

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/268505766>

The Impact of Using Scoring Rubrics in Peer Assessment on Improving Iraqi University EFL Learners' Speaking Skill

Article · January 2013

DOI: 10.13140/2.1.4141.0241

CITATIONS

0

READS

81

2 authors, including:



Zainab Jaafar Auda

University of Basrah

23 PUBLICATIONS 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Improving the Linguistic Capacity of MSC Researchers [View project](#)



An English Course to the Students of Marine Sciences [View project](#)

The Impact of Using Scoring Rubrics in Peer Assessment on Improving Iraqi University EFL Learners' Speaking Skill

Zainab Jaafar Auda

Prof. Balqis I.G. Rashid

Department Of English Language, College of Education,
University Of Basrah

Abstract

In the mid 1990s, the field of assessment and evaluation witnessed a shift from traditional to alternative assessment. The shift presented a number of new practices aiming at linking teaching and learning processes with assessment process. Previous studies investigating the impact of scoring rubrics and peer assessment proved them to have a positive impact on learners, teachers, and curriculum. The present paper¹ investigates the impacts of rubrics and peer assessment on Iraqi university EFL students' speaking performance in conversation classes. A pretest-posttest technique is used between which the two practices were applied on EFL learners in the Department of English Language, College of Education, University of Basrah. A paired t-test analysis was applied showing a positive impact of the rubrics and peer assessment on the performance of the participants. Moreover, the alternative practices helped creating an interactive student-centered classroom with highly motivated students.

1. Introduction

In the last two decades, the field of assessment and evaluation witnessed a remarkable progress moving it from the so-called traditional era to the alternative (or authentic) one. Following that, new practices and techniques flourished in the field, aiming at improving the learning and teaching processes, let alone the assessment process. Applying the new assessment practices inside classroom, teachers' role became more cooperative than being controlling in the classroom, and students started to take some responsibility for their learning. The learners' passive roles as being "novices" merely receiving knowledge from "experts" (i.e.

¹ This paper is based on the first researcher's MA thesis that is supervised by the second researcher. The first researcher is indebted to her supervisor with great thanks and gratitude for guiding her along the study with clear instructions and encouragement.

teachers) have been changed into more interactive ones (Anderson, 1998: 8).

Among the new practices of alternative assessment, being focused upon in this paper, are instructional rubrics and peer assessment. The study concentrates on their use in the assessment of EFL students' speaking performance in conversation classes. Beside being assessment tools, rubrics and peer assessment are found to be effective instructional tools as well. Previous works (Topping, 1998; Andrade, 2000; Wheeler et al, 2005; Jonsson and Svingby, 2007; Naksuhara, 2007; White, 2009; Reddy and Andrade, 2010; Raza, 2011) found these practices to be successful in improving both the learning and teaching processes.

The hypothesis being investigated here is that using instructional rubrics in peer assessment is effective in improving students' speaking performance and motivation and in creating a student-centered classroom in conversation classes. The paper starts by introducing basic aspects concerning the shift from traditional to alternative assessment, integration between assessment and instruction, and basic remarks on the two practices of rubrics and peer assessment.

2. Traditional Assessment and Alternative Assessment

Alternative assessment consists of any method of examining what students know or can do that is intended to show growth and inform instruction. It is an alternative to traditional forms of testing, namely multiple-choice test (Stiggins, 1991, cited in O'Malley and Pierce, 1996:1). The term "alternative assessment" is usually used by researchers as interchangeable with the term "authentic assessment". O'Malley and Pierce (1996: 1-2) propose that "[a]lternative assessment is by definition criterion-referenced and is typically authentic because it is based on activities that represent classroom and life-long settings."

In the mid- 1990s, a shift from traditional assessment to alternative assessment started in the United States' public k-12 schools as a substitute for the standardized tests. There, teachers were encouraged to use the practices of alternative assessment to evaluate the real learning of the children in authentic situations (Lombardi, 2008:4).

The need to shift from traditional assessment towards alternative assessment is being highly appreciated in educational development. By connecting teaching, learning, and assessment, alternative assessment is seen to promote students' learning and help teachers to be "fair, thoughtful, and creative when assessing students' work" Anderson (1998: 13). In this respect, Lombardi (2008:3) states that such a shift occurred due to several factors, among which he lists "economic conditions, new

scholarship on learning, and a student population with new expectations of educational institutions." The main differences between traditional assessment and alternative assessment are illustrated by Lombardi (2008:5) in Table (1).

Table (1) Traditional vs. Authentic Assessment

No	Traditional Assessment	Authentic Assessment
1.	Generally relies on forced-choice, written measures	Promotes integration of various written and performance measures
2.	Relies on proxy measures of student learning to represent target skills	Relies on direct measures of target skills
3.	Encourage memorization of correct answers	Encourage divergent thinking in generating possible answers
4.	Goal is to measure acquisition of knowledge	Goal is to enhance development of meaningful skills
5.	Curriculum direct assessment	Assessment directs curriculum
6.	Emphasis on developing a body of language	Emphasis on ensuring proficiency at real-world tasks
7.	Promotes "what" knowledge	Promotes "how" knowledge
8.	Provides a one-time snapshot of student understanding	Provides an examination of learning over time
9.	Emphasize competition	Emphasize cooperation
10.	Targets simplistic skills or tasks in a concrete, singular fashion	Prepares students for ambiguous and exceptions that are found in realistic problem settings
11.	Priority on summative outcomes or product	Priority on learning sequence or process

From Table (1) given above, it can be noticed that authentic assessment does not focus merely on assessment goals, instead it covers all aspects of learning, teaching, curricula design, and assessment. Hence, it focuses upon promoting students' learning, thinking and behaviour, refining curricula, and achieving assessment, which is more objective. This is unlike the emphasis of the traditional assessment on final summative results that are of little assistance to each of the learning, teaching, and assessment processes.

3. Alternative Assessment as an Integral Part of Instruction

Unlike traditional assessment, the practices of alternative assessment assure the benefit of giving feedback that extends along the learning process. Stiggins and Chappuis (2005: 12) assert that, in order to be functional, feedback should be continuous in the classroom and not solely a result of a final course test. This does not mean that the learning process

should abandon final assessment for testing purposes. On the contrary, if tests and exercises are set clearly to meet the course goals, they will give considerable information of students' levels (Black and Wiliam, 1998: 8). Thus, instructors should be aware of the importance of linking feedback with instruction and redirecting their instruction to its specified aim. If they are aware of the goals of their teaching process, "[i]nstead of "teaching to the test," teachers are more accurately "testing what they teach." (Guskey, 2003: 7) Students, on the other hand, should have the potential of feedback to identify their weaknesses and strengths and make more efforts to meet the goals of the learning process.

By evidence, studies found that when the practices of classroom assessment are integrated with the instruction/learning process, the results for both teachers and students would be positive (Black & Wiliam, 1998; Meisels, Atkins- Burnett, Xue, & Bickel, 2003; Newman, Bryk, & Nagaoka, 2001; Rodriguez, 2004, cited in Brookhart et al 2009: 53). So, enhancing the students' achievements and reducing the score gaps, seem to be within the reach of teachers (as suggested by Stiggins and Chappuis, 2005: 14), if they

1. Focus on clear purposes,
2. Provide accurate reflections of achievement,
3. Provide students with continuous access to descriptive feedback on improvement in their work (versus infrequent judgement feedback), and
4. Bring students into the classroom assessment processes.

