The Effect of Teacher, Peer, and Self-Assessment on Subsequent Learning: The Case of Oral Presentations*

Mohammad Salehi**
Sharif University of Technology
Afsaneh Farhang***
Sharif University of Technology

Abstract

There has been a burgeoning of research in teacher, peer, and self-assessment in terms of reliability and validity, the use of rubrics, the implementation of these kinds of assessments in different academic settings, the effect of demographic variables in these assessments, and the benefits of these assessments. However, the effect of these assessments on subsequent learning has rarely been explored. To fill that void, this study set out to examine the effects of these kinds of assessments on subsequent learning as evidenced by students' final exam scores. Two intact classes were chosen and teacher, peer, and self-assessments were employed in a general English class. The scores achieved on the students' oral presentations exhibited a statistically significant correlation between the teacher awarded scores and those on the final examination of general English proficiency as a course. Peer awarded scores could also predict final exam scores, though not as good as teacher awarded ones. Self-assessment failed to correlate significantly with final exam scores and teacher/peer assessments. The reasons for the findings are discussed further in the paper.
Key Words: self-assessment, peer-assessment, teacher-assessment, regression analysis, subsequent learning

1. Introduction

Innovative assessment practices can potentially bring about a significant paradigm shift in the way universities function (Kearney, 2013). Accordingly, adoption of well-organized assessment tasks designed to let the student work actively and collegially in self- and peer-assessment, opens up the opportunity to engage students in the assessment process. As Kearney (2013) contended, the students' disintegration from the assessment process, as a detrimental consequence of traditional assessments, has led to the unaddressed key issues of learning; thus, the students possess almost insufficient skills in relation to the evaluation of their own efforts.

Given all the far-reaching and growing attempts to achieve the goals concerning educational quality and standards, language teaching and learning approaches have undergone a gradual shift from teacher-focused to learner-focused instruction which consequently resulted in an increased emphasis on alternative forms of assessment (Prapphal, 2008).

Teacher-based assessment, as one of the widely embraced assessment approaches during decades, despite the absence of a widely accepted definition, was distinguished from other forms of assessment, as "it involves the teacher from the beginning to the end: from planning the assessment program, through to identifying and/or developing appropriate assessment tasks right through to making the assessment judgments... it allows the teacher to give immediate and constructive feedback to students" (Davison & Leung, 2009, pp. 395-396). In spite of the grounds that teacher assessment/ judgment gained over decades, it has been found to be almost unreliable and inconsistent (Brindley, 2001; Rea-Dickins & Gardner, 2000).

As Dlaska and Krekeler (2008) stated, a large body of research has been conducted to scrutinize the possibility, if any, of the learners' meaningful contribution to their own evaluation. They further referred to self-assessment (alternatively termed self-rating, self-evaluation or self-appraisal) as a tool that is often used to support student-centered learning whose fundamental objective is to increase insight into the

learning process, to promote active learning, and consequently, to save the teacher correction time. The primary characteristic of self-assessment is to engage students in specifying the criteria based on which they determine the extent to which they have met the required learning standards.

As another alternative to traditional teacher-assessment, peer-assessment has been regarded as a helpful approach in higher education to assessment (Falchikov, 2003, 2005; Falchikov & Goldfinch, 2000). A definition and typology of peer-assessment was proposed by Topping (1998) as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal status learners. Products to be assessed can include writing, oral presentations, portfolios, test performance, or other skilled behaviors" (Topping, 1998, p. 20). Given the fact that many teachers have sought to adopt an approach to assessment that is both reliable and efficient, peer-assessment has been employed as an efficient means if it is adopted as a supplementary to teacher feedback.

Research on peer-assessment could also benefit from a vast body of knowledge gained over the past two decades on collaborative learning issues, hence, according to Kollar and Fischer (2010), it could be argued that peer-assessment is a collaborative activity that occurs between at least two peers. With the advent of alternative or non-traditional forms of assessment such as peer-assessment, self-assessment, and so forth, teachers, teacher educators, and researchers have increasingly applied if not entirely replaced the alternative forms.

Despite all the controversies on whether peer and self-assessment could be accurate compared with professional assessment, and the extent to which the peer-awarded and self-awarded scores may correlate with teacher-awarded scores (Kwan & Leung 1996; Grez, Valcke, & Roozen, 2012; Rian, Hinkelman, & Cotter, 2015; Diab, 2016), up to the present, there has been no study conducted to investigate the student-awarded scores' potential to predict their subsequent language performance. The present study is aimed at addressing this gap in the literature by explaining how the teacher, peer, and self-awarded scores

on oral presentations can provide an almost accurate prediction of their future final exam performance.

To meet this end, investigating the existing correlation between the teacher-assessment and peer-assessment is of primary importance so as to consider its probable link to the enhanced learning coupled with a better future academic performance. In this regard, a number of researchers attempted to compare these forms of assessment in the academic context. A good illustration is the comparison drawn by Author (2014) between peer-assessments and traditional teacher-assessment in oral presentations. The results indicated a significant correlation between the two forms of assessment. However, up to now, the question concerning the effect of reciprocal peer-assessment of oral presentations on the subsequent language performance and its potential to predict future performance is still unanswered.

