity and fairness of peer ratings are described as 'relational effects' and termed as 'reciprocation'. In experimental studies in this field, reciprocation implies 'the tendency for two people who are involved in rating each other to be influenced in their rating behaviour by social interactions between the two' (Magin, 2001a). Lawrence (2001) argues in a study that peer bias might be in different ways in group project evaluation and exemplifies it using various scenarios. As asserted in this research, bias may result from over-generous behaviour of a peer-assessor group member or his/her effort to inflate his/her own mark by assigning inappropriately low marks to all other group members (creative accounting). Besides, another type of bias is present when most or all group members assign very low marks to one assessed student in the group. This may stem from the inadequate contribution of that particular group member in the project or his/her being a weak student. As a solution to such cases, Lawrence proposed the concepts of bias factor and normalisation factor. Thus, bias factor is obtained by dividing the rating given to others by average effort rating, whereas normalisation factor is calculated by dividing one (1) by bias factor. As argued by the researcher, the normalisation process is designed to iron out human biases during the peer assessment in group projects. Ballantyne et al. (2002) assert that codes can be used in peer assessment for the sake of anonymity, which might help eliminating student concerns about bias and unfair rating. For Tseng and Tsai (2007), the on-line peer assessment system could ensure a higher degree of anonymity of peer assessment. However, in their comparative research on peer assessment of posters, Smith et al. (2002) said that although every poster is anonymous on display, students have argued that close friends may identify posters by their topic.

In this research, self, peer and teacher assessment were used to evaluate the pre-service science teachers' presentation skills in an elementary science education course.

The curricular reform in Turkey requires the use of self and peer assessment approaches at primary and elementary education level and it is the teachers who are supposed to use these two assessment tools in classrooms. Teachers need to be informed and perform exemplary applications about these methods to implement them effectively, efficiently, and properly in their classes. Wen and Tsai (2008) argue that teachers should become familiar with new teaching and assessment strategies before they actually use the methods with their students. In this respect, the research examined the relationship between self, peer and teacher assessment. It also examined whether peer ratings involved bias. The following research questions were investigated in this research:

1. What is the relationship between self, peer and teacher ratings used to assess presentation skills in elementary science education instruction?
2. Are there any statistically significant differences between the self, peer and teacher ratings used to assess presentation skills in elementary science instruction?
3. To what extent do the peer ratings involve reciprocity bias?

2. Material and Methods

The research is a quantitative research that employs descriptive and inferential statistical methods. It was conducted in a medium-scale university in the Aegean region in Turkey. The research lasted for a semester (10 weeks) and were carried out as a part of the course Science Education II, a compulsory course offered in the second semester of the third year in the primary teaching programs of education faculties in Turkey.

2.1. Participants

The research's participants include a total of 203 students from five different classes taught by the researcher. Table 1 presents information about the student groups in the sample. Most of the groups consisted of 4 students. Also, there were several 3 or 5-student groups. Students' ages ranged from 19 to 24 years and 95 (46.8%) of the students were *female*, 108 (53.2%) male. Participants were recruited on a volunteer basis.

Table 1: Frequencies and percentages regarding the classrooms of the students in the sample and the number of group members

Classrooms	Number of students	Group sizes		
		3 member	4 member	5 member
3 A1	41 (20.2%)	1 (10.0%)	7 (70.0%)	2 (20.0%)
3 B2	43 (21.2%)	0 (0%)	7 (70.0%)	3 (30.0%)
3 B1	42 (20.7%)	0 (0%)	8 (80.0%)	2 (20.0%)
3 C2	40 (19.7%)	1 (10.0%)	8 (80.0%)	1 (10.0%)
3 D2	37 (18.2%)	3 (30.0%)	7 (70.0%)	0 (0%)
Total	203 (100.0%)	5 (10.0%)	37 (74.0%)	8 (16.0%)

2.2. Procedure

Before starting the applications, the pre-service science teachers were informed in detail about the process of self and peer assessment by the teacher (the author was the course

teacher). The students were handed out the self and peer assessment forms and provided with explanations about the points they had difficulty in understanding. This informing process lasted for 2 course sessions. During the following course sessions, two volunteering pre-service science teachers identified during the earlier week made a presentation about a topic for which they believed to be informed well and prepared for one week. Following the application, their presentation skills were evaluated through peer, self and teacher assessment. This application was performed in each of the five classrooms before the actual applications started. Thus, a small-scale pilot study was conducted for the applications to be continued for a semester.