4. Alternative Assessment in Higher Education

Though started its application in kindergarten, authentic assessment gained wide acceptance among educators and students in higher education. Different practices were used since the evolvement of authentic assessment that proved special success in enhancing students' learning in different fields of knowledge. What is more important to note is that involving students in assessment practices highly raises their motivation, thinking and achievement (Vu & A'lba, 2008: 7). In their exploration for an authentic approach to assessment to enhance student learning, Vu & A'lba (2008: 7-8) highlight five points that make successful the use of authentic assessment in higher education. In brief, the points are summarized as follows:

- 1 Involving students in assessment practices can provide them with "opportunities to synthesise and demonstrate what it means to become skilful professionals." (This point, in particular, is highly appreciated in the current study, because the assessment practices involved will be applied on students who are prepared for future professions of instruction, in simple words, to be teachers of English.)

- 2 Experiencing assessment tasks "can also provide space for students to challenge outdated ideas, routinised practices, and their own as well as public assumptions." Such practices can expand students' understanding and awareness of their future profession.
- 3 "Assessment should be integrated with learning tasks", so that students can direct their efforts towards the expected learning goals.
- 4 Assessment tasks ensure interactions between students and teachers that may allow a "timely relevant feedback" to be used as a basis for guiding the learning process.
- 5 It is crucial that teachers explain the "objectives, procedures and outcomes" of assessment clearly in advance in order to get the most of the assessment and learning processes.

5. Practices of Alternative Assessment

Different kinds of alternative assessment practices are wide spread and applied in numerous fields of knowledge. Lombardi, (2008: 6) introduces four basic practices. They are:

1. **Rubric:** a set of scoring guidelines that is disclosed to students—or, in many cases, actively developed in collaboration with students. A good rubric identifies the criteria by which a work is judged and describes the difference between excellent and weak work.
2. **Peer Assessments:** Topping (2012: 3) defines peer assessment as "an arrangement for peers to consider the level, value or worth of the products or outcomes of learning of their equal-status peers."
3. **Research Portfolio:** "... an appropriate mechanism for monitoring student progress on extended, multifaceted projects requiring higher-order thinking skills." (Lombardi, 2008: 7-8)
4. **Group work:** "Assignments that involve significant group work often come closer to the dynamics of real-world practice than those challenge students to work on projects independently." (Lombardi, 2008: 8)

6. Rubrics

6.1. Definition of "Rubric"

As the practices of alternative assessment developed widely in the 1990s, the word rubric started to gain its popularity among the educational mediums. However, looking for the term 'rubric' in dictionaries does not reveal the intended meaning of the word. Instead, they show the historical meaning of the term as being associated with the colour red in the Christian scripts (Stevens and Levi, 2005: 3).

In the field of assessment and evaluation, the term rubric refers to a rating scale that helps describing the students' levels of a targeted skill (Popham, 1997: 2; Brualdi 1998: 2). However, in addition to its main task as an assessment tool, a well-constructed rubric can be a teaching tool that helps improving students' skills as well as an assessment tool (Andrade, 2001: 1). In this regard, Andrade (2001 and 2005) differentiates between scoring rubrics and instructional rubrics. While scoring rubrics are those used by teachers merely to assign grades of students' performance, instructional rubrics are those shared with students for self- and/or- peer assessment, giving and receiving feedback, and finally assigning grades (Andrade, 2005: 29). Since they bridge the gap between instruction and assessment, instructional rubrics are highly encouraged for better educational results (Andrade, 2000:1).

6.2. Composition of Rubrics

Teaching with rubrics needs an experience and knowledge on how to create and use them appropriately. What makes a "good" rubric is the ability of a teacher to construct the suitable rubric that meets his and his students' needs at the first place, and to know how to use it appropriately. Thus, these two basic conditions, good construction and way of use, affect and decide the value of any rubric (Andrade, 2005:27).

The composition of any rubric includes two basic features: **evaluative criteria** and **quality definition** (Popham, 1997:72; Andrade, 2000:1). Along with those, Popham (1997: 72) adds a third feature- **scoring strategy**- that defines the type of the rubric (holistic or analytic) according to the scoring of the criteria.

- a. **Evaluative Criteria:** a list of criteria, or "what counts" (Andrade, 2000:1) in a task. Teachers and researchers find it always useful to engage students in defining those criteria and creating the rubrics (Brualdi, 1998: 2; Scott: 41: 2006).

The following points are suggested by the Rubric Design Guide 'RDG' (RDG, 2007:2) to be followed when identifying the grading criteria to be included in a rubric:

1. Determine the learning outcomes for the assignment by asking the following questions:
 - a. What is the intended learning that is to occur?
 - b. How can such learning be measured?
 - c. Are there any given conditions that should be considered for each outcome?

2. Create a separate item in the grading rubric for each learning outcome.
 3. Determine the importance for each of the grading criteria.
 4. Communicate such criteria to the students prior to the completion of the task so that they know what is expected of them.
- b. **Quality Definition (or "grading quality"):** it describes the qualitative levels of students' performance that are to be judged (Popham, 1997: 72). They are also called "**descriptors**" because they describe what should be achieved at each level of importance of the criteria (RDG, 2007:3).
Table (2) illustrates the basic format of an analytic rubric.

Table (2) Basic format of an analytic rubric

Title of the rubric			
<i>Description of the task being assessed</i>			
List of criteria	Criterion Level 1	Criterion Level 2	Criterion Level 3
Criterion 1	<i>Level description</i>	<i>Level description</i>	<i>Level description</i>
Criterion 2	<i>Level description</i>	<i>Level description</i>	<i>Level description</i>
Criterion 3	<i>Level description</i>	<i>Level description</i>	<i>Level description</i>

It must be mentioned that Table (2) given above is only an illustration to the basic format and not a static design of a rubric. Accordingly, each of the levels and criteria dimensions can vary according to the teachers' perspective of the students' quality of performance in their classes. The levels should simulate true performances of the students being assessed, and the quality definition should be clear enough to be understood. However, the levels are not preferred to exceed five columns. As far as the criteria are concerned, they can vary in number according to the skill and task being assessed. In simple words, teachers can tell their students their expectations and what they want them to focus upon through the criteria dimension.

When setting a rubric, teachers find that the hard part in composing a rubric is the quality definition. It should be made clear enough to be understood by all raters (including students when used in self- or peer assessment).

To define the quality definitions, RDG suggests the following:

1. Assign the descriptors to each level:
 - Describe the best work.
 - Describe the worst work.
 - Describe the levels in-between.
2. Determine the scoring scale (qualitative, quantitative, or both):
 - Qualitative: a scale of weak, satisfactory, strong.
 - Quantitative: a scale of 1-5.
 - Both: a scale of 1-4 or beginning, developing, accomplished, and exemplary.

6.3. Types of Rubric

Rubrics are of two types: **holistic** and **analytic**. In a holistic rubric, the criteria being evaluated are considered totally in combination and an overall judgment is made on a single descriptive scale (Popham, 1997: 72; Moskal, 2000: 4; Scott, 2006: 41). Whereas an analytic rubric focuses on "more specific aspects of performance" (Scott, 2006: 41) and requires the scorer to render criterion-by-criterion scores..." (Popham, 1997: 72). In other words, the skill being assessed is analyzed into its constructed criteria for each of which there are several levels of proficiency included from the least to the most.

6.4. Advantages and Disadvantages of Rubrics

Rubrics are not pure rating scales. Since they can merge instruction with assessment, teachers are encouraged not to limit their effectiveness by taking them simply as mere scoring guides. However, rubrics still have their pros and cons.

