

Interrater Scoring of Public Speaking Performances in English Language Teacher Education Program

Aynur Kesen Mutlu¹

¹ English Language Teaching Department, Istanbul Medipol University, Turkey

Correspondence: Aynur Kesen Mutlu, English Language Teaching Department, Istanbul Medipol University, Turkey. E-mail: aynur.kesen@gmail.com

Received: April 12, 2018

Accepted: May 25, 2018

Online Published: August 28, 2018

doi:10.5539/ies.v11n9p12

URL: <https://doi.org/10.5539/ies.v11n9p12>

Abstract

Based on the constructivist learning principles, self-assessment has been a targeted topic for many studies in the field of teacher education. Its importance and its leading to learner empowerment have been discussed for long. This current study in this line tries to move one step further by adding a correlative comparison between instructors' and students teachers' grading as well as searching into students' views on self-assessment in Oral Communication Skills Course at English Language Teaching Department of a private university in Turkey. Interrater consistency was examined throughout the study. This study involves 21 student teachers who assessed their speaking performances five times using a micro-analytic rating scale. In the analysis of data, both qualitative and quantitative methods were utilized. Both data sets suggest that there is a high correlation between instructor and student teachers grading. The study has got some implications for curriculum designers, instructors and teacher candidates.

Keywords: formative assessment, interrater scoring, learner empowerment, self-assessment, teacher education program

1. Introduction

Formative assessment methods have recently been voiced in educational context as a result of the mainstream of constructive psychology. The demands of current era have shifted the role of education from transmitting limited knowledge in a specific domain to making them lifelong learners (Dochy, Segers, & Sluijsmans, 1999; Falchikov & Goldfinch, 2000). In doing so self-assessment helps learners to become reflective practitioners (Schön, 1982) makes them active participants of their own learning (Falchikov & Goldfinch, 2000). Though self-assessment sounds promising in learning, broad range of literature stresses the validity and accuracy construct of this method. The studies indicate that the accuracy of self-assessment is enabled through experience in assessment (Falchikov & Goldfinch, 2000) and having clear criteria of assessment (Sadler & Good, 2006). In the absence of experience, the problem in self-assessment is the inflated grades given by learners who are unskilled in terms of assessment (Kruger & Dunning, 1999). It is also indicated in some studies (Ross, 1998; Ward, Gruppen, & Regehr, 2002) that students' self-assessment lacked reliability when compared to some external factors one of which is teacher assessment. For this very specific reason, searching into consistencies and inconsistencies between students' self-assessment and teacher assessment has been one of the popular areas of research in recent years (Barrot, 2015; Ünalı, 2016; Baleghizadeh & Hajizadeh, 2014).

1.1 Self-Assessment and Teacher Assessment

It is well known that with the popularity of alternative assessment and constructivism in language teaching, self-assessment has become an important component of language classroom assessment techniques. From the perspective of Bailey (1998), self-assessment could be defined as "procedures by which the learners themselves evaluate their language skills and knowledge". In a similar way, Mousavi (2012) defines self-assessment as "an individual's own evaluation of his/her language ability and this evaluation are generally based on how good the individual is at specific language skills or how well the individual can use different styles of the language". Self-assessment provides many advantages to language learners in many ways. By being involved in self-assessment, learners can "monitor their learning process, become active learners, develop their metacognitive knowledge, enhance their own learning, develop a better understanding of the purpose of the assignment and the assignment criteria, increase their motivation and involvement, take responsibility for their own learning, and and

think critically” (Searby & Ewers, 1997; Sluijsmans, Dochy, & Moerkerke, 1999; Allam, 2004; Rourke, 2013; Orsmond & Merry, 1997; AlFallay, 2004; Butler & Lee, 2010; Black, 2009; Brown & Hudson, 1998). Despite all the advantages listed above, self-assessment could be perceived as an assessment technique which does not get the required attention as it should. The reasons why students are not given much chance to evaluate themselves include “students’ tendency to overestimate or underestimate their performances as compared to teacher assessment, students’ lack of assessment skills, students’ tendency to perform the assessment based on potential rather than actual ability” (Lee, 2016; Brown & Hudson, 1998; Karnilowicz, 2012; Lew, Alwis, & Schmidt, 2010). Teacher assessment as opposed to self-assessment “was accompanied in students’ minds by instructors, since they are the authorized persons to make decisions about students’ progress and achievement” (Thawabieh, 2017). Louis and Harada (2012) present a comparison of teacher and self- assessment as follows:

Table 1. Comparison between students’ self-assessment and teacher-assessment

Teacher focus	Student focus
Teacher tells and student listens	Teachers and students are co-learners
Teacher uses summative assessment	Teacher and student together use formative
Teacher is uncertain of the student ability to assess his work	Teacher believes that self-assessment is a learnable skill.

Many researchers conducted studies pertaining to comparison between self-assessment and teacher assessment.

1.2 Studies on Teacher Assessment and Students’ Self-Assessment

Many studies have investigated the correlation between teacher assessment and self-assessment in various fields of education. Since the aim of this paper is to focus on interrater consistency between students’ self-grades and teacher grades in assessing public speaking performances of student teachers enrolled in the department of English language teaching, related literature comprises studies conducted on teacher assessment and self-assessment in speaking performances of language learners. In a study by Lee (2016), self-assessment and teachers’ assessment were compared. The participants of the study were teachers and students of a Korean-English Program. It was concluded in the study that students’ self-grades were similar to the teachers’ evaluation; however, there were no similarities pertaining to content of the teacher evaluation and students’ assessment. The study of Dlaska and Krekeler (2008) searched into students’ ability to accurately assess their own pronunciation skills in comparison with their teachers’ assessment. It was concluded in the study that students had difficulty in assessing their pronunciation skills. In another study by Lundquist, Momary, and Rogers (2013), students’ self-assessment of their communication skills was compared with that of their teachers. The findings showed that students’ self-assessment scores were lower as compared to scores given by the teachers. In a study by Chen (2008), students’ self-assessment and teacher assessment were compared on the basis of a training program designed for students. The findings indicated that as a result of the training program students’ self-assessment scores displayed a higher correlation towards the end of the training. In his study, Tavakoli (2010) explored the correlation between self-assessment and teacher assessment in a speaking test and he concluded that the correlation between students’ self-rating and teacher-rating was high.

1.3 Interrater Reliability in the Context of Self-Assessment and Teacher Assessment

Interrater reliability could be defined as “one type of internal consistency measure that differs from scale reliability in that it evaluates the level of agreement among raters versus the reliability of the assessment itself” (Porter & Jelinek, 2011). There are certain factors that affect interrater reliability and these include “rater training, rater selection, accountability for accurate rating, rubric design, type of rubric scale, and pilot programs and redesign” (Graham, Milanowski, & Miller, 2012). As for the interrater reliability between self-assessment and teacher assessment, Barrot (2015) lists the factors that affect interrater reliability as follows:

- Type of language skill
- Rating scale
- Rator factors
- Assessment training
- Students’ skill
- Students’ proficiency level

Interrater reliability has been investigated in some studies searching into the correlation between self-assessment

and teacher assessment (Alfallay, 2004; Butler & Lee, 2006; Chang, Tseng, & Lou, 2012 in Barrot, 2015). Despite numerous studies on the comparison between self-assessment and teacher assessment (Chang, Tseng, & Lou, 2012; Karnilowicz, 2012; Ünalı, 2016; Tavakoli, 2010; Lee, 2016), studies on interrater reliability between self-assessment and teacher assessment of speaking skill seems scarce in Turkish context where English is the medium of instruction in most educational institutions. There seems to be a gap in exploring the correlation between students' self-assessment and teachers' assessment in English speaking classes in Turkey. This study attempts to fill this gap by searching into interrater consistency between students' self-grades and teacher grades in assessing public speaking performances of student teachers enrolled in the department of English language teaching. The present study also differs from the previous similar studies on interrater consistency in that it searches into the differences and/or similarities between students' self-assessment and teacher assessment in regard to who chooses the presentation topic. As to provide some in-depth perspectives on the issue and contribute to the related literature on Turkish studies, this study examines the following research questions:

- 1) Is there any interrater consistency between students' self-grades and teacher grades in assessing public speaking performances of student teachers over a five-week period?
- 2) Is there any difference and/or similarities between students' self-assessment and teacher assessment on the basis of who chooses the topic of the presentation?
- 3) What are the reflections of student teachers in regard to self-grading?