The course Science Education II is taught for 4 course sessions a week. The teacher uses two of these four sessions to present theoretical knowledge. During the remaining two course sessions, pre-service teachers make presentations in accordance with the teaching method/approach/model/technique taught by the teacher in the previous week. This could be exemplified in two ways: the teacher teaches the students about problem-based learning approach during the first two sessions of the course Science Education II. During the first two sessions of the following week, the teacher teaches another science subject, the inquiry-based learning approach, while during the next two sessions, the pre-service science teachers make a presentation about the problem-based learning approach taught the previous week. Since student groups are determined at the beginning of the semester and students are informed at the beginning of the semester by the teacher about the topics to be taught each week, each group knows in what week and on which topic they are going to make a presentation. This kind of application allows pre-service teachers to learn about both the theoretical background and micro applications of the science methods and approaches in question. Throughout the semester, besides the two abovementioned examples about science (inquiry- and problem-based learning approaches), the students were also taught about many different subjects such as concept maps, alternative assessment approaches, learning cycle, analogies, inquiry-based learning, misconceptions, and cooperative learning. The science teaching methods (e.g., inquiry-based learning) and science topics (e.g., heat and temperature) chosen by the participants in this research were based on familiarity and general interest. Implementing these science teaching methods provides students with opportunities to apply their knowledge. It could also be argued that such application process contributes to the creation of a cooperative and student-centered learning environment.

2.3. Peer and Self Assessment Form

The same assessment form was used for assessment by the students (peer and self assessment) and the teacher (teacher assessment). Evaluation was based on 25 items in the assessment form. The items in the instruments were closely related to the “teaching abilities” necessary for professional competence and teaching career.

The assessment form is a five-point Likert-type instrument. Assessors can state their opinions for each item in the form by using the following expressions: “very good”, “good”, “average”, “poor” and “very poor”. Total scores were obtained by

assigning the statements with ratings of 5, 4, 3, 2 and 1. Thus, the minimum score to be obtained by a student in each form (self or peer assessment) is 25, while the maximum rating is 125. In the research, each student was assessed by all of the other students in their class and ratings were averaged to produce an overall peer rating. At the same time, each student in a group making a presentation was assessed independently for all statements. For instance, on the statement "He/She could communicate with the students", one of the assessment criteria under the "communication" subheading, a student in a group may receive 5 points, while another group member might be assigned 3 points.

The form mainly consists of two sub-criteria involving different aspects of assessment. The first is "science content knowledge and science teaching knowledge", and the second is "teaching-learning process". The second criterion consists of three different categories, which are "teaching process", "class management" and "communication". Besides, the form also contains entries such as name, grade, classroom, gender, the unit and method/approach covered in the presentation. This form was published by the Higher Education Council (1998) and adapted for group assessment by Bagci-Kilic and Cakan (2007). The assessment form used in this research is the adapted version of Bagci-Kilic and Cakan (2007). Yet, the form was re-organized in two ways for the research as the authors used it only for peer assessment, while the present research employed it both for peer and self assessment. Below are two examples of the items in self and peer assessment forms:

> *"He/she used appropriate language and visual aids (figures, charts, graphs, etc.)." (Peer assessment)*
> *"I used appropriate language and visual aids (figures, charts, graphs, etc.)." (Self assessment)*
> *"He/she determined the misconceptions of the students on the science subject." (Peer assessment)*
> *"I determined the misconceptions of the students on the science subject." (Self assessment)*

2.4. Analyses

To calculate the reliability of peer assessment scores, a calculation method based on ANOVA (analysis of variance) was used, a method which was developed by Ebel (1951, p.410) and also used by Magin and Helmore (2001, 293). In this calculation method, if F is the result of ANOVA, reliability of the scores obtained from N number of raters in the class is computed by the following formula

$$r_{nn} = (F - 1) / F \text{ , while}$$

reliability of the scores obtained from a single rater is determined by formula

$$r_{11} = (F - 1) / (F + N - 1).$$

Since applications were carried out in five different classrooms in this research, the reliability coefficients obtained were subjected to Fisher's transformation (Hays, 1994, p. 649) to obtain their means and standard deviations and 95% confidence intervals were calculated and Table 2 shows the reliability of peer assessment scores.