When it comes to the use of rubric as a popular assessment tool among teachers and instructors, the use of an effective rubric for peer-assessment assumes critical importance in the current study. As the previous researchers (Andrade, 2000; Popham, 1997; Quinlan, 2006) mentioned, rubrics as a popular means of communicating expectations for an assignment, are adopted to provide focused feedback on the ongoing performance, and to grade final products. However, Andrade, Du, and Wang (2008) referred to the common intention among educators to define the term "rubric" in slightly different ways. They further provided an almost unanimously accepted definition of the term rubric as a document for establishment of the assessment criteria required for description of the levels of quality that range from poor to excellent.

In addition to the evaluative function of rubrics, there is another important role that they may serve other than assessment; that is, a teaching function. Stiggins (2001) highlighted the potential role of rubrics to enable the learners to broaden the knowledge or skill they possess as well as to assess the quality of their performance. Thus, the extent to which professional assessment may differ from peer and self-assessment is explored on the basis of a rubric for assessing oral

presentation. And, whether each of the assessments may contribute to prediction of their subsequent exam performance will be answered accordingly.

Stiggins (2001) referred to the formative approach to assessment that contributes to make a progress towards the standards articulated in the rubrics. In the present study, the researchers' prime focus is on formative assessment which employs rubrics to promote students' learning as a result of the enhanced reflection on their own or a peer's work (Andrade & Valtcheva, 2009; Panadero, Alonso-Tapia, & Reche, 2013; Sadler & Good, 2006).

The rubric used in the present study (see the appendix) demonstrates that oral presentation involves a combination of delivery skills and language skills like grammar and vocabulary. As a matter of fact, the rubric falls into two categories. One is content and the other one is delivery. The researchers believe that a combination of content and delivery can predict final exam scores. The final exam lacks a presentation but is logically related to language skills that are tapped in oral presentation.

2. The Review of Related Literature
2.1. The Assessment Approaches on the move

To begin with an almost straightforward conceptualization of assessment, Barber and Hill (2014) defined assessment simply as the process to establish what students know and are able to do, which is broadly classified into two categories; assessment designed to support teaching and learning in classrooms; and assessment programs for public reporting, certification, for selection and for system accountability (Barber & Hill, 2014).

As the role of assessment as a driving force towards learning has been widely recognized (Boud, 1990; Lamprianou & Athanasou, 2009), and its remarkable impact on the learners' attention and concentration was underlined by various researchers (Dochy, Gijbels, & Segers, 2006), recognition of the significance of assessment of learning and assessment for learning, summative assessment and formative

assessment, respectively, have been of central concern in studies aimed at reformations in higher education in recent years. Therefore, as Looney, Cumming, van Der Kleij, and Harris (2018) argued, considerable research revolves around teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning (Looney, Cumming, van Der Kleij, Harris, 2018).The assessment of oral presentations is one of the fields in which peer and self-assessment could be applied.

During the last decades, the assessment in higher education has undergone a shift from traditional assessment of academic achievement or assessment of learning towards "assessment for learning" (Dochy, Gijbels, & Segers, 2006). In this regard, Boud (2000) referred to the necessity of proposing new methods of assessment to address the quality of education. Thus, he underscored the importance of sustainable learning and suitable methods to assess that learning. As he suggested, the students need to learn to undertake assessment of their own learning tasks so as to become effective lifelong learners. To this end, they have to decide whether they have fulfilled the appropriate standards for the task, and look for various channels of feedback from their environment, such as their peers, practitioners, and other different forms of sources. He further highlighted the independence of students from teachers or other formal sources as a direct outcome of such practice as learn to work with other learners and employ the expertise and skills in hand in a reciprocal manner (Boud, 2000). Therefore, he deemed it essential that a high quality formative assessment practice be introduced, as it promotes a lifelong learning as a result of the learners' engagement with these practices.

Poehner and Lantolf (2005) introduced summative assessment as a medium to capture the culmination of the learning process, as defined by a given curriculum. Further, they argued for the lack of a forward-looking assessment program that may potentially promote learning. They underlined the problematicity or insufficiency of the summative

assessment without any mediation during the course and any interest in future. Moreover, they highlighted the role of formative assessment (FA) in providing information on the degree of success during the learning process with a central focus on strengths and weaknesses in relation to the given curricular objectives.

With regard to the benefits of formative assessment, Rea-Dickins and Gardner (2000) referred to the ways in which FA serves teachers and practitioners. Accordingly, it helps the teachers manage their teaching program, it provides evidence of student learning, it indexes the extent to which both the teachers and the students have attained the goals and objectives prescribed in the curriculum, and it provides them with sufficient evidence for evaluating their own teaching practice (Rea-Dickins & Gardner, 2000).