Andrade (2005: 27-30) discussed all the aspects of rubrics, "the good", "the bad" and "the ugly" ones. They are summarized below:

A. The Good: instructional rubrics are good for both teachers and students. For teachers, they are good because they:

1. *Observe and make clear the teaching goals.*
2. *Help in designing the teaching methods that meet the teaching goals.*
3. *Communicate the goals to students.*
4. *Supervise students' feedback (weaknesses and strengths) and progress in little time.*
5. *Grade students' final performance according to whether they have met the goals or not.*
6. *Keep teachers "fair and unbiased" in their grading.*

For students, instructional rubrics are good because they:

1. *Tell them their teachers' goals, and so they focus their efforts towards meeting those goals. Applying that, Andrade (2005: 29) states, "I never hear a student complain that she 'didn't know what I want'".*
2. *If used in self-and-peer assessment, help students to get varieties of feedback that work as resources of "insight and help instead of ... reward and punishment" (Shepard, 2000:10, cited in Andrade, 2005: 29).*

B. The Bad: Rubrics may be **bad**, because:

1. *"They are not self-explanatory", so still teachers need to explain to students how to use them.*
2. *They are not "replacement for good instruction"*
3. *At their first experience with peer and self-assessment, "students are not always good" and need training and observation.*

C. The Ugly: In their worst descriptions, rubrics may be **ugly** because they are still issued to aspects of validity, reliability, and fairness.

1. **Validity:** *a valid rubric is that which goes along with "reasonable and respectable standards and with the curriculum being taught".*
2. **Reliability:** *a rubric is reliable when it is used by different raters reaching similar judgments.*
3. **Fairness:** *issues of "gender, race, ethnicity, or socioeconomic status" might affect the use of rubrics.*

For the sake of improving a rubric, Andrade suggests to compare it with "published standards", consult another teacher, or ask a colleague to co-assess the performance of the same students. In addition to that, consulting the students is yet another good way as well. This may happen by having a time to work with students prior to applying the rubric. Here, Andrade agrees with what is stated by Payne (2003, cited in Andrade, 2005: 30): "sitting and listening to students' critique assessments can be the best source of information about how good evaluations really are."

7. Peer Assessment

Peer assessment is considered as a principal practice in formative assessment (White, 2009: 3). As defined by Topping, peer assessment is "an arrangement of peers to consider the amount, level, value, worth, quality of successfulness of the products or outcomes of learning of others of similar status." (Topping, 1998: 250; Topping, 2012: 3) By

"similar status", Topping means students "usually in the same course and often in the same year" (Topping, 1998: 250). For formative assessment to be more productive in raising students' confidence and motivation, students' involvement in assessment is highly encouraged.

For the purpose of filling the gap in literature of peer assessment in higher education, Topping (1998) conducted a review surveying peer assessment between students in college and university in the period 1980-1996. He included 109 papers focusing upon peer assessment among students in higher education. According to the results arrived at by Topping, it was found that peer assessment is highly recommended in raising the learning performance of students. Researchers encourage involving students in such experiences and argue that teachers should help their students to "see, understand, contribute to, and appreciate their own journey of achievement success ... rather than [to be] victimized by, the assessment process." (Stiggins and Chappuis, 2005: 13). Such an idea is highlighted by Caban (2003: 5) in stating that "[r]ather than view themselves as unwilling victims of the test, learners should see themselves as active participants who have responsibility for their own learning outcomes."

7.1. Advantages and Disadvantages of Peer Assessment

The implementation of peer assessment in classrooms has several advantages as far as students are concerned and the learning process in general. In addition to their educational outcomes, peer assessment proved to have behavioural gains for students as well. In that sense, Topping (2012: 3) asserts that peer assessment has immediate advantages "in terms of learning and achievement", long term advantages "in terms of transferable skills in communication and collaboration, which will be in demand in later life", and some additional advantages "in terms of the self-regulation of one's own learning." As far as classroom is concerned, peer assessment is highly supported in building interaction inside classroom. Wheater et al (2005: 13) suggest that peer assessment implemented in classrooms make more interactive classes and help in a better understanding of materials.

In spite of the advantages of peer assessment and evidence that it can be effective in learning, there are several disadvantages that are inevitable in the process. Among them, Wheater et al (2005: 13) cite problems like teachers managing inexperienced assessors and use of valuable class time and other issues of validity and reliability. Table (3), which is adopted from White (2009: 5), illustrates potential advantages and disadvantages of peer assessment.

Table (3). *Potential advantages and disadvantages of peer assessment*

Advantages
<ol style="list-style-type: none"> 1. Helps students to become more autonomous, responsible and involved. 2. Encourages students to critically analyze work done by others, rather than simply seeing a mark. 3. Helps clarify assessment criteria. 4. Gives students a wider range of feedback. 5. More closely parallels possible career situations where judgment is made by a group. 6. Reduces the marking load on the lecturer. 7. Several groups can be run at once as not all groups require the lecturer's Presence.
Disadvantages
<ol style="list-style-type: none"> 1. Students may lack the ability to evaluate each other. 2. Students may not take it seriously, allowing friendships, entertainment value, et cetera, to influence their marking. 3. Students may not like peer marking because of the possibility of being discriminated against, being misunderstood, et cetera. 4. Without lecturer intervention, students may misinform each other (Peer Assessment, 2007, University of Technology Sydney).

Apart from its disadvantages, peer assessment is still worth application in classrooms since it proved enhancing learners' achievements that is a typical goal of any learning process. The following section show them.

8. Literature Review

A considerable amount of literature has been published on rubrics use and self-and/or- peer assessment. In this concern, an online search was made through some main educational libraries and databases including IVSL (Iraqi Virtual Scientific Library) that opens the gate to several significant databases like Science Direct, JSTOR, Project MUSE, Springer, and Citeseer, and the world largest digital library, ERIC (Education Resources Information Center). The search revealed a vast number of works on rubrics and self and/or peer assessment distributed in different disciplines including teachers education, computer science and technology, medical sciences, nursing, arts, assessments and evaluation, psychology, business, engineering, and others. In addition to that, it was found that the use of rubrics and self and/or peer assessment was applied in almost all levels of education, starting from kindergarten to higher education.

While all studies focused on the use of rubrics, the purpose behind them varied. Some studies investigated the effects of rubrics and peer

assessment on students' performance, others focused on students and/or teachers' perception of the use of rubrics and self and/or peer assessment, and some others examined the issues of validity and reliability in rubrics use and self and/or peer assessments. Three reviews and three studies are of interest to this study, they are mentioned here. The reviews are introduced firstly since they cover a number of the studies appeared in the online search.

The first review is done by Topping in 1998. As introduced earlier, Topping (1998) conducted a review covering 109 studies on peer assessment in different subject areas. In conclusion to his review, here is a summary of Topping's findings (Topping, 1998: 267-8):

1. There is positive impact of peers' feedback on students' grades improvement.
2. Reliability and validity of peer assessment is achieved in various subjects' areas.
3. Peer assessment does not gain acceptability among all students.
4. Although peer assessment requires efforts from students, it reduces their anxiety.
5. Peer assessment frequently improves students' performance and confidence.
6. Peer assessment of writing proved to be similar to (and in some cases better than) teachers' assessment.
7. Studies on peer assessment of group and project work gained positive results.

The second review is of Jonsson and Svingby (2007). A comprehensive review exploring the validity and reliability of rubrics and their impact on students' learning and instruction's improvement was done by Jonsson and Svingby in 2007. The review covered 75 studies in various disciplines and was distributed in all the levels of education, from kindergarten to higher education. Among the works reviewed, the review covered studies concentrating on whether rubrics can promote students' learning, self and peer assessment, students' perspectives towards the use of rubrics, and whether rubrics improve instruction. In conclusion, the following points are arrived at ([Jonsson and Svingby, 2007: 141](#)):

1. The reliable scoring of performance assessments can be enhanced by the use of rubrics, especially if they are analytic, topic-specific, and complemented with exemplars and/or rater training.

2. Rubrics do not facilitate valid judgment of performance assessments per se. However, valid assessment could be facilitated by using a more comprehensive framework of validity when validating the rubric.
3. Rubrics seem to have the potential of promoting learning and/or improve instruction. The main reason for this potential lies in the fact that rubrics make expectations and criteria explicit, which also facilitates [SIC] feedback and self-assessment.

The third review is of Reddy and Andrade (2010). Reddy and Andrade conducted a review on rubric's use in higher education including 20 empirical studies and doctoral theses in 2009. The published studies showed that rubrics have a wide use in different disciplines in higher education including "the liberal arts, information literacy, medicine, nursing, management, dentistry, food technology, teacher education, and film technology" (Reddy and Andrade, 2010: 437). Among the questions being investigated in the review is a question of whether rubrics can promote learning rather than merely assess it was included.

In answering the question on whether rubrics can be effective formative assessments used for promoting students attainment, Reddy and Andrade declare that the results were "inconclusive" due to little quality research conducted in this concern. Hence, while two studies (Petkov and Petkova, 2006; and Reitmeier, Svendsen, and Vrch, 2004, cited in Reddy and Andrade, 2010: 445) found out that the use of rubrics has reflected academic importance in students' achievement, one study (Green and Bowser, 2006, cited in Reddy and Andrade, 2010: 445) showed no difference between students' performance before and after the use of rubrics. Yet, Reddy and Andrade suggest that teachers should not only hand rubrics to students to get good results. Instead, they should teach them how to use them actively in self and peer assessment and in revising their works. Moreover, Reddy and Andrade accuse teachers of using rubrics as mere "scoring guides" and encourage using them as instruction guides to be shared with students in self and peer assessment to give best results in improving their products, with this point Reddy and Andrade show their full agreement.

Concerning the studies investigating the use of rubric and peer assessment in the assessment of speaking performance, the following are found. Firstly, in 2007, Naksuhara carried out a study on the use of rubrics in assessing speaking proficiency in Japan. Prior to the study, he

reviewed the available rating scales and rubrics, examined the marking categories and descriptors, and drafted a new scale 'an analytic rubric' including five criteria (pronunciation and intonation, grammar, vocabulary, fluency, and interaction communication). The researcher applied the rubric in a pilot study on upper-secondary students ($N= 42$) and two raters. Naksuhara tested the students (and video- recorded them) while participating in group discussion tasks (including information gab, ranking, and free discussion). Each test extended for 15-20 minutes, and tapes were gathered for a later rating with the rubric sheets. One upper secondary teacher, along with the researcher, assessed the students' performance after having an hour of discussion for explaining the rubric's criteria.

Multifaceted Rasch analysis was performed to examine the examinees' performance, rating and rating categories, and all levels of the rubric criteria. After that, Naksuhara found out that the rubric was a good indicator for students' speaking performance; it achieved reliability between the two raters, and its descriptors were clear. Nevertheless, Naksuhara suggested the use of the rubric with a greater number of raters to get a more effective test of reliability.

Secondly, following Naksuhara (2007), Raza (2011) carried out a study on the use of rubrics in assessing speaking proficiency in Pakistan. Starting his research, he surveyed the available rating scales in Pakistan and worldwide and drafted a new scale, 'a rubric' to apply it on university students ($N= 44$) and eight raters. Raza set to the students two tasks, the first was to be interviewed by their teachers for 15 minutes and the second was to let them have pair discussions each for 20 minutes. All the interactions were audio-recorded and handed to eight raters (each group of tasks was given to four raters) accompanied by the analytic rubric designed by Raza that included five criteria (interaction communication, fluency, grammar, pronunciation, and vocabulary). Receiving the ratings, Raza applied the FACET analysis to examine the examinees' speaking performance, rating severity, and item difficulty. His findings revealed that the rubric was a suitable measure for students' English abilities; however, he calls for raters training to get more reliability in assessment.

Thirdly, in 2009, White conducted an empirical study examining the students' perception of peer assessment and its impact on their learning. White applied peer assessment in a public speaking course (14 weeks) with third year EFL female students ($N= 55$) in Tokyo university. The peer assessment sheet was based on a previous study including

numeric values for the five levels of each criterion with no descriptors (i.e. 5, 4, 3, 2, 1 from best level to the least). The criteria being examined were voice control, body language, contents of presentation, effectiveness, and visuals).

Students had to present a topic, taken from news, in front of their peers in a minimum presentation that lasts for 2-3 minutes, using computer slideshows. At the end of each presentation, peers start to assess the performance and give feedback to the presenters. This feedback, White proposes, enabled students to figure out the best criteria of public speaking. In conclusion, White found that peer assessment has truly led to the promotion of students' learning.

9. Study Methodology

9.1. Context of the Study

The study was conducted in the academic year 2011-2012 at the department of English, College of Education, University of Basrah. Conversation classes are selected among the other classes for the application of the study because they are the most suitable classes at which students can practice their speaking skill regardless of any difficulty of the subject materials.

9.2. The Pilot Study

Prior to the main experimental work of the study, the researchers needed to conduct a pilot study focusing on the application of the analytic rubric inside the classroom of conversation classes with a sample of the target participants. This was based on an advice to the researcher by Dr. Heidi G. Andrade, the assistant professor at the University of Albany/New York. In fact, such a study was of primary significance and benefit to the researchers in getting training in the use of rubrics, how peer assessment works and techniques of teaching conversation, in addition to having a close view of students' performance in classroom.

9.2.1. Participants and Setting of the Pilot Study

The pilot study was conducted in the first semester of the year of the study. The participants were one group of the third stage students (group B consisting of 25 females). The first researcher instructed the students along ten lectures and video-recorded all the lectures (after getting the students' permission and promising them not to share the videos publically). Prior to the pilot lectures, the researcher attended three lectures as an observer of the students' speaking proficiency and interaction inside classroom. She found that the interview technique was

applied inside the classroom, and the participants in the three lectures were frequently the same few students, while the majority were passively attending the lectures.

9.2.2. The Rubric of the Pilot Study

According to the students' performance observed in the three lectures and the first video of the lectures she instructed, the first researcher designed a rubric concentrating on the linguistic criteria of the speaking skill. They were five criteria, namely: fluency, grammar, pronunciation, vocabulary, and comprehension, each with four levels of proficiency. Two grades were given to each criterion, making the sum of 20 for the total rubric. After explaining the dimensions of the rubric to the students in the second lecture, she explained the procedure she will follow with them.

9.2.3. The Peer Assessment

The interview technique was followed along the pilot study for which the first researcher created handouts in different topics (advertisements, gossip, sports, fashion, shopping, technology, and liberty). Peer assessment was introduced in the fourth lecture and students were asked to peer assess their mates starting from the fifth lecture and hand the rubrics back to the researcher at the end of each lecture. Having the peers' assessed rubrics; the researcher graded them and gave them as feedback to the assessed ones in the next lecture. The students who were not assessed by their peers were given zero and the same is for each unassessed criteria. In addition to that, the researcher used the same rubric in assessing the students' daily performances.

9.2.4. Results of the Pilot Study

The pilot study was successful in motivating passive students to participate in the lectures' discussions (for the sake of being peer assessed). However, due to time limit, active learners could not have the chance to participate in discussions for more than one time in order not to take others' chances of participation. Thus, the interview technique was proved unsuccessful with peer assessment in the limited 45 minutes of class and the large number of 25 students. Following that, the researcher made some discussions with several researchers and teachers (via emails, Academia.com and LinkedIn.com) about how to make the best of rubrics and peer assessment with such a large class. In conclusion, group work was important to apply and more authentic tasks to practice in the classroom.

9.3. The Empirical Study

9.3.1. The Participants

The participants were all the students of the third stage in the department of English, College of education. The total number of the students was seventy-four, sixty-eight females and six males. Since the number of males was very few in comparison to the number of females, the gender factor has not been taken into account in this study.

The participants were meant to be a representative sample of the target population, i.e. Iraqi university EFL learners. However, females' number always dominates the number of males significantly in the department of English language, College of Education. Thus, though may not be an adequate representation concerning the gender factor, the participants represent the convenience sample of the targeted population.

9.3.2. The Setting

The study was conducted in the second semester of the academic year 2011- 2012 and extended along two months. Conversation classes were given twice a week for each of the three groups of the third stage (A, B, and C). Therefore, the total number of the lectures of the study was 12 for each group and the first researcher had to attend six lectures weekly to cover the whole stage as she lectured all the stage by herself.

9.3.3. The Procedure

Seeing that students were passive in conversation classes, the goal of the teaching process was not concentrating on the linguistic criteria of the speaking skill per se, but to enhancing the motivation inside the classroom and creating a student-centered class. The group work technique, which is one of the alternative assessment's practices, was followed in all the lectures to help achieving that goal, especially for it gives opportunity of involving all the students in the limited time of class. Primarily, each group in the stage (A, B, & C) were divided into four subgroups inside the classroom, each consisting of about 6-7 students (AA, AB, AC, AD), (BA, BB, BC, BD) and (CA, CB, CC, CD).

9.3.4. The Pre-test

In the first lecture, the first researcher clarified to the students the study procedure they will be enrolled in. She explained the group-work technique according to which they will have to make their subgroups and the peer assessment strategy. The students chose their subgroups and peers. In each lecture, they have to nominate a new discussion leader for starting the topic, managing the discussion, and giving a conclusion. The

discussions extended only for ten minutes through which the researcher observed the subgroups and helped them whenever there is a breakdown in interaction or a need to interrupt. At the end of the first lecture, she asked the students to be prepared for the pretest that will be held in the second lecture. The students had to choose their topics and nominate the leaders. She also informed them that the pretest will be video-recorded and explained to them why it is needed. Besides, she comforted them that the videos will not be shared and kept only for the study purposes.

In the days of the pre-tests, the students were asked to start their discussions with the topics they have prepared. One subgroup after another took their time of ten minutes of discussion while the researcher video-recorded them without participating in the discussion.

Having got those videos helped the researcher to recognize the students' capabilities in conversation. Some of the students were active enough while others were not willing to participate at all, and there were those in between. Moreover, the repetition of the recordings helped to establish a basis for the students' performance and developed the suitable rubric. After all, the pretest videos were used by the researcher in assessing their proficiency levels according to the rubric that was designed and shared with the students in the empirical study.

9.3.5. The Pre-test Videos as Sample of the Students' Performance

Giving the students authentic products of their own helped in enhancing their awareness of their strengths and weaknesses. For this reason, the researcher presented the students their pre-test videos, in the third lecture, and asked them to criticize their own performance. Her primary goal was to help the students understand the criteria descriptors of the rubric that should be associated with authentic samples of their own performance.

9.3.6. Teaching Technique and Lecture Format

Being one of the practices of the alternative assessment, the group technique was uniquely helpful in achieving the new goals. As introduced before, the students freely chose their sub-groupings. Therefore, the researcher asked them to keep those sub-groupings along the period of her instruction. However, some changes were made whenever there were absentees in the group. The main condition was to keep the sub-group in an even number (for the purpose of the peer assessment).

All the lectures had the same format. At the first 10-15 minutes, the first researcher introduces to the students some conversational strategies

that aimed at helping them manage the subgroup discussions and improve their fluency. In each lecture, a list of strategies is written on the board including statements of how to start and end a discussion, keeping the floor, agreeing and disagreeing, asking for and giving opinions, etc. In the next 5 minutes, the researcher introduces the task that the students have to discuss within their sub-groups. The discussions were given only 10 or 15 minutes according to the type of the task. The researcher used a countdown watch to help the students, especially sub-groups leaders, to get accustomed to time management.

The researcher did not participate in these discussions and her role was primarily a facilitator and an advisor when a discussion breaks down, and an active listener and observer of students' performance. She used to circulate around the class while discussions were taking place and help students maintain the discussion whenever she sees a group finish so early by supporting them with further ideas and questions. Additionally, she used to assist the leaders who seemed to be confused at the beginning of their sub-group management by encouraging them and motivating their mates to help them in their missions. At the end of the sub-group discussions, the students peer assess each other within their sub-groups and collect their rubrics with the leaders. The peer assessment takes about 5 minutes to finish.

By the time allotted to the sub-groups discussions ends, the researcher used to ask the sub-groups' leaders to give the summary, the opinions being raised, and the concluding remarks of their discussions. Then, she manages a whole group discussion by asking each sub-group what they thought about the other sub-groups' opinions, findings and/or decisions. This open discussion usually lasts to the end of the lecture and takes between 10 to 15 minutes, according to the time left from the previous activities.

9.3.7. The Classroom Activities

Given that the practices of the alternative assessment encourage the use of tasks-based activities, the researcher used specific tasks along the lectures of the study that aimed at engaging the students in interactive discussions. In all of the tasks, the students have to end the discussions with certain conclusions. Meanwhile, no handouts were shared with the students, but the vocabulary of the tasks were explained, when necessary, before starting the discussions to make sure that all the class share the same understanding of the activities.

The total number of tasks used in the study was six, some were selected from books, like problem solving (Ur, 1996: 127), shopping list (Ur, 1996: 126-127), balloon debate (Hermer, 2001: 273), and ranking (Gammidge, 2004:69), and others were designed by the researcher herself (Technology and journey plan).

9.3.8. The Rubric

Once the goals of the study were changed, the contents and descriptors of the rubric used in the pilot study were modified to meet the new goals. The first researcher had valuable discussions with her supervisor in setting the wordings of the rubric. The help of the supervisor was indispensable in making the final rubric's design. (see Appendix I for the rubric). The basic goal of the empirical study was to help creating an interactive student-centered classroom, that is why the new rubric included criteria like the following in addition to *fluency*, *grammar*, *pronunciation*, and *vocabulary*.

1. ***Interaction***: to direct the students within their sub-group discussions and encourage them for participating in various ways.
2. ***Content***: to encourage the students giving valuable information in regard to the subject matter, and not to take a passive role.
3. ***Humor***: to keep the discussion as formal as possible and not to deviate away from the classroom atmosphere. Since students have the control over their give-and-take, this criterion was important to direct them within the academic discussion.
4. ***Assessment Criteria of the discussion leader (leading criteria and time control)***: to guide and direct the leader through the discussions as well as to keep the allotted time prescribed earlier.

As far as the scoring of the rubric is concerned, each criterion was given two scores except the use of Arabic and the interaction that were given four scores for each. This particular scoring aimed at raising the students' awareness of the use of English language inside the classroom and taking an active part- by all means- in the interaction (since many of them had a problem in talking in English fluently, they code switch to Arabic frequently). Concerning the discussion leaders, the criteria listed in their table gained two scores for each and they are to substitute the interaction criteria listed in the main table. Ultimately, the total score of the rubric for any student was 20 marks. Table (4) illustrates the scoring of the new rubric.

Table (4). Scoring of the new rubric's criteria

Criteria		Superior	Advanced	Intermediate	Novice
Fluency		2	1.5	1	0.5
Grammar		2	1.5	1	0.5
Pronunciation		2	1.5	1	0.5
Vocabulary	No Use of Arabic	4	3	2	1
	Use of English	2	1.5	1	0.5
Interaction*		4	3	2	1
Content	Content details	2	1.5	1	0.5
	Humor	2	1.5	1	0.5
Discussion leader*		2	1.5	1	0.5
		2	1.5	1	0.5
Total score of the rubric		20	* the criteria of the discussion leader are to replace the interaction criterion		

9.3.9. The Peer Assessment

Since peering has not proved by earlier researches to be necessarily static (Topping, 1998: 251), the researcher gave the freedom to the students to choose their peers along the empirical study. Some of them kept their peers while others used to change them throughout the lectures.

Since the technique followed in the classroom was sub-group discussions, then the peer assessment was a basic tool for ensuring students' performance inside the crowded classroom. The teacher in any case could not keep an eye on each-and-every student's performance, thus such an assessment was a bare necessity in assisting the teacher to control and guide the large class. Moreover, to guarantee the objectivity of the rubric, the researcher explained to the students that the words 'some' and 'hardly' will be defined as three downwards, while 'many' means more than three. This was the only choice to make sure that the students understand the rubric in the same way, and as a result, they will assess each other- most probably- in the same way.

In the sub-groups, each two students were peer assessing each other, while all the members used to assess the performance of their leaders (only in the leadership table). In the case of odd sub-groups, the leaders are excluded from the peer assessment, but were still assessed by all the members in the table specified for them.

9.3.10. The Post-test

Having finished the lectures with the students, the total number of the lectures in using the rubric with the peer assessment was six. Before those, the first four lectures were allotted to the introductory sessions, the

pre-tests, watching the videos and introducing the practices of the rubric and the peer assessment. With the end of the tenth lecture, the time period of the second semester was over and the students were about to have their second term exams within few days. For this reason, the researcher had to end the experiment and told the students to be prepared for the post-test. She informed them that the post-test will take the same procedure of the pre-test, i.e. a ten-minute discussion within the sub-groups, one after another, and their assessment will be done according to the same rubric used for the peer assessment, so they should keep their performance according to its components.

Once it was the time of the second term exams, the researcher asked the permission of the department and the teachers of conversation of the third stage to have the opportunity of testing the students in the subgroups format and have the company of those teachers in the tests so that they can assess the students performance for the second term results while she assesses them for the study purposes. Getting the approval, she prepared the topics of the discussion (Appendix II) and presented them in advance to the head of the department and the teachers concerned. The students were informed of the day of the post-test and their sub-groups lists were announced according to their usual sub-groupings.

In the day of the post-test, the researcher- accompanied by each group's teacher- started to call for one sub-group after the other, giving each a time of 10 minutes, assessing them directly on the rubric, and video recording them while discussing their topics. Their teachers, on the other hand, used to give the students a score of 20 which was the total mark of the second term exam (and which was the same total score of the rubric).

10. Results and Analysis

The Paired-Samples *T*-Test procedure was used to measure the differences between the students' performance before and after applying the study methodology with the participants' students. *T*-Test is the most common test used in comparing the means of two data and it is found with three types: one sample *t*-test, two-sample *t*-test, and paired *t*-test. In this case, the paired *T*-Test was chosen among the three types of *T*-Test because it is capable of comparing the means of two variables for the same group (like, the pre and posttests of the students). It is believed that Paired *T*-Test "computes the differences between values of the two variables for each case and tests whether the average differs from 0." (Elliott and Woodward, 2007:70).

The statistics part of this study was done with the assistance of Dr. Wesal Fakhri Hassan from Marine Science Centre, University of Basrah, for whom the researcher is indebted with great thanks and appreciation. After calculating the students' final scores according to the rubric, the researcher gained two lists for only 66 students (out of 74). This was due to the absence of eight students either in the pretest or posttest. Accordingly, the statistics applied was done covering only the scores of these 66 students. The following sections detail the results of the analysis.

1. Differences distribution: Prior to the paired t -test analysis, the first step was to simply observe the distribution of the differences using a boxplot. Figure (1) shows the plots for the pretest and posttest data in a way of simple comparison.

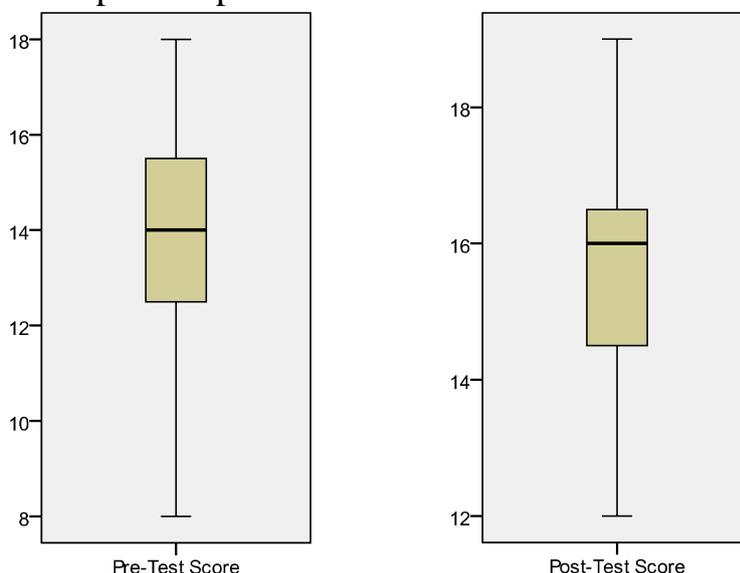


Figure (1) Two boxplots showing the distribution differences of the Pre and Post Tests scores

From figure (1) above, it can be seen that the scores of the students in the pretest are distributed within the range of 8+ to 18, with a concentration of scores at score 14. On the other hand, the posttest scores are distributed within the range of 12 to 18+, with a concentration at score 16. This is a clear indication that the students' performance are improved between the two periods of the tests.

2. Paired t -test: In order to determine whether the study methodology was effective, a paired t -test was performed using the data of the pre and posttests introduced earlier. To get adequate results, a calculation of *Post minus Pre* tests scores should be conducted to find out the differences in

the students' scores. In this case, there are two hypotheses to test. They are:

1. $H_0: \mu = 0$ (The null hypothesis: the mean of the differences is zero, i.e. the study methodology is ineffective.)
2. $H_a: \mu > 0$ (The alternative hypothesis: the mean of the differences is positive, i.e. the methodology is effective.)

This was done through calculating the paired t -test statistics and the results obtained are shown in Table (5) given below.

Table (5) Paired Samples Test of the Post-Pre test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Post- Pre test	1.6716	1.8961	.2316	1.2092	2.1341	7.216	66	.000

From table (5) above, it can be seen that the Post-Pre test mean is 1.6716 with a standard deviation of the differences given by 1.8961. The calculated t -statistics with the differences of the 66 scores (66 df) is given by 7.216, which has a p-value of 0.01. Out of these results, it can be noted that the mean of the Post minus Pre test scores differences is positive, which is supportive to the alternative hypothesis introduced earlier. Accordingly, the null hypothesis is rejected in this case, and in conclusion, these results provide evidence that the study methodology is effective in promoting the students' speaking skill.

11. Discussion

According to the results of the students' scores, it is found that applying the rubric in peer assessment and following the group discussion technique was effective positively. Adding to the increase in the students' scores, the goal behind the procedure of the empirical study was fulfilled. The group discussion technique helped significantly in creating motivated students with clear interest in participating in classroom discussions. Students were eager to take part in the authentic tasks and to give final conclusions. In certain tasks (like shopping list and balloon debate), the classroom was full of enthusiasm and in-depth discussions. However, when some students have a problem in managing their subgroups, they find cooperation and encouragement from their mates. In conclusion, a student-centered classes was actually created inside the classroom.

The study findings agree with the previous studies conducted earlier like Petkove and Petkova (2006, cited in Reddy and Andrade, 2010: 445), Reitmeier, Svendsen, and Vrch (2004, cited in Reddy and Andrade, 2010: 445), Naksuhara (2007), White (2009), Raza (2011) and those reviewed by Jonsson and Svingby (2007). They also ensure that working with rubrics need some training for both instructors and students. When used in peer assessment, rubrics are highly effective and more productive. Students need to have the chance to try their own chances of discovering their weaknesses by themselves and not to rely solely on the teacher. Consequently, they will be more attentive to their performance and progresses along the learning process. Also, having the peer assessment in mind, this can encourage them to perform in their best that helps in turn to create highly motivated students.

12. Conclusion and Recommendations

Having positive impacts on students' performance, rubrics and peer assessment were investigated by an empirical study with a sample of EFL learners at the university level. The study showed that the use of these two practices was effective particularly with the application of the group technique. Together, the three alternative practices were not merely positive in improving the students' performance, but also in raising their motivation and cooperation inside classroom and in creating students'-centered classes as well.

Finally, the researchers recommend the following points to the teachers of conversation:

1. Implementing the use of rubrics, peer assessment, and group work technique in conversation classes to help teachers manage large classes and create students-centered classrooms.
2. Analytic rubrics are highly recommended than holistic rubrics, since they detail the skill being assessed and make clear the reason behind a specific assessment.

13. References

- Anderson, Rebecca S. (1998). "Why Talk about Different Ways to Grade? The Shift from Traditional Assessment to Alternative Assessment". New Directions for Teaching, (74), Summer, 5- 16.
- Andrade, Heidi and Ying Du. (2005). "Student' Perspective on Rubric-Referenced Assessment". Practical Assessment, Research and Evaluation [PAREOnline]. 10 (3), April. n. Pag.
- Andrade, Heidi Goodrich. (2000). "Using Rubrics to Promote Thinking and Learning". Educational Leadership. 57 (5), 13-18.
- _____ (2001). "The Effects of Instructional Rubrics on Learning to Write". Current Issues in Education, [Online], 4 (4), n. Pag. Available at: <http://cie.ed.asu.edu/volume4/number4>
- Black, Paul and Wiliam Dylon. (1998). "Inside the Black Box: Raising Standards Through Classroom Assessment". Kappan Professional Journal, [Online], BERA short Final Draft, 1- 14. Accessed in Jan. 2012, from <http://www.pdkintl.org/kappan/kbla9810.htm>
- Brookhart, Susan M. Et al. (2009). "Promoting Student Ownership of Learning Through High-Impact Formative Assessment Practices". Journal of MultiDisciplinary Evaluation, 6 (12), 52- 67.
- Brualdi, Amy. (1998). "Implementing performance assessment in the classroom". ERIC/ AE digest series EDO- TM- 98- 0, August, 1-2.
- Caban, Heather L. (2003). "Rate Group Bias in the Speaking Assessment of Four L1 Japanese ESL Students". Second Language Studies. 21 (2), Spring, 1-44.
- Elliott, Alan C. and Wayne A. Woodward. (2007). Statistical Analysis: Quick Reference Guidebook. London: Sage Publications, Inc.
- Gammidge, Mick. (2004). Speaking Extra: a Resource Book of Multi-Level Skills Activities. Cambridge: Cambridge University Press.

- Guskey, Thomas R. (2003). "How Classroom Assessment Improve Learning". Educational Leadership. 60 (5), February, 6- 11.
- Hermer, Jeremy. (2001). The Practice of English Language Teaching. 3rd ed. London: Longman press.
- Jonsson, Anders and Gunilla Svingby. 2007. "The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences". Educational Research Review. (2), 130-144.
- Lombardi, Marilyn M. (2008). "Making the Grade: The Role of Assessment in Authentic Learning". Educause Learning Initiative: advancing learning through IT innovation. White paper. 1-16. Accessed in Dec. 2011, from <http://creativecommons.org/licenses/by-nc-nd/3.0/>
- Moskal, Barbara M. and Jon A. Leydens. (2000). "Scoring Rubric Development: Validity and Reliability". Practical Assessment, Research and Evaluation, 7 (10), n. Pag. Accessed in May, 2011 from <http://PAREonline.net/getvn.asp?v=7&n=10>
- Nakatsuhara, Fumiyo. (2007). "Developing a Rating Scale to Assess English Speaking Skills of Japanese Upper-Secondary Students". Essex Graduate Students Papers in Language and Linguistics. (9), 83-103.
- O'Malley, J. Michael and Lorraine Valdez Pierce. (1996). Authentic Assessment for English Language Learners. USA: Addison-Wesley Publishing.
- Popham, James W. (1997). "What's Wrong- and What's Right- with Rubrics". Educational Leadership, October,72- 75.
- Raza, W. (2011). "A Rating Scale to Assess English Speaking Proficiency of University Pakistan". Teaching and Learning

- Language Current Trends and Practices. University Sains Malaysia Press, Malaysia. 242-256. (gained via email from the author).
- Reddy, Y. Malini and Heidi Andrade. (2010). "A Review of Rubric Use in Higher Education". Assessment and Evaluation in Higher Education, 35(4), 435- 448, first published on 07 August 2009.
- Rubric Design Guide 'RDG': Establishing Rubrics to Guide your Grading. (2007). Center for Innovative Teaching and Learning CITL.
- Scott, J. (2006). "Authentic Assessment Tools". California Adult Literacy Professional Development Project. 33- 48. Accessed in August, 2011, from <http://www.calpro-online.org/eric/docs/custer/custer5.pdf>
- Stevens, Dannelle D. and Antonia Levi. (2005). Introduction to Rubrics. USA: Virginia, Stylus Publishing.
- Stiggins, Rick and Jan Chappuis. (2005). "Using Student-Involved Classroom Assessment to Close Achievement Gaps". Theory into Practice, 44 (1), 11-18. Accessed in March, 2012 from <http://www.jstore.org.tiger.sempertool.dk/stable/3496986>
- Topping, Keith J. (1998). "Peer Assessment between Students in College and University". Review of Educational Research, 63 (3), 249-276.
- _____. (2012). "Peers as a Source of Formative and Summative Assessment". In, J. H. McMillan (Ed.), Sage Handbook of Research on Classroom Assessment. Sage: Thousand Oaks. (Unpublished Book. This Chapter is gained by direct contact with the author via email on Dec. 2011)
- Ur, Penny. (1996). A Course in Language Teaching: Practice and Theory. Cambridge: Cambridge university press.

- Vu, Thu Thuy & Gloria Dall' Alba. (2008). "Exploring an Authentic Approach to Assessment for Enhancing Student Learning". (Refereed paper in a symposium entitled 'Re-imagining higher education pedagogies' AARE conference, Brisbane)
- Wheater, C. Philip et al. (2005). "Students Assessing Students: Case Studies on Peer Assessment". Planet, (15), December, 13- 15.
- White, Eddy. (2009). "Students' Perspectives of Peer Assessment for Learning in a Public Speaking Course". Asian EFL Journal, (33), January, 1-31.

[Speaking Rubric for Conversation Classes]

group:

Date:

Topic

My name is:

I am assessing:

Discussion leader:

Criteria	Superior	Advanced	Intermediate	Novice
Fluency	<input type="checkbox"/> continuous, clear and well organized speech	<input type="checkbox"/> continuous speech with some pauses to search for suitable words and/or organize thoughts	<input type="checkbox"/> frequent pauses to organize thoughts and/or for lack of vocabulary	<input type="checkbox"/> frequent long pauses with incomplete utterances
Grammar	<input type="checkbox"/> no errors	<input type="checkbox"/> some minor errors that do not obscure meaning	<input type="checkbox"/> many errors that do not obscure meaning	<input type="checkbox"/> major errors that obscure meaning
Pronunciation	<input type="checkbox"/> no errors	<input type="checkbox"/> some minor errors that do not affect communication	<input type="checkbox"/> many errors that do not affect communication	<input type="checkbox"/> major errors that affect comprehension
Vocabulary	<input type="checkbox"/> no use of Arabic	<input type="checkbox"/> hardly uses Arabic words	<input type="checkbox"/> almost half the speech is in Arabic	<input type="checkbox"/> most of the speech is in Arabic
	<input type="checkbox"/> Rich use of vocabulary (including fixed expressions and phrases)	<input type="checkbox"/> often adequate and accurate use of vocabulary	<input type="checkbox"/> repetition of words because of limited adequate and accurate vocabulary	<input type="checkbox"/> Inadequate and inaccurate use of vocabulary
Interaction	<input type="checkbox"/> Raises questions, responds, comments, helps to clarify others' ideas, and encourages others to participate in discussion	<input type="checkbox"/> only raises questions and responds to others	<input type="checkbox"/> Only responds to questions if asked by name	<input type="checkbox"/> no attempt to participate in discussion
Content	<input type="checkbox"/> Interesting details related to the topic including facts, evidences, personal stories,... <input type="checkbox"/> uses interesting humor	<input type="checkbox"/> questions and responses related to the topic with no details <input type="checkbox"/> restricted use of humor	<input type="checkbox"/> Circumlocutes around the topic with no valuable content <input type="checkbox"/> focuses on humor in each statement	<input type="checkbox"/> very short answers (yes/no or 2-5 words responses) <input type="checkbox"/> humor dominates speech
Discussion leader	<input type="checkbox"/> Starts, develops, maintains and closes discussion effectively within time limit	<input type="checkbox"/> Starts, develops, and maintains discussion with no closing within time limit	<input type="checkbox"/> Starts discussion but loses control over discussion & time	<input type="checkbox"/> unable to start or control discussion
	<input type="checkbox"/> Engages all members in discussion (by asking questions, requesting for clarifications, commenting, clarifying misunderstandings, correcting others' utterances, etc)	<input type="checkbox"/> Engages all members in discussion by asking questions only	<input type="checkbox"/> Engages specific members by asking questions and/or dominates discussion him/herself	<input type="checkbox"/> unable to engage others in discussion

Appendix I: the scoring rubric of the study

Appendix II: Posttest Topics

1. Women can do anything that men do in the world of work. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
2. I think that parents only get so offended by television because they rely on it as a babysitter and the sole educator of their kids. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
3. Old people (like grandfathers and parents) are not good friends to spend free time with. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
4. Mobile phones are irreplaceable devices. I cannot imagine my life without a mobile phone. Discuss with your mates and find out who agrees or disagrees and why. Try to find out the advantages and disadvantages of such devices. Then make a summary of your discussion.
5. Happiness is nothing more than good health and a bad memory. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
6. A celebrity is a person who works hard all his life to become well known, then wears dark glasses to avoid being recognized. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
7. The only time people dislike gossip is when you gossip about them. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
8. I find television very educating. Every time somebody turns on the set, I go into the other room and read a book. Discuss with your mates and find out who agrees or disagrees and why. Try to find out the advantages and disadvantages of such devices. Then make a summary of your discussion.
9. Family isn't about whose blood you have. It's about who you care about. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
10. A great marriage is not when the 'perfect couple' comes together. It is when an imperfect couple learns to enjoy their differences. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
11. Lack of friends means, stranger in one's own country. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.

12. A friend cannot be considered a friend unless he is tested on three occasions: in time of need, behind your back and after your death. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.
13. Music is essentially useless, it is a waste of time. Discuss with your mates and find out who agrees or disagrees and why. Then make a summary of your discussion.

تأثير استخدام الجداول التقييمية في تقييم النظير لتطوير مهارة التكلم لدى طلبة الجامعة العراقية من متعلمي اللغة الانجليزية

زينب جعفر عودة أ.د. بلقيس عيسى كاطع راشد

قسم اللغة الانجليزية، كلية التربية، جامعة البصرة

الخلاصه

شهد مجال القياس و التقييم منتصف التسعينيات تحولاً مما كان يسمى بالتقييم التقليدي إلى ما أصبح يدعى بالتقييم البديل. و نتج عن هذا التحول العديد من التطبيقات التي تهدف إلى ربط مجال التقييم بمجالي التعلّم و التعليم. (و من ضمن هذه التطبيقات، تطبيق الجدول التقييمي "Rubric" و تطبيق تقييم النظير "Peer Assessment"). فقد أثبتت دراسات سابقة التأثير الايجابي لهذين التطبيقين على كل من المتعلمين و المدرسين و المناهج. لذا تهدف الدراسة الحالية للتحقق من تأثير تطبيق الجدول التقييمي "Rubric" و تطبيق تقييم النظير "Peer Assessment" على مهارة التكلم لدى طلبة الجامعة من متعلمي اللغة الإنجليزية في دروس المحادثة. تم إتباع طريقة الاختبار ما قبل الدراسة و بعدها في تحصيل النتائج الإحصائية، إذ طُبِّقَ كل من التطبيقين البديلين فيما بين الاختبارين على طلبة الجامعة من متعلمي اللغة الانجليزية في قسم اللغة الانجليزية في كلية التربية، جامعة البصرة. حُلَّتْ نتائج الدراسة باستخدام الاختبار "ت" T -test و أظهرت النتائج تأثيرات إيجابية لتطبيق الجدول التقييمي "Rubric" و تطبيق تقييم النظير "Peer Assessment" على إنجاز عينة الطلبة البالغة ست و ستين طالباً و طالبةً. بالإضافة إلى ذلك، ساعدت التطبيقات البديلة على خلق صف تفاعلي ذي فعالية عاليه محوره الطلبة.