Table 2: Reliability data about the peer ratings

Classrooms	r_{11}	r_{nn}
3 A1	0.25	0.93
3 B2	0.40	0.97
3 B1	0.27	0.94
3 C2	0.35	0.96
3 D2	0.17	0.88
Mean	0.29	0.94
95% Confidence interval	0.11 – 0.45	0.86 – 0.98

As a result of ANOVA-based calculation method performed to determine the reliability of peer assessment, the reliability coefficients obtained for each of the five classes (r_{11}) were in the range of 0.11 – 0.45 and had a mean of 0.29 following Fisher's transformation. When the ratings obtained from all the students in the classes were used, on the other hand, these values (r_{nn}) were in the range of 0.86 – 0.98 and had a mean of 0.94. In other words, reliability value increases with a higher number of raters. It is an anticipated result that scores obtained from more raters have higher reliability.

Correlation coefficients were computed to address the first research problem, while ANOVA was employed to analyze the data regarding the second research problem. To examine the last research problem, correlation coefficients were calculated among the scores assigned by the students to each other's presentations, as was done by Magin (2001a). These coefficients were also subjected to Fisher's transformation and their means and 95% confidence intervals have been reported.

3. Results

The relationships between peer, self and teacher assessments were revealed by correlation coefficients. Table 3 shows the correlation coefficients between the ratings. All coefficients given in the table are statistically significant.

Table 3: Correlation coefficients between peer, self and teacher ratings

	Peer	**Self**	**Teacher**
Peer		.52	.69
Self			.61
Teacher			

Another research question is whether there is a statistically significant difference between self and peer ratings, and teacher ratings. First of all, the descriptive statistics regarding the obtained ratings are presented in the Table 4.

Table 4: The means and standard deviations of the self and peer ratings and teacher ratings

	N	M	SD
Peer	199	106.49	9.27
Self	200	105.53	12.36
Teacher	200	94.84	14.44
Total	599	102.28	13.29

As revealed by the descriptive analysis in the table, the peer and self ratings are similar, while they are higher in comparison with the teacher ratings. ANOVA was employed to examine whether these rating differences are statistically significant. As obtained by the ANOVA results, the findings indicate significant differences (F=55.99, $p < 0.01$). Tukey's HSD test and LSD test were used to determine between which groups the rating differences were significant. These two tests yielded the same results. Table 5 presents the results of Tukey's HSD test.

Table 5: The results of the post-hoc tests performed following the ANOVA comparing the peer and self ratings and the scores assigned by the teacher

Groups	Mean difference	Std. error	p	Cohen's d
Peer – Self	0.96	1.22	.71	0.08
Peer – Teacher	11.65	1.22	.00	0.95
Self – Teacher	10.69	1.22	.00	0.88

The results of Tukey's HSD test demonstrate that the differences between the peer and self ratings are insignificant, while there were statistically significant differences between teacher ratings and self and peer ratings. Cohen's d was computed to examine the practical significance of these rating differences. It was observed that the peer ratings were approximately 0.95 standard deviation higher than the teacher ratings. Similarly, the self ratings were found to be about 0.88 standard deviation higher than the teacher ratings. In the literature, differences of this size are considered as 'large effect' size differences (Cohen, 1992). As a consequence, with regard to this research problem, the correlations between the self and peer ratings and the teacher ratings are above medium (moderate), while the rating differences are both statistically and practically significant. This suggests that the students differentiated between themselves and their peers in terms of their presentation skills in science course, but they found themselves and their peers more successful when compared to the teacher's assessments.

For the last research problem, the correlations between the students' ratings were examined to reveal if there is any reciprocity bias, as was done by Magin (2001a). The correlations obtained from the five classes ranged between -0.12 and 0.14. As a result of Fisher's transformation, the mean correlation was found to be 0.10 and 95% confidence interval was obtained as [-0.11, 0.24], which indicates that the ratings were almost uninfluenced by bias, or the peers did not favour each other and table 6 shows the correlation coefficients between the ratings.

Table 6: The correlation coefficients between the students' ratings

Classrooms	Valid pairs*	r
3 A1	484	0.14
3 B2	527	-0.12
3 B1	482	0.14
3 C2	482	0.07
3 D2	372	0.10
Mean		0.07
95% Confidence interval		-0.11 – 0.24

* The number of valid pairs was reduced due to absenteeism (as absent students failed to assess their peers who made presentations during the course sessions they missed).

