

The Effect of Using Peer Assessment Training on Writing Performance among Arab EFL High School Students in Malaysia

Abdallah Mohammad Salem Almahasneh
Centre for Fundamental and Liberal Education
Universiti Malaysia Terengganu, Malaysia

Samsiah Abdul- Hamid
Centre for Fundamental and Liberal Education
Universiti Malaysia Terengganu, Malaysia

Abstract

Peer assessment training has appeared as potential new tools for enhancing Arab English as foreign language (EFL) high school students on writing performance. The purpose of this research was to investigate the effect of using peer assessment training on writing performance among Arab EFL high school students. One hundred and twenty students aged fifteen and sixteen years old participated in this study with an equal number of male and female students. The students were from two Arab high schools in Malaysia. This study employed a quasi-experimental research design. Data was collected using Analytical Marking Scale (Alderson, et al., 1995) to assess student's writing performance in the pretest and posttest. The result of this study shows that there is a significant difference in the writing performances between the experimental and control groups at $P < 0.001$. The findings suggests that students who have been exposed to peer assessment training write a better draft compared with those students who were only given conventional training in writing an essay, and without the benefit of peer feedback.

Keywords: Arab EFL, peer assessment, writing performance

Cite as: Almahasneh, A.M.S., & Abdul- Hamid, S. (2019). The Effect of Using Peer Assessment Training on Writing Performance among Arab EFL High School Students in Malaysia. *Arab World English Journal*, 10 (1)105-115.

DOI: <https://dx.doi.org/10.24093/awej/vol10no1.10>

1. Introduction

Peer assessment (PA) is a communicative strategy that are used by teachers in language classrooms to actively engage students in the learning process by having them assess their peers' work based on the teacher's benchmark. PA is a form of formative assessment where feedback on the learning is collected from the students themselves. Later, after training students on peer assessment, this students-generated feedback is exchanged with their peers and is used to improve their performance, be it in either written or spoken activities. As a form of assessment for learning, PA has been regarded by Topping (2009) as "an arrangement for the learner to consider and specify the level, value or quality of a product or performance of other equal-status learners" (p. 20). Based on this scenario, teachers guide their students to take an active role towards their learning and their peers' production language. This practice also enables students to assume more responsibilities in making decisions that will ultimately influence the peers' work and performance.

Previous studies conducted on the effect of peer assessment (PA) training at the university level have shown to impact positively on students' writing performance and feedback exchange in most aspects of academic activities (Al-Barakat & Al-Hassan, 2009; van den Berg, Admiraal, & Pilot, 2006; Weaver & Esposito, 2011). It should be noted that most of these studies were focused on undergraduate and postgraduate students. However, there is still a difficult use of peer assessment among high school students. that looks into the effectiveness of PA training in improving writing performance (Yurdabakan, 2010). This investigation explores the effect of using PA training on the writing performance of Arab EFL high school students.

2. Literature review

Peer assessment is a form of formative assessment wherein students are given feedback on their learning process by other students in the same level instead of the teacher (Topping, 2009). It is a continuous process which has been reported to not only reinforce students' strengths and build on their weaknesses but also to empower students with the abilities and skills to evaluate and examine the work of their peers and ultimately of their own (Weaver & Esposito, 2011).

The benefits of giving feedback exceed the benefits of getting feedback from the teacher to students as the first involved activating students' metacognition and the ability to question their learning development. Fletcher & Shaw (2012) note that students are given the responsibility to take charge of their learning and the assessment process, they did better and scored higher compared to their peers in the teacher directed assessment group. Lafave et al. (2013) describe this type of assessment as a learning autonomy exercise for the students while Bryant & Carless (2010) named it as part of a self-regulated learning process wherein students benefit from giving and receiving feedback from each other in a group work.

Peer assessment is a form of collaborative work which promotes social interaction and provides room for students to help and learn from each other's behavior and ideas (Anson & Goodman, 2013, p. 27). For such a type of collaboration to be effective, a pre-set criteria for evaluation is needed as a standard to help students organize their thoughts and pass useful judgments on their peers' work. Consequently, during this collaborative work, peer-assessors will engage in "cognitively demanding activities" which will ultimately reinforce their performance that leads to better understanding of the topic.