When it comes to the implementation of an assessment system in educational programs, a range of questions may arise regarding the tensions between the summative and formative purposes of assessment and the validity and reliability of teacher constructed assessment tasks. Brindley (2001), in particular, examined the ways in which such questions could be addressed by using two instances of education in Australia. Having exemplified the occasions when the outcome-based approaches to assessment proven problematic, Brindley (1998, 2001) referred to a number of political, technical, and practical justifications including tensions in relation to the measurement qualities of teacher-constructed assessment tasks, the manner in which they are administered, doubts surrounding comparability of teacher-developed assessments, and the high expenses and time demands of developing and administering individualized performance assessments (Brindley, 1998, 2001). Thus, he regarded the full adoption of teacher-assessments to be almost premature, or insufficient, which requires an ongoing interplay between expertise, time, and energy to make it well qualified.

In the current study, giving oral presentations, as a compulsory individual task, is viewed as pertinent to formative assessment whose outcomes are investigated in terms of their predictive potential with regard to summative assessment or teacher awarded-scores on final

exam. As noted earlier, the researchers seek to explore the relation between teacher, peer, and self-awarded scores on oral presentations as a formative assessment and the extent to which each could be probably used as an alternative to teacher-awarded scores on summative assessment. As a matter of fact, what is important here is to investigate the predictive power of oral presentation scores which are regressed on final exam scores. The researchers believe that oral presentations, as a part of compulsory individual tasks, are related to formative assessment that is done on an ongoing basis, and to a good extent could reflect one's language proficiency skills whose components are assessed through both oral presentations and final exams.

As the first alternative to wholesome teacher assessment, one could refer to self-assessment whose effect on teaching and learning have been widely researched in the literature. According to Dlaska and Krekeler (2008), in order to promote student centered learning in which providing insight into the learning process, and encouraging active learning are of outmost importance, self-assessment (variously termed self-rating, self-evaluation, or self-appraisal) is often used. Moreover, self-assessment has a central role in saving the teacher correction time and supporting students in dealing with various weaknesses (Dlaska & Krekeler, 2008). They also referred to the studies on self-assessment which are predominantly concerned with the correlations between self-assessment and teacher-assessment. They investigated the reliability of L2 learners' assessment of their own articulation of different speech sounds compared with the sounds produced by a native speaker. The results revealed the students' high recognition of correct sounds in about 90% of the cases, as they were advanced learners of the German language. However, recognition of inaccurate sounds turned out to be significantly more difficult; they could identify only less than half of inaccurate sounds. Furthermore, Dlaska and Krekeler (2008) argued, "errors in judgements are integral to self-assessment. Whether or not self-assessments are considered suitable for language tuition also depends on the importance attached to other factors: self-assessment procedures can enhance the awareness of one's performance, they can

increase learner motivation, and shift the decision making process in the direction of the learner" (Dlaska & Krekeler, 2008, p. 515).

Another efficient supplementary to professional assessment is peer-assessment. Falchikov and Goldfinch (2000) deemed it essential to differentiate between self-assessment and peer- assessment; they argued that in the former, students judge their own work, while the latter they judge the work of their peers. Thus, the recognition of peer-assessment as a means to promote learning as a result of provision of detailed peer feedback, paved the way for introducing peer-assessment into courses.

In a more recent study, Adachi,Tai, and Dawson (2018) investigated both values and challenges that peer and self-assessment introduce in higher education from academics' point of view. The researchers attempted to shed lights on the challenges and inhibitions the academics face while incorporating peer and self-assessment in classrooms. Also, the academics in their study, almost unanimously, referred to the teaching function of peer and self- assessment which provide a learning opportunity for the students (Adachi,Tai, & Dawson, 2018).

In studies that combine peer-assessment and self-assessment of oral performance in L1 universities, student ratings and teacher/professional ratings indicated both disagreement (De Grez, Valcke, & Roozen, 2012; Suñol, Arbat, Pujol, Feliu, Fraguell, & Planas-Lladó, 2016) and agreement (Lanning, Brickhouse, Gunsolley, Ranson, & Willett, 2011s). In this regard, Gessa-Perera (2011) conducted a peer-assessment experiment applied to oral presentations. Having proposed the assessment tasks as learning tasks, he compared the results of the peer-assessment with the teacher-assessment outcomes. It was revealed that peer-assessment contributes to the enhancement of grades, involves students in learning process and allows the teacher to propose changes to improve learning (Gessa-Perera, 2011).

In another study, Suñol, Arbat, Pujol, Feliu, Fraguell, and Planas-Lladó, (2016) aimed to analyze the use of peer and self-assessment in oral presentations as complementary tools for assessment. On the basis

of a study conducted at the University of Girona (Spain), two instruments were implemented to measure students' peer and self-assessments. Moreover, a rubric was designed to guide the assessment process. The results were compared with the marks awarded by the professor. The findings of the analysis turned out in contrast with studies by other authors, in that significant deviations were observed between the two assessment systems. Give the higher marks awarded by students, they contended that "this may be attributed to several factors, including the students' lack of experience in assessment, their level of knowledge and observation skills, or feeling of solidarity with peers and not wanting to do them any wrong in peer assessment processes" (Suñol, Arbat, Pujol, Feliu, Fraguell, and Planas-Lladó, 2016, p. 622).