4. Discussion

The correlation coefficient analysis was used to determine the relationship between self, peer and teacher scores. In this research founded that a moderately high correlation (0.61, 0.69) between the teacher scores and self and peer scores. The literature contains studies that demonstrate similar or different results on the issue. Pope's (2005) study reported a correlation value of 0.59 between self assessment and tutor assessment and a correlation value of 0.60 between peer assessment and tutor assessment. Obviously, these values are very close to those obtained in the present research (in particular, the correlation between self assessment and teacher assessment). In Segers and Dochy's (2001) research, the peer and tutor scores are significantly interrelated (approximately 0.7). In their study, the self scores are, to a minor extent, related to peer and tutor scores. In Parti's (2002) research, moderately low (0.50, 0.46) correlations were found between teacher assessment and self assessment both in the experiment and control groups. As for peer assessment, the correlation was low (0.49) in the control group, while a high level correlation (0.85) was reported for the experiment group.

In this research founded that mean peer and self assessment scores as significantly higher than mean teacher assessment scores. ANOVA analysis was used to analyze peer, self and teacher scores in detail. This analysis revealed no significant difference between self and peer assessment, but statistically significant differences between teacher ratings and self and peer scores. In other words, the pre-service science teachers assigned to themselves and their peers higher scores than the teacher did. Similarly, in a study by Langan et al. (2005) which employed peer assessment to evaluate oral presentations, the students' peer scores were considerably higher than the scores they obtained from their tutors. Miller (1999) conducted a study on self and peer assessment with the students of a physical therapist programme, in which he applied the problem-based learning approach to six different problems and obtained similar results. His research results indicate that there was no significant difference between the peer assessment scores and the self assessment scores for each case problem. In Matsuno's (2009) study, in parallel to this research, the averaged peer scores were higher than the averaged teachers' scores; yet, as different from this study, the averaged self assessment scores and averaged teachers' ratings were the same in the research. In Suñol et al. (2015)'s study show that in general, students (in both peer and self

assessment) award themselves higher marks than those awarded by the teacher. Similar to this research, in Parti's (2002) research, the lowest scores also came from teacher assessment, while the highest averaged scores were obtained in peer assessment.

A large number of studies are found in the literature examining the relationship of teacher/ instructor/ tutor/ lecturer/ expert assessment scores with peer scores in different fields; yet, the studies are scarce on the relationship between self scores and instructor/tutor/lecturer/expert scores. Therefore, comparisons on the subject are usually based on the peer assessment studies in the literature. Accordingly, my finding about the correlation coefficient between peer assessment and teacher's assessment (0.69) is lower than that of MacAlpine's research. In MacAlpine's (1999) study, in which student presentations were rated in the four main sections of knowledge, body-language, voice and overall impression, peer assessment marks were compared to lecturers' marks and the research reported a correlation coefficient of 0.80. In Bagci-Kilic and Cakan's study on peer assessment, the correlation between peer and instructor in the first application was 0.69, the same value obtained in the present research, while the second application yielded a lower (0.37) value (Bagci-Kilic & Cakan 2007). Tseng and Tsai's (2007) research on on-line peer assessment found correlations between the scores assigned by the students and the experts/teachers that ranged from 0.49 to 0.79.

Another important point in peer assessment studies concerns whether raters behave sufficiently objective or biased. There might be many variables that may influence, whether unconsciously or deliberately, the bias of the students as raters. Magin (2001a) argues that the bias is seen to arise as a result of friendships and social interactions accompanying group task activities. Langan et al. (2008) identified the main biases in peer assessment as gender (male markers favoured male speakers), institutional affinity (slight bias towards those from their own university) and peer-group status (anecdotal evidence of 'popular' individuals receiving higher peer marks). In this research, the correlation levels between the scores assigned by students were examined in order to reveal reciprocity bias level. These levels ranged from -0.12 to +0.14 in five classes. Fisher's transformation yielded a mean correlation of 0.10 and a 95% confidence interval of [-0.11, +0.24]. This result indicates that the scores assigned/received by peers were almost unaffected by reciprocity; in other words, the friends did not favour each other. This finding of the research is compatible with that of Magin's (2001) study. In Magin's study carried out with 16 groups, the correlation levels ranged from -0.02 to + 0.31, with a mean of 0.110. As a result of Fisher's transformation, the mean correlation obtained was +0.11 and the 95% confidence interval was [+0.07, +0.15]. Reciprocity effects (rater-ratee relational effects), therefore, account for only 1% of the variance in scores. In this research, on the other hand, this effect was found as 0.005% (explained variance 0.10). Weaver and Esposto (2012)'s study which examined only peer assessment found that within groups, students who received higher marks from their peers generally awarded marks to their peers across a wider range, whereas students who received lower average grades often awarded the same mark to all team members. In another research when students are asked to self-assess their own work and compare their marks with the marks given by their peers,