2. Methodology

2.1 Context of the Study

The present study was conducted in English Language Teaching Department of a private university in Istanbul/Turkey. Data of the study were collected during Oral Communication Skills Course over a five-week period.

2.2 Participants of the Study

The participants of the current study all of whom were freshmen were 21 teacher candidates who assessed their own speaking performances ten times using a micro-analytic rating scale. 5 of the participants were male and the rest were female. At the time of the study, the participants were taking Oral Communication Skills Course and Advanced Reading and Writing Course. Their linguistic background was supposed to be similar since they all passed Proficiency Exam prepared and administered by English Preparatory Program of the university they were enrolled in. All the participants took part in the training session designed for using the micro-analytic rating scale for their oral performances. The teacher of the participants who was teaching "Oral Communication Skills Course" also took part in the study.

2.3 Data Collection Tools

The data for the current study were collected through a micro-analytic rating scale designed for speaking performances (Barrot, 2015) and students' reflection papers. The rating scale consisted of 27 items 11 of which were content-related and the rest were delivery related. There were two versions of the scale: one designed for the teacher and one designed for students. All the items were the same in both versions except for the point of view used [WU2]. The scale was designed as a 5 point Likert Scale (See Appendix). The second data collection tool used in the study was the reflection papers that students wrote upon completion of each oral performance and the assessment. Participants were told to reflect their opinions pertaining to their self-assessment and teacher assessment after each presentation session. Participants were told that there were no limitations regarding the time they spent for writing the reflection papers.

2.4 Procedure

Before collecting data through students' presentations in class, all the participants attended the training session during which they were informed about both the purpose of the study and the use of rating scale they were to use after their oral performances in class. A detailed introduction of the assessment rubric was presented by the researcher. Each of the components was explained and the first step of the training session lasted until the researcher was sure that every item in the scale was clear. As the second step, all the participants watched a video-taped presentation and assessed the performance of the speaker on the basis of the rating scale to be used in the study. The rationale behind involving the participants in such a task was to make sure that they had no difficulty in assessing the speaking performance. The discussion about the assessment of the demo presentation lasted until participants had no questions. After the training session, a list of presentation topics was given to all participants and each of the participants chose the presentation topic for each week. As to search into differences and/or

similarities between students' self-assessment and teacher assessment regarding who chooses the topic, each participant chose three presentation topics on his own and the other two presentation topics were assigned by the teacher. As to collect data through students' oral performances, each participant was given 15 minutes to present his/her topic in actual classroom setting. Teacher assessment through the rating scale was carried out during the performance of each student. Each student was asked to assess his/her own performance using the rating scale. The performance of each participant was video-taped and the student had a chance to either assess his/her oral performance right after the presentation or at home. The same procedure went on for a 5-week period during which all the participants delivered five different oral performances. As to explore participants' feelings and opinions about their self-assessment, they were asked to write reflection papers at home for each oral performance they delivered in class. At the end of the 5-week period, all the rating scales completed by the participants and the reflection papers were collected and compared with the teachers' assessment of the oral performances through the use of the same rating scale.

2.5 Data Analysis

Data for the current study were analyzed both quantitatively and qualitatively. Analysis of the quantitative data obtained through micro-analytic rating scale was carried out by utilizing Pearson Correlation Analysis along with descriptive analysis via SPSS 15.0 for Windows program. Thematic analysis was utilized as to analyze qualitative data collected from students' reflection papers.

3. Results

3.1 Qualitative Data Analysis

Analysis of the rating scale:

Prior to the parametric data analysis, normality assumptions were checked with Shapiro-Wilk Test and Q-Q Plots and the data is found to be normal ($p > .05$). Outlier values are controlled with the help of Skewness and Kurtosis calculations and outliers are excluded from the further analysis with the use of z values. The data analyzed with the use of the paired sample t-test in order to assess the difference between the scoring of student and teacher ratings of the presentation performance. Each week's student and teacher scores on content, delivery, and total basis were paired. There were 21 pairs on the mentioned basis (see Table 2).

Table 2. Paired Samples Statistics of student and teacher ratings of the rubric

		<i>M</i>	<i>N</i>	<i>SD</i>
Pair 1	student content 1st week	39.28	18	5.60
	teacher content 1st week	38.44	18	6.33
Pair 2	student content 2nd week	43.89	18	6.08
	teacher content 2nd week	42.06	18	6.22
Pair 3	student content 3rd week	43.67	18	5.36
	teacher content 3rd week	44.83	18	5.43
Pair 4	student content 4th week	43.94	18	5.15
	teacher content 4th week	45.28	18	5.78
Pair 5	student content 5th week	44.50	18	6.50
	teacher content 5th week	44.72	18	5.88
Pair 6	student delivery 1st week	54.00	18	8.58
	teacher delivery 1st week	58.61	18	8.69
Pair 7	student delivery 2nd week	61.11	18	10.03
	teacher delivery 2nd week	63.39	18	7.28
Pair 8	student delivery 3rd week	63.06	18	7.73
	teacher delivery 3rd week	65.11	18	8.04
Pair 9	student delivery 4th week	62.56	18	7.36
	teacher delivery 4th week	65.06	18	7.33
Pair 10	student delivery 5th week	64.56	18	9.82
	teacher delivery 5th week	63.72	18	8.37
Pair 11	student total 1st week	93.28	18	12.82
	teacher total 1st week	97.06	18	14.32
Pair 12	student total 2nd week	105.00	18	15.60
	teacher total 2nd week	105.44	18	13.12
Pair 13	student total 3rd week	106.72	18	12.64
	teacher total 3rd week	109.94	18	12.95
Pair 14	student total 4th week	108.67	18	15.01
	teacher total 4th week	110.33	18	12.76
Pair 15	student total 5th week	109.06	18	15.82
	teacher total 5th week	108.44	18	13.97
Pair 16	student total 1st week	93.28	18	12.82
	student total 5th week	109.06	18	15.82
Pair 17	teacher total 1st week	97.06	18	14.31
	teacher total 5th week	108.44	18	13.97
Pair 18	student content 1st week	39.28	18	5.59
	student content 5th week	44.50	18	6.50
Pair 19	teacher content 1st week	38.44	18	6.33
	teacher content 5th week	44.72	18	5.88
Pair 20	student delivery 1st week	54.00	18	8.57
	student delivery 5th week	64.56	18	9.82
Pair 21	teacher delivery 1st week	58.61	18	8.69
	teacher delivery 5th week	63.72	18	8.37

According to the results of the paired samples t-test seen in Table 1, there was no difference between weekly comparisons of the teacher and student ratings ($p > .01$). While this finding is in line with some studies which concluded that there was a consistency between students' self-assessment and teacher assessment (Chen, 2008; Butler & Lee, 2006; Tavakoli, 2010; Chang, Tseng, & Lou, 2012; Barrot, 2015), it is contradictory with the findings of the studies in which inconsistency was found between self-assessment of students and teacher assessment (Ross, 1998; Kruger & Dunning, 1999; Dlaska & Krekeler, 2008; Lundquist, Momary, & Rodgers, 2013). The findings of the current study also revealed that ratings of students and teacher differed significantly between first week and the fifth week on the basis of content, delivery and total rating. More specifically, students' ratings of the first week were significantly different from their ratings in fifth week on all content ($t(17) = -3.57, p < .003$), delivery ($t(17) = -5.36, p < .001$), and total ratings ($t(17) = -5.01, p < .001$). Accordingly, students rated

their fifth week's presentations significantly higher on content ($M_{1st} = 39.28$, $M_{5th} = 44.50$), delivery ($M_{1st} = 54$, $M_{5th} = 64.56$), and total ratings ($M_{1st} = 93.28$, $M_{5th} = 109.06$). In the similar direction, teacher's ratings of the first week were significantly different from the ratings in fifth week on all content ($t(17) = -6.12$, $p < .001$), delivery ($t(17) = -3.57$, $p < .003$), and total ratings ($t(17) = -5.12$, $p < .001$). Specifically, teachers rated students' fifth week's presentations significantly higher on content ($M_{1st} = 38.44$, $M_{5th} = 44.72$), delivery ($M_{1st} = 58.61$, $M_{5th} = 63.72$), and total ratings ($M_{1st} = 97.06$, $M_{5th} = 108.44$).

In addition, on the basis of task type (two types: presentation topic chosen by the student or teacher), ratings of the students significantly differed on the grounds of who chooses the topic of the presentation (see Table 3). More specifically, students' ratings significantly differ on content ($t(17) = -2.21$, $p < .03$), delivery ($t(17) = -2.55$, $p < .03$), and total ($t(17) = -3.35$, $p < .01$) in terms of who chooses the topic. Accordingly, students rated the presentations the topics of which were chosen by themselves higher on all dimensions: content ($M_{student} = 87.83$, $M_{teacher} = 82.94$), delivery ($M_{student} = 123.67$, $M_{teacher} = 117.06$), and total ($M_{student} = 213.67$, $M_{teacher} = 200$). On the other hand, teacher's ratings significantly differ on content ($t(17) = -2.33$, $p < .04$), and total ($t(17) = -2.20$, $p < .05$) in terms of who chooses the topic. Accordingly, the teacher also rated the presentations the topics of which were chosen by students themselves higher on content ($M_{student} = 87.33$, $M_{teacher} = 83.28$) and total ($M_{student} = 215.78$, $M_{teacher} = 207$). Content ($M_{student} = 87.33$, $M_{teacher} = 83.28$) and total ($M_{student} = 215.78$, $M_{teacher} = 207$).

Table 3. Paired samples statistics of student and teacher ratings of the rubric in terms of chosen subject

		M	N	SD
Pair 1	task_teacher_student_content	82.94	18	9.23
	task_student_student_content	87.83	18	8.85
Pair 2	task_teacher_student_delivery	117.06	18	14.51
	task_student_student_delivery	123.67	18	14.68
Pair 3	task_teacher_teacher_content	83.28	18	9.39
	task_student_teacher_content	87.33	18	10.3
Pair 4	task_teacher_teacher_delivery	123.72	18	15.53
	task_student_teacher_delivery	128.44	18	13.18
Pair 5	task_teacher_teacher_total	207.00	18	24.54
	task_student_teacher_total	215.78	18	23.11
Pair 6	task_teacher_student_total	200	18	22.32
	task_student_student_total	213.67	18	22.90

3.2 Qualitative Data Analysis

Analysis of participants' reflection papers:

Reflection papers through which participants reflected their views regarding self-assessment after each presentation were analyzed via thematic analysis (Creswell, 2007). Thematic analysis is a useful method that could be used to investigate different participants' perspectives by presenting similarities and differences as well as unanticipated insights. In addition, thematic analysis helps the researcher in summarizing important features of a large data set (Braun & Clarke 2006; King, 2004). In analyzing the data via thematic analysis, the phases below were followed:

- 1) Familiarizing with data
- 2) Generating initial codes
- 3) Searching for themes
- 4) Reviewing themes
- 5) Defining and naming themes (Nowell et al., 2017)

As to provide validity and reliability of the themes and related responses, the reflection papers of the participants were also analyzed by the second researcher and the same procedures were followed. Upon completion of the analysis of participants' views jointly carried out by two researchers, emerging themes and sample responses were listed. Table 4 displays the themes and a sample response for each theme.

Table 4. Participants' views about self-assessment

<i>Theme</i>	<i>Sample quotation</i>
<i>Help for future career</i>	<i>As a teacher candidate, having a chance to evaluate myself is really helpful. I feel as if I am the teacher.</i>
<i>Fun</i>	<i>Assessing my own performance is fun</i>
<i>Personal development</i>	<i>At first I was stressful because it was the first time for me to evaluate myself. When I compared teacher's assessment with my own assessment, I saw that my grades were lower and this really surprised me. With this kind of evaluation, I have been able to see my weaknesses and I try to improve myself. I believe this self-evaluation process will contribute to my future presentations.</i>
<i>Better understanding of assessment</i>	<i>By being involved in such a technique, I had a better Picture of teacher's assessment. I mean, most of the things are much clearer now and I can understand my teacher's views much better.</i>
<i>Motivation</i>	<i>Having a chance to evaluate my own performance makes me feel much more motivated because I am involved in the assessment.</i>
<i>Self-confidence</i>	<i>Assessing my own performance contributes to my self-confidence. When I evaluate my own performance, I feel that I have the ability to do what teachers do most of the time.</i>
<i>Stress</i>	<i>Assessing my own performance is usually helpful. However, it becomes stressful because I have to do it after every presentation.</i>
<i>Lack of training</i>	<i>Although I am happy with assessing myself and comparing my assessment with that of my teacher, sometimes I feel that it would have been better if I had been involved in such alternative assessment techniques before. I sometimes have difficulty in deciding on which grade I should give.</i>
<i>Objectivity</i>	<i>It is good to evaluate myself but most of the time it is difficult to be objective.</i>
<i>Time</i>	<i>I enjoy presentations; however, in the classroom setting, I sometimes feel that it takes much time to get ready for it. In addition, self-assessment takes a lot of time</i>

The analysis of participants' views pertaining to self-assessment yielded ten general themes under which participants' responses were grouped (see Table 4). We see in Table 4 that the themes fall into two categories: themes having a positive connotation (help for future career, fun, personal development, better understanding of assessment, motivation, and self-confidence) and themes having a negative connotation (stressful, lack of training, objectivity, and time consuming). It is clearly seen that the number of themes which imply participants' positive feelings about self-assessment is higher than that of the themes which imply participants' discomfort and dislike about their own assessment. The findings of the current study are in line with two groups of previous studies which (1) imply positive attitudes of students towards self-assessment (Stefani, 1994; Shahrakipour, 2012; Siow, 2015; Hanrahan & Isaacs, 2018) and (2) underline students' negative attitudes towards their assessing themselves (Falchikov, 1986; Munoz & Alvarez, 2007; Hanrahan & Isaacs, 2018). We can infer from the findings of the present study that from the perspective of the participants, self-assessment is a valuable technique as it helps students with their personal development, self-confidence and future career. In addition, participants find self-assessment as an efficient tool that leads to a better understanding of the assessment process. As opposed to positive effects of self-assessment, the participants believe that being involved in assessing their own performances may cause stress and difficulty in being objective in assessment. It was also stresses by the participants that not being involved in a training program makes it difficult for some to assess their own performances which are also viewed as a time-consuming activity.

4. Discussion

The aim of the current study was to search into the interrater consistency between students' self-grades and teacher grades in assessing public speaking performances of student teachers over a five-week period. The study also investigated the differences and/or similarities between students' self-assessment and teacher assessment on the basis of who chooses the topic of the presentation [WU3]. The last but not the least, students' views regarding self-assessment was explored in the study. Findings reveal that there was no difference between weekly comparisons of the teacher and student ratings, which could be attributed to the assessment training or the scale used in the assessment process. In addition, results of the current study revealed that the ratings of the participating students and the teacher differed significantly between the first week and the fifth week on the basis of content, delivery and total rating. More specifically, students' ratings of the first week were significantly different from their ratings in the fifth week on all content. Accordingly, students rated their fifth week's presentations significantly higher on content, delivery, and total ratings. In the similar direction, teacher's ratings of the first week were significantly different from the ratings in the fifth week on all content, delivery, and total ratings. Specifically, teachers rated students' fifth week's presentations significantly higher on content, delivery, and total ratings. The increase in the ratings of both students and the teacher could stem from becoming more familiar with

the assessment process over a five-week period. In addition, on the basis of task type (two types: presentation topic chosen by the student or the teacher), ratings of the students significantly differed on the grounds of who chooses the topic of the presentation. More specifically, students' ratings significantly differ on content, delivery, and total in terms of who chooses the topic. Accordingly, students rated the presentations the topics of which were chosen by themselves higher on all dimensions: content, delivery, and total. On the other hand, teacher's ratings significantly differ on content and total in terms of who chooses the topic. Accordingly, the teacher also rated the presentations the topics of which were chosen by students themselves higher on content and total. Such findings could be attributed to the possibility of students' feeling more motivated and performing better when they choose the presentation topics. As for the views of participating students on self-assessment, findings of the study revealed that while some students have a positive attitude towards being involved in self-assessment, others are of the opinion that self-assessment is disadvantageous as it requires special training and objectivity and could be time-consuming and stressful.

The current study; however, has some limitations that need to be considered. One of the limitations that may have affected the results of the study specifically the findings pertaining to consistency between students' self-assessment and teacher assessment could be the training that students received only for a very short time. Other factors that may have influenced the findings could be the rating scale, students' background, language skill, and the presentation topics. For these very specific limitations, in further studies, these factors should be considered from a wider perspective. In addition, further research could be conducted at a longer period with a larger group of participants to reach more generalizable results.

The current study also has some implications that foreign language teachers should consider in designing their oral communication skills course. As the study has revealed, self-assessment is viewed as a useful tool though it may be disadvantageous for some students. Since the disadvantageous aspects such as being time consuming and the difficulty in being objective could be overcome by some training, the self as well as teacher assessment should be implemented in teaching/learning process in order to raise awareness on the learning process.

References

- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32, 407-425. <https://doi.org/10.1016/j.system.2004.04.006>
- Allam, S. (2004). *Educational Alternative Assessment: Theoretical and methodological foundations and field applications*. Cairo, Dar Al Arab Thought.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Heinle and Heinle: US.
- Baleghizadeh, S., & Hajizadeh, T. (2014). Self and teacher-assessment in an EFL writing class. *Gist Education and Learning Research Journal*.
- Barkaoui, K. (2010). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *TESOL Quarterly*, 44(1), 31-57. <https://doi.org/10.5054/tq.2010.214047>
- Barrot, J. (2015). Comparing the linguistic complexity in receptive and productive modes. *GEMA Online Journal of Language Studies*, 15(2), 65-81. <https://doi.org/10.17576/gema-2015-1502-05>
- Black, P. (2009). Formative assessment issues across the curriculum: The theory and the practice. *TESOL Quarterly*, 43(3), 519-524. <https://doi.org/10.1002/j.1545-7249.2009.tb00248.x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, J., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675. <https://doi.org/10.2307/3587999>
- Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, 90(4), 506-518. <https://doi.org/10.1111/j.1540-4781.2006.00463.x>
- Butler, Y. G., & Lee, J. (2010). The effect of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31. <https://doi.org/10.1177/0265532209346370>
- Chang, C. C., Tseng, K. H., & Lou, S. J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. *Computers and Education*, 58(1), 303-320.

- <https://doi.org/10.1016/j.compedu.2011.08.005>
- Chen, Y. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262. <https://doi.org/10.1177/1362168807086293>
- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five traditions* (2nd ed.). Thousand Oaks, CA: Sage.
- Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, 36(4), 506-516. <https://doi.org/10.1016/j.system.2008.03.003>
- Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, 11(2), 146-166. <https://doi.org/10.1080/0260293860110206>
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of educational research*, 70(3), 287-322. <https://doi.org/10.3102/00346543070003287>
- Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: Center for Educator Compensation Reform.
- Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings*. Online Submission.
- Hanrahan, S. J., & Isaacs, G. (2001). Assessing Self- and Peer-assessment: The students views. *Higher Education Research & Development*, 20(1), 53-70. <https://doi.org/10.1080/07294360123776>
- Karnilowicz, W. (2012). A comparison of self-assessment and tutor assessment of undergraduate psychology students. *Social Behavior and Personality: An International Journal*, 40(4), 591-604. <https://doi.org/10.2224/sbp.2012.40.4.591>
- King, N. (2004). Using templates in the thematic analysis of text. In C. Cassels, & G. Symon (Eds.), *Essential Guide to Qualitative Methods in Organizational Research*, 256-270. <https://doi.org/10.4135/9781446280119.n21>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), 1121. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Lee, I. (2016). Teacher Education on Feedback in EFL Writing: Issues, Challenges, and Future Directions. *TESOL Quarterly*, 50(2), 518-527. <https://doi.org/10.1002/tesq.303>
- Lew, M. D., Alwis, W. A. M., & Schmidt, H.G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment and Evaluation in Higher Education*, 35(2), 135-156. <https://doi.org/10.1080/02602930802687737>
- Louis, P., & Harada, V. (2012). Did students get it?: Self-assessment as key to learning. *School Library Monthly*, 29(3), 13-16.
- Lundquist, L. M., Shogbon, A. O., Momary, K. M., & Rogers, H. K. (2013). A Comparison of Students' Self-Assessments with Faculty Evaluations of Their Communication Skills. *American Journal of Pharmaceutical Education*, 77(4). <https://doi.org/10.5688/ajpe77472>
- Mousavi, S. A. (2012). *An encyclopedic dictionary of language testing*. Tehran: Rahnama.
- Muñoz, Ana P. & Álvarez, Marta E. (2007). Students' Objectivity and Perception of Self-Assessment in an EFL Classroom. *The Journal of Asia, TEFL*, 4(2), 1-25.
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, 16(1). <https://doi.org/10.1177/1609406917733847>
- Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education*, 22(4), 357-368. <https://doi.org/10.1080/0260293970220401>

- Porter, J. M., & Jelinek, D. (2011). Evaluating inter-rater reliability of a national assessment model for teacher performance. *International Journal of Educational Policies*, 5(2), 74-87
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20. <https://doi.org/10.1177/026553229801500101>
- Rourke, J., A. (2013) 'Assessment 'as' Learning: The Role that Peer and Self-Review can Play towards Enhancing Student Learning'. *The International Journal of Technology, Knowledge, and Society*, 8(3), 2-12. <https://doi.org/10.18848/1832-3669/CGP/v08i03/56302>
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31. https://doi.org/10.1207/s15326977ea1101_1
- Schon, D. A. 1982. Some of what a planner knows. *Journal of the American Planning Association*, 48(3), 351-364. <https://doi.org/10.1080/01944368208976184>
- Searby, M., & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the school of music, Kingston University. *Assessment & Evaluation in Higher Education*, 22(4), 371-383. <https://doi.org/10.1080/0260293970220402>
- Shahrakipour, H. (2012). On the impact of self-assessment on EFL learners' receptive skills performance. *International Research Journal of Arts and Humanities*, 40, 143-164.
- Siow, L. F. (2015). Students' perceptions on self- and peer-assessment in enhancing learning experience. *The Malaysian Online Journal of Education Science*, 3(2), 21-35.
- Sluijsmans, D. M. A., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self- peer- and co-assessment. *Learning Environments Research*, 1, 293-319. <https://doi.org/10.1023/A:1009932704458>
- Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19(1), 69-75. <https://doi.org/10.1080/03075079412331382153>
- Tavakoli, M. (2010). Investigating the relationship between self-assessment and teacher-assessment in academic contexts: A Case of Iranian university students. *The Asian EFL Journal Quarterly*, 12(1), 234-260
- Thawabieh, A. M. (2017). A Comparison between Students Self-Assessment and Teachers Assessment. *Journal of Curriculum and Teaching*, 6(1), 14. <https://doi.org/10.5430/jct.v6n1p14>
- Ünalı, İ. (2016). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education*, 41(1), 67-80. <https://doi.org/10.1080/02602938.2014.980223>
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1), 63-80. <https://doi.org/10.1023/A:1014585522084>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).