Topping's (1998) qualitative review of peer-assessment provides an in depth integrative overview of peer-assessment studies in higher education in which students awarded marks or grades to their peers, in group works and projects on various professional skills, whether for their performance on simple multiple-choice tests or on more complex activities such as oral presentations and written works. As an overall summary of the outcomes, Topping (1998) concluded that peer-assessment exerts a positive formative effect on the enhancement of scores and subjective perceptions either through provision of simple quantitative feedback or detailed open-feedback which were not mutually exclusive. Drawing a comparison between the two forms of feedback, he argued "quantitative feedback seems more likely to be unidirectional, distant, and anonymous. Detailed feedback seems more likely to involve personal contact and to be reciprocal or mutual, personalized, and sometimes public. Peer-assessment seems equally likely to contribute to or not contribute to the assessee's final official grade" (Topping, 1998, p. 267). Concerning the validity and reliability of assessment, he further viewed peer-assessment as adequately reliable and valid in a variety of applications; however, according to the reviewed literature, he assumes the detailed and formative assessments less reliable than simple numerical marks or grades. Regarding the

students' attitudes towards assessment of their peers, he concluded that the students perceive this practice a demanding but anxiety reducing task that maximizes learning gains. He also referred to the improved confidence and enhanced appraisal and presentation that the students reported with regard to peer-assessment (Topping, 1998). While peer- and self-assessment differ with regard to whose work the students are evaluating, their own or that of a peer, "both are student-led assessment processes designed to enable students to understand and apply quality criteria to their work, freeing them from teacher dependence" (Brown & Harris, 2013, p. 112).

All things considered, however, as far as the students are not fully cognizant of the assessment criteria and mechanisms, this confusion and insufficient understanding may contribute to inaccurate or unreliable marking. Thus, the crucial role of rubric or a clearly defined criterion for assessment is accentuated.

2.2. The use of Rubrics in Assessment

Andrade, Du, and Wang (2008) defined rubric as a "document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor" (Andrade, Du, &Wang, 2008, p. 3). Further, they investigated the effect of a written assignment as a model, a list of criteria for the assignment, and self-assessing according to a rubric, with respect to the participants' gender, time spent on writing, and their prior rubric use. So as to begin the process of generating a list of criteria for an effective essay, to receive a written rubric, and finally to employ the rubric to self-assessment of the first drafts, the treatment involved application of a model paper. The results indicated that adoption of a model to generate criteria for an effective assignment and employing a rubric for self-assessment has a positive influence on the elementary school students' production of effective writing (Andrade, Du, & Wang, 2008). Therefore, as in an earlier study by Andrade and Du (2005) students reported that their active use of rubrics promoted their learning and academic performance as a result of their focused discussions on the ways in which the rubrics are used to plan an approach to an assignment,

check their work, and reflect on feedback from others, it could be argued that their study provided an indirect support for the learning advantages of rubrics for undergraduates.

Although the widely acknowledged function of rubrics is to grade student work, there are some assertions that they can serve another, however more important, function as well. Stiggins (2001), for instance, argued for the teaching function of rubrics in addition to their evaluative function. Stiggins (2001) asserted that, when rubrics are used as part of a formative, student-centered approach to assessment, they can potentially assist students' development of recognition skills and knowledge as well as the ability to make dependable judgments about the quality of their performance. He also underscored the role of students, as the key users of assessment, who use assessments in the same ways that teachers do in clarification of the standards for quality and providing ongoing feedback on their progress toward those standards.

Andrade and Valtcheva (2009), also explored the effect of criteria-referenced self-assessment on learners' achievement. Having drawn upon the earlier studies on the benefits of self-assessment, they asserted that "the use of criteria-referenced self-assessment can have powerful effects on learning. The effect can be both short-term, as when self-assessment influences student performance on a particular assignment, as well as long-term, as students become more self-regulated in their learning"(Andrade & Valtcheva, 2009, p. 17).

2.3. Assessment of Oral Presentations

Owing to the fact that the leading role of oral presentations, as the researchers' preferred language skill, cannot be overlooked in this study, it is worthwhile to make a reference to the studies in which the evaluation of oral presentations as well as the challenges faced in this regard are taken into account.

Regarding the advantages of a criterion-referenced assessment of oral skills, Cooper (2005) contends that all the standards and criteria which students are expected to achieve to meet certain objectives are

explicitly identified. Also, when assessing oral presentations, Race (2001) deemed it essential that the criteria be clear from the outset, so that the students understand the weighting of each criterion. Cooper (2005) briefly summarized the assessment of oral presentation framework by addressing the assessment qualities of being valid, reliable, fair, unambiguous, manageable, ethical, and being helpful in terms of students' development and provision of a timely formative feedback to teachers. Moreover, she referred to some of the inherent challenges when incorporating this assessment framework such as inadequate skills instruction in verbal and non-verbal communication, hence, it couldn't be valid to test them on these skills. Furthermore, low manageability of the assessment tasks for both teaching staff and students is another challenge that she underlined in addition to the tension between large student cohorts and a quality assessment task, and insufficiency of information to understand what characterizes performance standards that apply to each marking criterion (Cooper, 2005).