students moderately underestimate the work of their colleagues and overestimate their own (El-Mowafy, 2014). Matsuno's (2009) study which examined self and peer assessment together found that teachers were neither lenient nor severe with the self- and peer-assessors but students generally evaluated their peers leniently and evaluated themselves severely. Author's research conducted in higher education on writing abilities also examined the issue of bias and employed a bias analysis of rater-writer interactions. The results of this analysis showed that the peer-assessors tended to award more lenient scores to lower-level writers and harsher scores to more proficient writers (Matsuno, 2009).

5. Conclusions and Implications

This research on a quantitative research of self, peer and teacher assessments of presentation skills in science teaching by Turkish university students. The research has certain important findings. First, the peer and self scores were in accordance with the teacher scores, which was indicated by significant correlations between self, peer and teacher scores in this research, but peer and self scores were significantly higher than teacher scores. Secondly, there are moderately high correlations between self, peer and teacher assessment. Therefore, it is suggested that in teacher education programmes, self and peer assessment approaches can be used in evaluating science teaching presentation skills.

Thirdly, analyses of reciprocity bias level revealed that the scores assigned or received by peers were almost unaffected by bias, or the friends did not favour each other. Then, this research suggests that in some contexts, peer and self assessments can serve useful in teacher education programmes. One of the most problematic aspects of self and peer assessment applications is the issue of bias. Therefore, beside the analysis method employed in this research to reveal bias, the use of alternative analysis methods might also be needed to support the results. Examples might include the normalization process proposed by Lawrence (2001), normalization process of Individual Weighting Factors proposed by Spatar et al. (2015) or the Multifaceted Rasch measurement method employed by Matsuno (2009). In addition, further studies could administer students a sociometry test to investigate the effect of certain variables (e.g., popularity in the classroom) on bias and objectivity. As for students' presentation skills, studies could examine the relationship between intelligence domains and these skills (e.g., do students with high verbal/linguistic intelligence or social/interpersonal intelligence also have improved presentation skills?) or the relationship between intelligence domains and self and peer assessment (i.e., does a student with high intrapersonal intelligence assign more objective self assessment scores?).

Another recommendation for further research is, as suggested by some studies (i.e., Kwan & Leung 1996), that participants can be active in identifying the criteria to be used in self and peer assessment. Also, the teacher's observations and interviews in the research revealed that the number of evaluation criteria is high for the students (25 items), which causes difficulties with time management. Even Jones and Alcock (2014)'s

peer assessment research suggest that assessment in the absence of criteria can be a useful approach. Thus, further studies could be recommended to use a smaller number of assessment criteria or another form of (like rubric) assessment criteria.

Many studies on self and peer assessment have demonstrated the importance of feedback. In such studies, it is important to find out the answer to the question "how should feedback organized?". Furthermore, using printed forms in application for each student making a presentation every week constitutes a limitation both the need for photocopies and waste of time. In addition, in studies with large samples (like this research), it is much time-consuming to transfer to electronic environment the data obtained from self and peer assessment forms called 'paper and pencil'. The use of web-based or online self and peer assessment process could be recommended as a solution to these drawbacks. By reviewing such studies in the literature (i.e., Freeman & McKenzie 2002; Li & Gao, 2015; Lin et al. 2001; Liu & Li, 2014; Tsai et al. 2001; Tseng & Tsai 2007; Willey & Gardner, 2010), a similar system could be employed to assess presentation skills in science education.

Acknowledgements
The author would like to thank Zafer Cepni for his comments and help in data analysis of this research.