The implementation of (PA) in school is one strategy that has resulted in positive learning outcomes (Yurdabakan, 2010). Studies on the effect of peer evaluation on students' writing autonomy by Moussaoui (2012) on 30 undergraduates in Algeria and on writing performance by Xiao & Lucking (2008) on 232 undergraduates in an American university revealed that there was a significant influence of peer assessment on the improvement of the students' writing performance and thinking skills in the first study and writing performance and students' satisfaction in the latter.

Similarly, Johnson and Winterbottom (2011) conducted an experimental study on female students at girls-only rural comprehensive school in the UK, where the ages of subjects ranged from 11 to 18 years old. The researchers examined how the use of self and peer assessment among girls influence their writing performance. The results showed that there is no significant influence of peer assessment on girls' school students.

Likewise, Van Gennip, Segers, and Tillema (2010) conducted an experimental study on 62 third year male students in Dutch secondary vocational education to measure the role of interpersonal variable and conception of peer assessment in vocational education. The researchers used a set of apparatuses such as psychological safety, trust, value diversity, interdependence and conceptions of peer assessment. The results showed that there is a significant influence of peer assessment as an assessment approach to support learning.

PA also has been studied by Gamlem and Smith (2013) in a study 11 school students. The researcher conducted research on five female and six male students aged between 13 to 15 years old in Norway. They attempted to explore the students' perception of classroom feedback through observations and interviews. The findings showed significant influence of classroom feedback on students' perception of classroom feedback. For PA to succeed, students are required to take a positive stance towards using this strategy in classrooms to improve their assessment practices and writing performance.

Peer assessment training has been shown to be effective to influence learning and to articulate judgment when it is conducted within a specific time framework. Nicolaidou (2013) conducted a case study on 24 primary school pupils to determine how student writing performance and peer feedback developed over time through the use of electronic portfolios. The results showed that there was a statistically significant improvement on student's writing performance and peer feedbacks improved from simple feedbacks to constructive feedbacks over time.

In summary, there are three pedagogical factors that have great influence on EFL students' writing performance. While most peer assessment studies have been done primarily on native speakers of English and with Asian EFL learners, not much research has been conducted to investigate how Arab EFL students would respond to such learning strategy. Thus, the aim of this study is to investigate the effects of using peer assessment training on the writing performance of Arab EFL high school students.

3. Methodology

3.2. Search Design

This study used a quasi-experimental research design. It is a type of experimental design between groups without random assignment of participants, as the purpose was to determine whether the program or intervention affects the participants of the study (Creswell (2002); Nyström (2008); Shadish, Cook, and Campbell (2002)). This design served the experimental nature of this study in finding the causal relationship between the independent variable and the dependent variable to see if there was a statistical difference in the effect of PA training on the writing performance to answer the following research question: What is the effect of using PA training on the writing performance of Arab EFL high school students?

3.3. Participants

The participants in this study were selected based on convenience sampling from the total population of the Arab EFL high school students (N= 950) who were studying in eleven schools for Arab students in Malaysia. The sample consisted of sixty female and sixty male students from two of the eleven schools. The students' first language was Arabic and they studied English as a second language. All of the 120 participants, they are between 15 and 16 years old. This age group was selected since it is the age range that reflects cognitive maturity where metacognition is active (Tsivitanidou, 2011).

3.4. Instrument

As the purpose of this study was to investigate the effect of using peer assessment training on the writing performance and writing improvement among Arab EFL high school students in Malaysia; an instrument named Analytical Marking Scale was used for data collection Alderson et al. (1995).

3.4.1. Analytical Marking Scale

The analytical marking scale developed by Alderson et al. (1995) was adapted in this study to mark the students' writing task in the pre-test and post-test to collect data in answering the research questions. The students in the experimental group used this instrument to score their peers' written tasks after receiving the training on PA. The Analytical marking scale took place in a form of a rubric that helped the students to mark the writing section in a consistent way and to minimize the possible discrepancy in rating without a rubric. It consists of seven criteria which are the content, organization, cohesion, vocabulary, grammar, mechanics and spelling. These criteria were designed to suit undergraduate students to see whether there could be different in writing performance.