Drawing upon Brown, Bull, and Pendlebury's (1997) argument concerning the students' enhancement of oral presentation skills by viewing their own presentation on video and reflection on their own performance in an entertaining and anxiety-free environment, Andeweg, van der Laaken, and Swennen (2005) highlighted a number of obstacles to be overcome in this regard -such as expensive cameras, demanding download time and site access problems, organizational costs, computer storage limits, and video file processing time- and called for new feasible ways of promoting feedback and self-evaluation skills for students in oral presentations.

Moreover, in spite of the usefulness of the feedback students receive from their teacher and fellow students immediately after their presentation, Andeweg, van der Laaken, and Swennen (2005) argued that this form of feedback might be even less than desired since the students find it almost difficult to focus just after their presentation, assessors do not usually take notes, which leads to problematic specification of what their comments are based on, and finally, students

are not always trained to give constructive feedback, consequently, they find the identification of underlying presentation techniques nearly difficult, and instead, focus on conspicuous presentation details such as fillers and fidgeting behavior per se; "this is the so-called halo-effect in which minor details overshadow the observation of more essential behavior" (Andeweg, van der Laaken, & Swennen, 2005, p. 4). However, as noted in the earlier sections, many of the problems concerned in this area could be presumably obviated in the case of presenting a criterion-referenced rubric of oral presentation assessment.

Regarding the adoption of a video recorded presentation for assessment, Andeweg, van der Laaken, and Swennen (2005) proposed the use of annotated recording; therefore, the feedback that the students receive and give would be considerably more concrete and objective that contributes to improved presentation techniques not only for the presenters who receive a more specific feedback, but also for the other students who develop a keener eye (Andeweg, van der Laaken, & Swennen, 2005).

As the aims and findings of the above mentioned studies suggest, the role of teacher, peer, and self-assessments in future performance are of central importance to be investigated from various perspectives in further studies. Thus, the research questions to be addressed in this area are as follows:

1) What is the effect of teacher, peer, and self-assessment on subsequent learning as determined by the students' final exam scores?

2) Is there any potential that each of teacher, peer and self-awarded scores might predict learners' future performance on the final exam?

3. Methodology
3.1. Setting and Participants
This study was conducted using two intact (non-randomized) general English classes in Sharif University of Technology in Tehran, Iran. The classes were held twice a week in three-hour sessions. One class that comprised 10 female and 20 male students (N=30) was held in the morning. The other class with 6 female and 22 male students (N=28)

was held in the afternoon. As one part of the study, the researchers provided these two classes with instructions. Moreover, giving oral presentations was a requirement of the course. The students ranged from 18-20 years of age.

3.2. Instrumentation

The main body of instruments used in this study was a rubric which consisted of 13 criteria that was used in a previous study (authors, 2017). After having made some modifications, the researchers employed the rubric in this study. For example, due to the fact that some students usually tend to copy their presentation materials from online sources, originality was also added to the criteria. The researchers distributed the rubric, whose language was English, in each of the 30 presentations to both the presenter (self-assessment) and the audience (peer-assessment). Students gave a score on the basis of a Likert scale from 0 to 5. The maximum possible score a presenter could get was 65 (13×5). The total number of the criteria was 13 seven of which were related to content and the remaining criteria were related to delivery or presentation skills (see the appendix for the rubric).

The book designated for instruction during the general English course was Inside Reading whose focus was on a series of passages followed by vocabulary and reading comprehension exercises. The oral presentations and classroom discussions served as a complementary to the instruction. Thus, in order to pass a three-credit course in general English, the students were required to give oral presentations in addition to taking a final exam which consisted of cloze passages, reading comprehension, and vocabulary items. The justification behind this requirement could be explained by Magin and Helmore (2001) who rightly mention that students do not take the assessments procedure seriously unless their assessments are counted towards the final grading. It is also useful to note that the purpose of the study was not revealed to the students.

3.3. Data Collection

The data collection procedure took place following each presentation. Besides the rubric according to which the students had to assess one

another's performance (see the appendix), the researchers also provided some spaces so as to let the students write additional comments about the presentations. In this study, as the students already knew one another and the presenter clearly knew who his/her rater was and vice versa, peer-assessment was not conducted anonymously. It is noteworthy to mention that prior to data collection procedure whose central focus was on teacher, peer, and self-assessment of oral presentation, the researchers deemed it essential to train the students how to assess one another's as well as their own performance according to the rubric provided. To this end, during the opening session of the program, the researchers explicated the notions of peer and self-assessment and elaborated on the significance of each criterion. The students were also given some time to reflect on each criterion and ask for further clarification if necessary. And, so as to avoid the problem of teacher variability, the same teacher instructed both classes.