About the Author
Fatma Sasmaz Oren is currently an associate professor at Manisa Celal Bayar University, Turkey. She graduated Department of Science Education at 1998, Master of Education 2001 at Manisa Celal Bayar University, Institute of Science. She began her professional life as a science and technology teacher, and then completed her doctoral research on 'the effect of the learning cycle approach on logical thinking skill, academic achievement and attitude' at Gazi University, Turkey in 2005. Her main field of study involves alternative assessment approaches in science education, constructivism, concept cartoons, concept mappings and learning cycle approach. There are a number of national and international articles published on the topics mentioned by the author. She works as an associate professor in the Department of Science Education of the Education Faculty at Manisa Celal Bayar University, Turkey.

References

1. Al-Barakat, A., & Al-Hassan, O. (2009). Peer assessment as a learning tool for enhancing student teachers' preparation. Asia-Pacific Journal of Teacher Education, 37(4), 399–413.
2. Bağcı-Kılıç, G., & Çakan, M. (2007). Peer assessment of elementary science teaching skills. Journal of Science Teacher Education, 18, 91–107.

3. Birjandi, P., & Tamjid N.H. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. Assessment & Evaluation in Higher Education, 37 (5), 513-533.
4. Brooks, C.M., & Ammons, J.L. (2003). Free riding in group project and the effect of timing, frequency, and specificity of criteria in peer assessments. Journal of Education for Business, 78(5), 268–272.
5. Carvalho, A. (2013). Students' perceptions of fairness in peer assessment: evidence from a problem-based learning course. Teaching in Higher Education, 18(5), 491-505.
6. Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1),155–159. http://psycnet.apa.org/?fa=main.doiLanding&doi=10.1037/0033-2909.112.1.155 (Retrieved from the website on September 9, 2009)
7. Crane, L., & Winterbottom, M. (2008). Plants and photosynthesis: peer assessment to help students learn. Journal of Biological Education, 42(4), 150-156.
8. Davies, P. (2000). Computerized peer assessment. Innovations in Education and Training International, 37(4), 346–355.
9. Ebel, R. (1951). Estimation of the reliability of ratings. Psychometrika, 16, 407–424.
10. El-Mowafy, A. (2014). Using peer assessment of fieldwork to enhance students' practical training. Assessment & Evaluation in Higher Education, 39(2), 223–241.
11. Freeman, M., & McKenzie, J. (2002). SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects. British Journal of Educational Technology, 33(5), 551–569.
12. Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. Assessment & Evaluation in Higher Education, 24(4), 365–377.
13. Hassan, O.A.G., Fox, A., & Hannah, G. (2014). Self- and peer- assessment: evidence from the accounting and finance discipline. Accounting Education: An International Journal, 23(3), 225-243.
14. Hays, W.L. (1994). Statistics. (5th Edition). Florida: Harcourt Brace.
15. Higher Education Council (1998). Faculty-School Collaboration. Ankara, Turkey: Author. http://www.yok.gov.tr/egitim/ogretmen/aday_ogretmen/bolum5.html (Retrieved from the website on February 7, 2008)
16. Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. Studies in Higher Education, 39(10), 1774–1787.
17. Kearney, S. (2013). Improving engagement: the use of 'authentic self- and peer - assessment for learning' to enhance the student learning experience. Assessment & Evaluation in Higher Education, 38(7), 875-891.
18. Kearney, S., Perkins, T., & Kennedy-Clark, S. (2015). Using self- and peer-assessments for summative purposes: analysing the relative validity of the AASL (authentic assessment for sustainable learning) model. Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2015.1039484

19. Kim, H.S. (2014). Uncertainty analysis for peer assessment: oral presentation skills for final year project. European Journal of Engineering Education, 39(1), 68-83.
20. Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. Assessment & Evaluation in Higher Education, 34(1), 105–114.
21. Kwan, K., & Leung, R. (1996). Tutor versus peer group assessment of student performance in a stimulation training exercise. Assessment & Evaluation in Higher Education, 21(3), 205–214.
22. Langan, A.M., Wheater, C.P., Shaw, E.M., Haines, B.J., Cullen, W.R., et al. (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. Assessment & Evaluation in Higher Education, 30(1), 21–34.
23. Langan, A.M., Shuker, D.M., Cullen, W.R., Penney, D., Preziosi, R.F., & Wheater, C.P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. Assessment & Evaluation in Higher Education, 33(2), 179–190.
24. Lawrence, K.Y.L. (2001). Some refinements on peer assessment of group projects. Assessment & Evaluation in Higher Education, 26(1), 5–18.
25. Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches. Assessment & Evaluation in Higher Education, 26(1), 61–72.
26. Li, L., & Gao, F. (2015). The effect of peer assessment on project performance of students at different learning levels. Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2015.1048185
27. Lin, S.S.J., Liu, E.Z.F., & Yuan, S.M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. Journal of Computer Assisted Learning, 17, 420–432.
28. Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. Active Learning in Higher Education, 7(1), 51–62.
29. Liu, X., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. Assessment & Evaluation in Higher Education, 39(3), 275-292.
30. Liow, J.L. (2008). Peer assessment in thesis oral presentation. European Journal of Engineering Education, 33, 525-537.
31. Lopez-Real, F., & Chan, R.Y.P. (1999). Peer assessment of a group project in primary mathematics education course. Assessment & Evaluation in Higher Education, 24(1), 67–79.
32. MacAlpine, J.M.K. (1999). Improving and encouraging peer assessment of student presentations. Assessment & Evaluation in Higher Education, 24(1), 15–25.
33. Magin, D. (2001a). Reciprocity as a source of bias in multiple peer assessment of group work. Studies in Higher Education, 26(1), 53–63.

34. Magin, D. (2001b). A novel technique for comparing the reliability of multiple peer assessment with that of single teacher assessments of group process work. Assessment & Evaluation in Higher Education, 26(2), 139–152.
35. Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? Studies in Higher Education, 26(3), 287–298.
36. Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing, 26(1), 75–100.
37. Miller, P.J. (1999). The agreement of peer assessment and self-assessment of learning processes in problem-based learning. Journal of Physical Therapy Education, 13(2), 26–30.
38. MoNE (Milli Eğitim Baknlığı) (2006). İlköğretim Fen ve Teknoloji Dersi (6, 7 ve 8. sınıflar) Öğretim Programı [Elementary Science and Technology Curriculum (for 6th, 7th, and 8th grades)], Ankara: Author.
39. Mok, M.M.C., Lung, C.L.C., Cheng, D.P.W., Cheung, R.H.P., & Ng, M.L. (2006). Self-assessment in higher education: experience in using a metacognitive approach in five case studies. Assessment & Evaluation in Higher Education, 31(4), 415–433.
40. Norcini, J.J. (2003). Peer assessment of competence. Medical Education, 37, 539–543.
41. Nurov, A. (2000). Self-assessment of foreign language achievement: the relationship between students' self-assessment, teachers' estimates and achievement test. Master's diss., University of Bilkent. Retrieved from http://teav.ankara.edu.tr/browse/1096/
42. Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. Assessment & Evaluation in Higher Education, 25(1), 23–38.
43. Patri, M. (2002). The influence of peer feedback on self- and peer- assessment of oral skills. Language Testing, 19(2), 109–131.
44. Patton, C. (2012). 'Some kind of weird, evil experiment': student perceptions of peer assessment. Assessment & Evaluation in Higher Education, 37(6), 719-731.
45. Patton, B.J., & Marty-Snyder, M. (2014). Practical applications for using peer assessment in physical education teacher education field experiences. Strategies: A Journal for Physical and Sport Educators, 27(6), 25-31.
46. Poon, W.Y., McNaught, C., Lam, P., & Kwan, H.S. (2009). Improving assessment methods in university science education with negotiated self and peer assessment. Assessment in Education: Principles, Policy & Practice, 16(3), 331-346.
47. Pope, N.K.L. (2005). The impact of stress in self-and peer assessment. Assessment & Evaluation in Higher Education, 30(1), 51–63.
48. Purchase, H.C. (2000). Learning about interface design through peer assessment. Assessment & Evaluation in Higher Education, 25(4), 341–352.