However, the instrument was improvised to suit the learning level and objectives for the targeted high school students. Thus, modifications made were and only five criteria of Alderson Analytical Marking Scale were used. The other two parts were removed because these criteria were not suitable for high school students. The five criteria used were contents, organization, vocabulary, grammar, and mechanics. The changes were in the structural questions and marks that were given to the students during training. Meanwhile, Alderson was given a high marks to assess undergraduate students.

In the content part of this instrument, a score ranging between 0-3 was given based on their ability to write about the main idea, supporting ideas and giving examples. In the organizational section, a score between 0-3 was given according to their ability to follow a required format (introduction, body, and conclusion). In the vocabulary part, a score range from 0-3 was given based on their choice to correct words that were related to the topic. In the grammatical part, a score between from 0-3 was given based on their correct usage of grammar, namely verb usage and sentence structure. In the mechanics section, a score between 0-3 was given based on the number of spelling and punctuation mistakes made by the writer. In totally, the instrument yields a score of fifteen.

In order to see the reliability of the two raters, the student t-test was used to compare the means of their rating. The test was used to find the variations between the marking score of two groups by comparing the mean values in two sets of data.

3.5. Research Procedures

The study was conducted over three stages. Pretest stage, Intervention stage or the Peer Assessment Training stage, and Posttest stage.

3.5.1. Pre-Test Stage

In this stage, 120 participants were purposively assigned into two groups, the control (N=60) and experimental groups (N=60). All participants in both groups were asked to complete a written task on a given topic that was prepared for this purpose. This topic was selected from the students' course material thus deemed suitable for the students' level in the 10th grade. The task was to write a descriptive essay to a friend describing your movement to a new school. The purpose of this test was to find out the students' writing performance before they were engaged in the PA training (Boon, 2013; Al-Barakat, 2009). They were given 45 minutes to complete the writing task and it was conducted during their class schedule.

The students' essays were examined and rated by two independent expert teachers in the area of writing. Please see Table 1 for the result of pre-test.

3.5.2. The Intervention: Traditional Instructions and PA Training

Following the pretest, students who participated in this study received a nine-week training course with a total of thirteen and a half hours. They were assigned accordingly into two groups, control and experimental with each group consisted of sixty students.

3.5.3. Traditional Instructions to the Control Group

The control group was taught to write English essays in nine weeks which was equivalent to thirteen hours. Hence, in the first lesson, the students were introduced to writing a descriptive essay. Then they were further taught how to address the content of the topic. After that, the teacher discussed the typical components of the essay specifically on how to write the introduction, the body, and the conclusion. The teacher guided the students on the use of vocabulary, grammar, spelling and other writing mechanics as necessary. After each lesson, students were asked to practice writing on a topic and submitted their work in the next class session to be marked by the teacher.

In a mode of traditional instruction, the teacher teaches and later students are asked to practice before they hand in their first draft to the teacher for evaluation (Devon, Paterson, Moffat, & McCrae, 2012). Thus, what are severely lacking in the traditional instructions are the feedbacks to improve the draft, to minimize errors, to develop underdeveloped ideas, correct inaccurate word choice and to reorganize cluttered thoughts. This is still the practice in many English classrooms (Saito, 2008), and this was the practice in the control group, which was to see how well these students write on their own after the writing lesson by the teacher, without peer feedback and peer assessment.

3.5.4. Peer Assessment Training for the Experimental Group

Students in the experimental group received the same lesson on descriptive essay writing by a teacher. The difference between both groups was the focus on peer assessment training PA training after the writing practice. Like the control group, the students in the experimental group were introduced to descriptive essay writing. In the first week, the teacher discussed the criteria of the Analytical Marking Scale, the main instrument for PA training, elicited the students' agreement on its areas, and discussed what feedback means, how it can be exchanged, and explained the task to the students and responded to their questions. Over the next week, students were introduced to the task where they learned how to address the content of the topic. After that, the teacher discussed the typical components of the essay specifically on how to write the introduction, the body, and the conclusion. During training, inputs on vocabulary selection, grammar usage, spelling and other writing mechanics were given and discussed. Later, like the control group, the students were asked to write a draft essay to be submitted in the next class session. The difference between the control and experimental groups lay here, where, instead of submitting to the teacher to mark their essay, the students submitted or in other words, exchange their draft essays with their peers in the classroom. Under the supervision of their teacher, students read their peers' essays and write their comments using the pre-set criteria as in the Analytical Marking Scale. Once the students received their papers with the feedback from their peers, they had to rewrite an improved second draft based on the comments.

3.5.5. Post-Test

The purpose of this stage was to determine the effect of using PA training on writing performance versus the traditional instruction and to gauge students' response on the on Peer Assessment training.

All participants (N=120) in both control and experimental groups were asked to write a posttest essay on a given topic deemed suitable for their age and linguistic level. They were given 45 minutes to write the essay. After completing the task, the essays were collected and photocopied. The original copies of the students' essays were marked and rated by two independent raters. The independent raters were also asked to comment and give feedbacks on each essay separately. The photocopies of the essays were redistributed to students and they were asked to assess the writing production.

The scores obtained by both groups as given by the two independent raters were recorded and later compared to the scores obtained from the essays in the pretest stage. The two independent raters' assessment and feedbacks were compared to the students' assessment to determine to the similarities and differences in students and teacher's rating.

The purpose of the test was to find out if there was an effect of using peer assessment training on writing performance among Arab's EFL high schools students in writing process after training students on PA in the classroom. Indeed, both groups have been set for the post-test (see table 3.1). The control group was given a traditional training on how to write a descriptive essay, while the experimental group was given a PA training while learning how to write a descriptive essay.

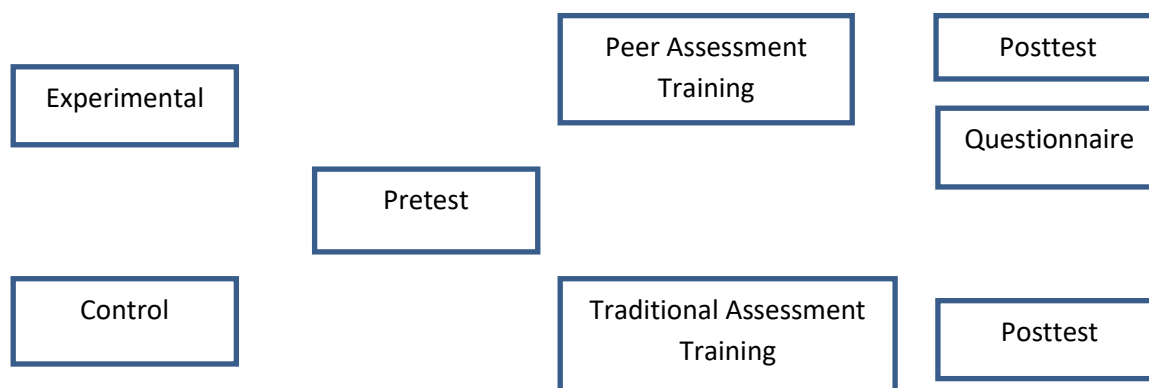


Figure 3.1: The procedures of the study

3.6. Research Question

To address the research question, "What is the effect of using PA training on the writing performance of Arab EFL high school students?" a pre-test was used to determine the students' writing performance prior to training. A pre-test was run in order to find out whether there was a significant difference between the control group and experimental group in their writing performance. The results of pretest and posttest are shown in Tables 1 and 2.

Table 1: *T-test of pretest scores between control and experimental groups.*

Group	Means	Sd	Df	t-v	p-v
Control	6.60	1.73	118	-5.325	0.000
Experimental	8.45	2.06			

F (118, t = -5.325, p<.01)

Independent samples t-test was employed in order to find out the significant difference between the pretest scores of the control and experimental groups. The purpose of this test was to find out the level of the students before engaging them in peer assessment training in the process of writing. Table 1 indicates that the mean score of pre-test for the control group is 6.60 and the standard deviation is 1.73 and the experimental group is 8.45 and the standard deviation is 2.06 in students' writing performance. The result indicates that students in experimental group scored significantly higher than students in control group F (118, t = -5.325, p<.01). In this study assigning the subjects randomly to intact groups was not possible as assigning them randomly, since it would affect their schedule and disrupt their classroom learning. Moreover, a quasi-experimental design was used to control the effect of the pre-test.

Table 2: Adjusted and unadjusted group means and variability for writing performance, using Post-test scores as a covariance.

Group	N	Unadjusted		Adjusted	
		M	Sd	M	SE
Control	60	10.53	1.56	11.57	1.83
Experimental	60	13.07	.83	13.88	.152

Table 2 shows the descriptive statistics for peer assessment training. The statistics indicate that the experimental group is more affected in writing performance than the control group after training them on peer assessment. Between the two groups, control group had the lowest means on peer assessment, whereas experimental group had the highest means on peer assessment (Baker, 2008; Friedman, Cox, & Maher, 2008). The means and standard deviation for the experimental and control groups on writing performance, before and after controlling for the peer assessment training. The mean score of the control group is 10.53 and the standard deviation is 1.56 and the experimental group is 13.07 and the standard deviation is 0.83. The researcher found that a single experience with peer assessment training was necessary for building Arab EFL high school students probability and confidence. As an evident from this table, virtually there is a difference between the experimental group and the control group after the peer assessment is controlled.

Analysis of covariance (ANCOVA) was further conducted to determine a statistically significant difference between the control and experimental groups on their posttest writing performance scores controlling for PA training. The idea behind controlling peer assessment is to test the extent of using peer assessment among Arab EFL high school students.

Table 3: Analysis of Covariance for Writing Performance among Groups, using post-test scores as a covariate.

Source	Df	Mean square	F	Sig	eta ²
Posttest	1	9.45	2.94	.094	.71
Group	1	32.17	10.83	.014	.61
Error	117				

F (1.117) =10.83, $p < .05$, eta squared = .61)

The result indicates that after controlling for the peer assessment training, there is a significant difference between the experimental and control groups in writing performance, $F(1.117) = 10.83$, $p > .05$, eta squared = .61). The result of eta squared (.61) indicates that there is a significant effect of peer assessment training among Arab EFL high school students. This result further suggests that engaging high school students in peer assessment training was useful and beneficial for Arab EFL high school students to improve their writing performance through peer feedback.

A paired samples t-test was run in order to find out if there is a significant difference in the experimental group before and after the training. Paired samples t-test is used to find out the significance difference of the two means scores of the same group.

Table 1.4: Paired t-test Result for Pretest and Posttest of PA Training.

	M	SD	T	Df	Sig.
Pretest	7.53	2.12	-31.33	119	P = .000
Posttest	11.80	1.79			

On average the students performed better in the posttest (M= 11.80, SD= 1, 79) as compared with the pretest (M=7.53, SD= 2.12). Based on the results of Paired Sample T-test t (119) = -31.33, p=.000, 95% CI [-4.55, -4.01]. The result of paired t-test shows that the peer assessment training had a significant effect on Arab EFL high school students.

4. Discussion

The research question was aimed at determining the effect of using peer assessment training on the writing performance. The finding indicates that there is a significant difference between the experimental and the control groups, after evaluating the effect of using PA training on the writing performance of Arab EFL high school students. Data shows that students in experimental group scored higher than the control group. The study revealed that PA approach allows students to use the success criteria to comment on the strength of each other's work. This helps students identify their own strength or performance. A result suggests that, when students properly received PA training, they can successfully improve their writing skill. This finding is consistent with the previous studies conducted by Birjandi and Hadidi Tamjid (2011); Xiao and Lucking (2008) and Lundstrom and Baker (2009). Lundstrom and Baker (2009) suggested that giving and receiving peers' feedback on writing seems to improve students' abilities in writing.

On other hand, this finding is not consistent with the study made by Wang (2014). The authors' claimed that students who do not have enough knowledge to assign essay topic could not develop their own feedback in writing performance. This may be due to different methods exploited in this study when compared to the work of Wang (2014). The researcher in this study used PA training and questionnaire, while Wang (2014) used only questionnaire and interview questions. However, the result from the present study indicates that students can develop their feedback in writing performance during peer assessment training. This might probably be due to different methods been exploited in this study when compared to the work of Wang (2014).

5. Conclusion

The current study investigates the effect of using peer assessment training on writing performance among Arab EFL high school students. The findings of this study indicate that peer assessment training poses a very practical learning tool to improve students' understanding and retention of lessons learnt from the teachers.

In addition, the findings of this study demonstrate that peer assessment could be suitable for high school students to help them to increase their performance in writing. Training students on peer assessment enhances practice and exchange of comments during classroom. Peer

assessment training can be a part of student activity in writing used to engage students in giving and receiving purposeful comments from their classmates. Peer assessment training increases students' interaction and boost their self-confidence.

About the Authors:

Abdallah Almahasneh is PhD candidate at Universiti Malaysia Terengganu. He holds a M.Sc in applied linguistics from universitit Malaysia Terengganu and a lecturer at university of Tabuk in Saudi Arabia. His research interest includes fields in language linguistics. ORCID ID: <https://orcid.org/0000-0001-6307-9762>

Samsiah Abdul-Hamid is Senior English language lecturer at Universiti Malaysia Terengganu. Her research interests include all language skills in EFL. She won a golden medal on her tenses/aspect teaching innovation at an Annual Research and Innovation Week in Kuala Terengganu, Malaysia. She has published a few books and her two most current publications are Mastering English for Employment Second Edition (2018) and Writing Research proposals in Education (2018). ORCID ID: <https://orcid.org/0000-0002-0347-0070>

6. References

- 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.
- Alias, M., Masek, A., & Salleh, H. H. M. (2015). Self, Peer and Teacher Assessments in Problem Based Learning: Are They in Agreements? *Procedia - Social and Behavioral Sciences*, 204, 309-317.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation: Ernst Klett Sprachen*. Cambridge University
- Baker, D. F. (2008). Peer assessment in small groups: A comparison of methods. *Journal of Management Education*, 32(2), 183-209.
- Birjandi, P., & Hadidi Tamjid, N. (2011). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533.
- Boon, S. I. (2015). The role of training in improving peer assessment skills amongst year six pupils in primary school writing: an action research enquiry. *Education 3-13*, 43(6), 666-682.
- Devon, J., Paterson, J. H., Moffat, D. C., & McCrae, J. (2012). Evaluation of student engagement with peer feedback based on student-generated MCQs. *Innovation in Teaching and Learning in Information and Computer Sciences*, 11(1), 27-37.
- Friedman, B. A., Cox, P. L., & Maher, L. E. (2008). An expectancy theory motivation approach to peer assessment. *Journal of Management Education*, 32(5), 580-612.
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150-169.
- Creswell, J. W. (2002). Educational research: Planning, conducting, and evaluating quantitative: *Prentice Hall Upper Saddle River, NJ*.

- Gefen, D., & Ridings, C. M. (2002). Implementation team responsiveness and user evaluation of customer relationship management: A quasi-experimental design study of social exchange theory. *Journal of Management Information Systems, 19(1)*, 47-70.
- Johnson, R. (2004). Peer assessments in physical education. *Journal of Physical Education, Recreation & Dance, 75(8)*, 33-40.
- Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: a study of peer and self assessment in a girls only class. *Educational Studies, 37(4)*, 391-403.
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18(1)*, 30-43.
- Nicolaidou, I. (2013). E-portfolios supporting primary students' writing performance and peer feedback. *Computers & Education, 68(0)*, 404-415.
- Nyström, J. (2008). A quasi-experimental evaluation of partnering. *Construction Management and Economics, 26(5)*, 531-541.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing, 25(4)*, 553-581.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference: *Wadsworth Cengage learning*.
- Sluijsmans, D. M. A., Brand-Gruwel, 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.
- Topping, K. J. (2009). Peer Assessment. *Theory Into Practice, 48(1)*, 20-27.
- Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21(4)*, 506-519.
- 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.
- van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: *The role of interpersonal variables and conceptions. Learning and Instruction, 20(4)*, 280-290.
- Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19(0)*, 80-96.
- Weaver, D., & Esposto, A. (2011). Peer assessment as a method of improving student engagement. *Assessment & Evaluation in Higher Education, 37(7)*, 805-816.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11(3-4)*, 186-193.
- Yurdabakan, I. (2010). The investigation of peer assessment in primary school cooperative learning groups with respect to gender. *Education 3-13, 39(2)*, 153-169.