Regarding the amount of time allocated to each presentation, all of the presenters were required to manage their speech within approximately 15-20 minutes. Also, the presenters made a serious attempt to finish the presentations within the time limitation as one of the criteria in the rubric was related to the students' time management skills.

As noted earlier, the researchers' primary objective was to explore the relation if any, between the teacher, peer, and self-awarded scores on oral presentations and the potential effect of each on the final exam performance. Obviously, it could be highly important to give a clear description of the final exam administrated in the closing phase of the study and its subtests. The final exam included 40 items which comprised eight cloze items, eight reading comprehension questions including three passages of almost similar length, and twenty-four vocabulary items. The time allocated to the final exam was 90 minutes overall.

3.4. Data Analysis

As the correlations between teacher-awarded scores and the two sets of peer and self-awarded scores on oral presentations are of central

importance to determine the validity of self and peer-assessment of oral presentations compared to teacher-assessment; and, to explore the probable effect of each on the prediction of learners' future performance, the researchers ran a Pearson product-moment correlation between the mean of teacher-awarded score and each of the two peer and self-awarded scores.

Furthermore, by performing a regression analysis on the collected data comprising the teacher, peer, and self-awarded scores on oral presentations in addition to the scores on the final exam, the researchers sought to determine whether or not each of the three sets of scores could potentially impact, or better said, predict the students' overall performance on the final exam.

3.5. Results and Discussion
3.5.1. Correlation Analysis

The relationship between peer-awarded scores and teacher awarded scores on oral presentations was investigated using Pearson product-moment correlation coefficient. The results revealed that there was a significantly high positive correlation between the two variables ($r = 0.8$, $n = 30$, $p < .001$).

The relationship between self-awarded scores and teacher awarded scores on oral presentations was also investigated. There was a moderate, positive correlation between the two variables ($r = 0.34$, $n = 30$, $p < .001$).

Additionally, the correlations between the final exam scores and the teacher, peer, and self-awarded scores on oral presentations are displayed in Table 1. Accordingly, there is a significantly high correlation between the final exam scores and teacher-awarded scores on oral presentations ($r = 0.8$, $p < .001$.), and between the final exam scores and peer-awarded scores on oral presentations ($r = 0.74$, $p < .001$). However, the observed correlation between the final exam scores and self-awarded scores on oral presentations was considerably low ($r = -.16$, $p < .001$).

Table 1. *The Pearson correlations between final exam scores and teacher, peer, and self-awarded scores on oral presentations*

		final	self	teacher	peer
Pearson Correlation	final	1.000	-.160	.801	.744
	self	-.160	1.000	-.177	.045
	teacher	.801	-.177	1.000	.809
	peer	.744	.045	.809	1.000
Sig. (1-tailed)	final	.	.199	.000	.000
	self	.199	.	.174	.406
	teacher	.000	.174	.	.000
	peer	.000	.406	.000	.
N	final	30	30	30	30
	self	30	30	30	30
	teacher	30	30	30	30
	peer	30	30	30	30

3.5.2. Regression Analysis

To explore the impact or predictive potential that each of the three sets of teacher, self, and peer-awarded scores might possess, a linear multiple regression was carried out. The contribution of each independent variable is presented in Table 2.

Table 2. *Standard multiple linear regression of teacher, self, and peer-assessment of oral presentations on the final exam scores*

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	6.770	2.158		3.137	.004
Self	-.019	.028	-.081	-.674	.506
Teacher	.119	.046	.526	2.572	.016
Peer	.089	.056	.322	1.597	.122

a. Dependent Variable: final

As presented in Table 2, there was a collective significant relationship between the teacher awarded scores on oral presentations and final exam scores (p = 0.16). However, this relationship was estimated to be approximately insignificant about peer-awarded scores (p = 0.12) and self-awarded scores (p= 0.50).

For teacher awarded scores, the recorded beta value was 0.526, while for peer-awarded scores this value was lower (b= 0.32, p<0.001). Also, the contribution of self-awarded scores in prediction of the learners' future performance was substantially insignificant with the beta value of 0.08.

Moreover, the R2 value was 0.67; thus, 67% of the variation in final exam score can be explained by the model containing teacher, peer, and self-awarded scores on oral presentations.

Needless to say that checking for the assumption concerning multicollinearity is of outmost importance in regression analysis, since a good multiple regression would not be achieved with the presence of multicollinearity. According to Pallant (2011) "multicollinearity exists when the independent variables are highly correlated (r= 0.9 and above)". In the current study, as presented in Table 1, a large correlation (r = 0.8) was observed between the two independent variables of teacher

and peer-awarded scores. However, this value despite being substantial, does not exceed the maximum value of r= 0.9 discussed above. Hence, the researchers ensured that the correlation between variables could not exert any potentially negative effect on the contribution of the regression model.