49. Raes, A., Vanderhoven, E., & Schellenswith, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. Studies in Higher Education, 40(1), 178–193.
50. Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. Assessment & Evaluation in Higher Education, DOI: 10.1080/ 02602938.2015.1008982
51. Sadeghi, K., & Khonbi, Z.A. (2015). Iranian university students' experiences of and attitudes towards alternatives in assessment. Assessment & Evaluation in Higher Education, 40(5), 641-665.
52. Schönrock-Adema, J., Heijne-Penninga, M., van Duijn, M.A.J., Geertsma, J., & Cohen-Schotanus, J. (2007). Assessment of professional behaviour in undergraduate medical education: peer assessment enhances performance. Medical Education, 41, 836–842.
53. Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: the value-added of the students' perspective. Studies in Higher Education, 26(3), 327–343.
54. Sitthiworachart, J., & Joy, M. (2008). Computer support of effective peer assessment in an undergraduate programming class. Journal of Computer Assisted Learning, 24, 217–231.
55. Sluijsmans, D.M.A., Brand-Gruwell, S., & van Merriënboer, J.J.G. (2002). Peer assessment training in teacher education: effects on performance and perceptions. Assessment & Evaluation in Higher Education, 27(5), 443–454.
56. Sluijsmans, D.M.A., Brand-Gruwell, S., van Merriënboer J.J.G., & Martens, R.L. (2004). Training teachers in peer-assessment skills: effective on performance and perceptions. Innovations in Education and Teaching International, 41(1), 59–78.
57. Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: a case for student and staff development. Innovations in Education and Teaching International, 39 (1), 71–81.
58. Spatar, C., Penna, N., Mills, H., Kutija, V., & Cooke, M. (2015). A robust approach for mapping group marks to individual marks using peer assessment. Assessment & Evaluation in Higher Education, 40(3), 371-389.
59. Speyer, R., Pilz, W., Kruis, J.V.D., & Brunings, J.W. (2011). Reliability and validity of student peer assessment in medical education: a systematic review. Medical Teacher, 33(11), e572–e585.
60. Suñol, J.J., Arbat, G., Pujol, J., Feliu, L., Fraguell, R.M., & Planas-Lladó, A. (2015). Peer and self-assessment applied to oral presentations from a multidisciplinary perspective. Assessment & Evaluation in Higher Education, DOI:10.1080/02602938.2015.1037720
61. Topping, K.J., Smith, E.F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. Assessment & Evaluation in Higher Education, 25(2), 149–169.

62. Tsai, C.C., Liu, E.Z.F., Lin, S.S.J., & Yuan, S.M. (2001). A networked peer assessment system based on a Vee heuristic. Innovations in Education and Teaching International, 38(3), 220–230.
63. Tseng, S.C., & Tsai, C.C. (2007). On-line peer assessment and the role of the peer feedback: a study of high school computer course. Computers & Education, 49, 1161–1174.
64. Tu, Y., & Lu, M. (2005). Peer-and-self assessment to reveal the ranking of each individual's contribution to a group project. Journal of Information System Education, 16(2), 197–205.
65. van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: evaluating seven course designs. Assessment & Evaluation in Higher Education, 31(1), 19–36.
66. van Zundert, M.J., Sluijsmans, D.M.A., Könings, K.D. & van Merriënboer, J.J.G. (2012). The differential effects of task complexity on domain-specific and peer assessment skills. Educational Psychology, 32(1), 127-145.
67. Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. Innovations in Education and Teaching International, 40(3), 281–290.
68. Walker, A. (2001). British psychology students' perceptions of group-work and peer assessment. Psychology Learning and Teaching, 1(1), 28-36.
69. Weaver, D., & Esposto, A. (2012). Peer assessment as a method of improving student engagement. Assessment & Evaluation in Higher Education, 37(7), 805–816.
70. Wen, M.L., & Tsai, C-C. (2008). Online peer assessment in an inservice science and mathematics teacher education course. Teaching in Higher Education, 13(1), 55–67.
71. Willcoxson, L.E. (2006). "It's not fair!": assessing the dynamics and resourcing of teamwork. Journal of Management Education, 30(6), 798–808.
72. Willey, K., & Gardner, A. (2010). Investigating the capacity of self and peer assessment activities to engage students and promote learning. European Journal of Engineering Education, 35(4), 429-443.
73. Woolhouse, M. (1999). Peer assessment: the participants' perception of two activities on a further education teacher education course. Journal of Further and Higher Education, 23(2), 211–219.
74. Yucel, R., Bird, F.L., Young, J., & Blanksby, T. (2014). The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. Assessment & Evaluation in Higher Education, 39(8), 971–986.