4. Discussion

The fact that assessment can lead to learning is in line with the ideas of Dochy, Gijbels, and Segers (2006) who came up with the notion of assessment for learning. The results of this study also confirm the ideas of Stiggins (2001) in that rubrics can assist students in terms of the quality of their assessments. Quality of assessment in the current research is materialized in the close affinity between the assessments of the teacher and those of the peers. Although the current paper makes no attempt to deal with the validity of assessments, indeed the assessments of the peers can be said to be valid as evidenced by high correlations between teacher awarded scores and those of peer awarded ones. Therefore, the results confirm those of authors (2017) and author (submitted).

Moreover, the approximately high correlation between teacher marking and those of the peers can be explained by the use of the rubrics. Rubrics can tune students to the criteria in the rubrics which will be beneficial for learning (Yucel, Bird, Young, & Blanksby, 2014). It could be argued that the associations observed between the teacher and peer assessments with final scores can lie in the underlying constructs of the rubrics employed and the final exam. It is also noteworthy to mention that the final exam comprised vocabulary, reading comprehension and cloze items. The rubrics consisted of a good number of criteria such as grammatical appropriateness, adequate knowledge of vocabulary, pronunciation, intonation and etc.

The reason why teacher's scores produced better correlations with final scores is that teacher scoring is usually looked at as a yardstick against which students' scorings can be gauged. The fact that teachers' scores can be regarded as a valid assessment is also echoed by researchers such as Cho, Shunn, and Wilson (2006) and Panadero,

Romero, and Strijbos (2013). But still, it is a moot point that teacher's scores can indeed be construed as a valid measurement. We believe that in case of highly seasoned professionals, they scoring can be viewed as valid and trustworthy.

The reasons behind low associations between self-assessment and teacher, peer and final scores can be accounted by the following: First and foremost, students might have over scored or underscored themselves which can lead to correlational discrepancies between self-assessment and teacher/peer assessments. Another justification might have to do with the fact that in each presentation, self-assessment consisted of just one score whereas peer assessment pertained to scorings of the whole class except for the teacher and the presenter. This aggregation of scores could have ironed out idiosyncrasies in scorings whereas the presenter's scoring was singular in nature.

As discussed in the earlier sections, the researchers deemed it essential to highlight the underlying role of rubrics as a part of formative assessment and the substantial effect that may exert on students' development of recognition skills, knowledge, and ability to judge the quality of their performance as well as their peer's (Stiggins, 2001). Furthermore, it bears significance to note the distinction between assessment of learning and assessment for learning, summative assessment and formative assessment, respectively, and the function of each in higher education (Dochy, Gijbels, & Segers, 2006). With these considerations in mind, the researchers sought to explore the correlation between teacher, peer, and self-awarded scores on oral presentations viewed as a formative assessment, and the extent to which each may probably serve as an alternative to summative assessment or teacher awarded-scores on final exam with regard to their predictive potential which was investigated during the study.

5. Conclusions and implications
Based on the results of the study, it can be concluded that teacher and peer assessments of oral presentations can predict end-of-course scores. In case of absenteeism on the part of the students of a class, trust can be placed on teacher and peer-assessments. As a matter of fact, the

distinction between teaching and assessment can be blurred via inclusion of formative assessment. Oral presentations in this study served as a formative assessment tool that served the good purpose of directing the students towards the ultimate goals of the class which eventuated in close proximity of scores of oral presentations and the-end-of course achievement test. In the present study, the researchers attempted to highlight the probability that in the absence of a wholesome summative assessment, peer-awarded scores of a formative assessment (of oral presentations in this study) could be a statistically and empirically proven alternative. However, some further research is still required to apply such findings into a larger scale. Though not documented in this paper, the achievement scores of the two classes were higher than the average of other classes taught in the center.

Teachers are well advised to incorporate alternative assessment in classes in the form of formative assessment. This can lead to better summative assessment. The existence of criteria in the rubrics employed can alert students to learning goals and objectives.

6. Limitations

Despite the fact that there are distinct advantages for the incorporation of peer and self-assessment practices in language learning classrooms, such as the learning that follows from the students' interpretation and evaluation of their peers (Topping, 1998), It is likely that the process of peer-assessment may involve limitations in the current study as in many other peer and self-assessment studies.

One of the most prevalent challenges the researchers encounter in this area is presence of friendship effect which intentionally or unintentionally influences the students' judgment of their peers' performance. It is probable that some students hesitate to criticize their peers, even if peer assessment is conducted anonymously. Additionally, some students might be intended to give extremely low scores to their peers so as to keep their own achievement at a high level. In the current study, the researchers observed a close friendship, three or four cases at most, between the students; hence, to make a firm decision regarding the accuracy and fairness of their assessment could be almost difficult.

Moreover, in the case of self-assessment, the students' evaluation of their own performance might be, to a large extent, influenced by various personality factors the most important of which could be self-confidence that is a contributing factor in scoring process; the students' may overscore or underscore themselves compared to their peers, it was an overriding concern in this study.

References

Adachi, C., Tai, J. H. M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 294-306.

Andeweg, B., van der Laaken, B., & Swennen, M. (2005). Watch yourself! Giving feedback on recorded presentations in a Virtual learning Environment.

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–18.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric- referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*(2), 3-13.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1-11.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, *48*(1), 12-19.

Assessment of Writing From Instructor and Student Perspectives. *Journal of Educational   Psychology, 98*(4), 891-960.

assessment: Effects on construct validity, performance, and perceptions of fairness and   comfort. *Studies in Educational Evaluation*, *39*(1), 195-203.

Barber, P., & Hill, M. (2014). *Preparing for a renaissance in assessment*. London: Pearson.

Boud, D. 1990. Assessment and the promotion of academic values. *Studies in Higher Education, 15*(1), 1–13.

Boud, D. 2000. Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151–67.

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programs: A review of the issues. *Language Testing*, *15*(1), 45–85.

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, *18*(4), 393–407.

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. New York: Routledge.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and Reliability of Scaffolded Peer

Cooper, D. (2005, July). Assessing what we have taught: The challenges faced with the assessment of oral presentation skills. In *paper from the Higher Education Research and Development Society of Australasia Conference*.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*(3), 393-415.

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(1)*,* 129–142. doi: 10.1177/1469787412441284 .

Diab, N. M. (2016). A comparison of peer, teacher and self-feedback on the reduction of language errors in student essays. *System*, *57*(1), 55-65.

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System, 36*(1), 506-516.

Dochy, F., Gijbels, D., & Segers, M. (2006). Learning and the emerging new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends*. Oxford, Amsterdam: Elsevier.

Falchikov, N. (2003). Involving students in assessment. *Psychology Learning and Teaching, 3*(2), 102–108.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Oxon, UK: Routledge.

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.

Gessa Perera, A. (2011). Co-assessment as a complementary alternative to learning assessment. Analysis and reflection in university classrooms. *Revista de Educación*, (354), 749-764.

Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education*, *13*(2), 129-142.

Harris, L. R., & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101-111.

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.

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*(4), 344-348.

Kwan, K. P., & Leung, R. W. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, *21*(3), 205-214.

Lamprianou, I., and Athanasou, J. (2009). *A teacher's guide to educational assessment*. Revised edition. Rotterdam: Sense.

Lanning, S. K., Brickhouse, T. H., Gunsolley, J. C., Ranson, S. L., & Willett, R. M. (2011). Communication skills instruction: An analysis of self, peer-group, student instructors and faculty assessment. *Patient Education and Counseling, 83*(1), 145–151. doi: 10.1016/jpec.2010.06.024 .

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467.

Magin, D. and Helmore, P. 2001. Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education*, *26*(1) 287–298.

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Crows Nest. New South Wales: Allen & Unwin.

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, *39*(3), 125-132.

Panadero, E., Romero, M., & Strijbos, J. (2013). The impact of a rubric and friendship on peer

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, *9*(3), 233-265.

Popham, J. W. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership, 55*(2), 72–75.

Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language Testing*, *25*(1), 127-143.

Quinlan, A. (2006). *Assessment made easy: Scoring rubrics for teachers from K-college.* Lanham, MD: Rowman and Littlefield Education.

Race, P. (2001). *The lecturer's toolkit. A practical guide to learning, teaching and assessment.* London: Kogan Page.

Rea-Dickins, P. & Gardner, S. (2000). Snares and silver bullets: Disentangling the construct of formative assessment. *Language Testing*, *(1)*17, 215–243.

Rian, J. P., Hinkelman, D., & Cotter, M. (2015). Self, peer, and teacher assessments of student presentation videos. In *JALT 2014 conference proceedings* (pp. 688-696).

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31.

Stiggins, R. 2001. *Student-involved classroom assessment,* (3rd ed.), Upper Saddle River, NJ: Merrill Prentice Hall.

Suñol, J. J., Arbat, G., Pujol, J., Feliu, L., Fraguell, R. M., & Planas-Lladó, A. (2016). Peer and self-assessment applied to oral presentations from a multidisciplinary perspective. *Assessment & Evaluation in Higher Education*, *41*(4), 622-637.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276.

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.

Appendix

An Evaluation Rubric

Presenter's Name…………………

Topic   …………………………………………

Your Name ……………………………………………………………

Assessing your peers and probably yourself is not an easy task. You need to try to be fair and objective. Use the following scale when assessing your fellow students and yourself.

1 = poor 2 = below average 3 = average 4 = above average 5 = excellent

	Criteria	**Comments**	**Scores**
Content	Structure of the Presentation		
	Evidence of Rehearsal		
	Pronunciation & Clarity of Expression		
	Intonation		
	Originality (e.g., not copied from internet)		
	Appropriate/Accurate Use of Grammar and Vocabulary		
	Quality of the Content (Was it informative?)		
Additional Criteria			

Delivery & Presentation Skills	Attention Engaging		
	Visual Aids		
	Interaction with the Audience (e.g., asking questions)		
	Confidence		
	Eye Contact/Voice/Gestures/ Movements		
	Timing & Pacing		

Additional Comments: