The Impact Of Using Scoring Rubric In Peer Assessment On Promoting Iraqi EFL Learners' Speaking Skill In The University Level

1 author:

Zainab Jaafar Auda
University of Basrah
37 PUBLICATIONS   **4** CITATIONS

SEE PROFILE

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

Project   The Impact of Using Scoring Rubrics in Peer Assessment in Promoting Iraqi EFL Learners' Speaking Skill at the University Level View project

Project   Academic Writing Course View project

Department of English Language
College of Education for Humanities
University of Basrah

THE IMPACT OF USING SCORING RUBRICS IN PEER ASSESSMENT ON PROMOTING IRAQI EFL LEARNERS' SPEAKING SKILL AT THE UNIVERSITY LEVEL

A THESIS SUBMITTED TO
THE COUNCIL OF THE COLLEGE OF EDUCATION FOR
HUMANITIES, UNIVERSITY OF BASRAH
IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF ARTS
IN ENGLISH LANGUAGE

BY

ZAINAB JAAFAR AUDA

SUPERVISED BY
PROF. DR. BALQIS I.G. AL-RASHID

2013 A.D.   **1434 A.H.**

Bism Allah Arrahman Arraheem

In the Name of Allah, the Merciful, the Compassionate

To my parents and all members of my family,

Thank you for all moments of support and prayer

SUPERVISOR'S CERTIFICATION

I, hereby, certify that this thesis which is entitled "The Impact of Using Scoring Rubrics in Peer Assessment on Promoting Iraqi EFL Learners' Speaking Skill at the University Level" has been prepared under my supervision at the Department of English, College of Education for Humanities, University of Basrah, in partial fulfillment of the requirements for the degree of Master of Arts in English Language.

Signature:

Supervisor: Prof. Dr. Balqis I.G. Al-Rashid

In view of the available recommendation, I forward this thesis for debate by the Examining Committee.

Signature:

Name: Prof. Dr. Balqis I.G. Al- Rashid

Head of the Department of English

Date: 15. Nov. 2012

REPORT OF THE EXAMINING COMMITTEE

We, hereby, certify that we have read this thesis which is entitled "The Impact of Using Scoring Rubrics in Peer Assessment on Promoting Iraqi EFL Learners' Speaking Skill at the University Level", and, being an Examining Committee, have examined the candidate in its contents, and that, in our opinion, it is appropriate for the partial fulfillment of the requirements for the degree of Master of Arts in English Language.
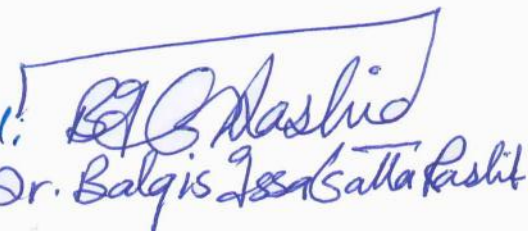
Signature:

Member: Dr. R.S. Al-Nawas

Signature:

Member: Assist. prof. Jameel Q. Hameed
28.1.2013

Signature: Prof. Dr. Riyadh Tariq Kadhim Al-Ameed

Chairman:

prof. Dr. Balqis Issa Satta Rashid

Approved by the Council of the College of Education for Humanities, University of Basrah

Signature:

Prof. Husain A. Hashiem

Dean of the College of Education for Humanities

Date:

CONTENTS

CHAPTER FIVE: RESULTS AND DISCUSSION

CHAPTER SIX: CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS

ACKNOWLEDGMENTS

ABSTRACT

In the mid 1990s, the field of assessment and evaluation witnessed a shift from traditional assessment to alternative assessment. The shift presented a number of new practices aiming at linking teaching and learning processes with assessment process for the sake of making the most of the assessment benefits. Among those practices, scoring rubrics and peer assessment have been found to have positive impacts on learners, teachers, and curricula. In addition, rubrics are found to be more effective in assessing the skills that are usually assessed subjectively, like the speaking skill. Thus, the primary goal of the study is to investigate the impact of using scoring rubrics on promoting the EFL learners' speaking skill in conversation classes.

Being used in the study as a mere assessment tool, no real effect was found of the scoring rubric on the students' performance. Therefore, the rubric was used in the students' peer assessment inside the classrooms. Accordingly, the study hypothesizes that, firstly, the scoring rubric, when used in peer assessment in conversational classes, has a positive impact in promoting the students' speaking skill inside the classroom. Secondly, when the two practices are implemented with the group work technique, they can be effective in creating students-centered classrooms. Thirdly, a well-formed rubric can achieve high reliability among different raters.

In order to test the hypotheses of the study, firstly, the researcher followed a pretest- posttest technique between which she applied the three practices of alternative assessment (i.e. rubric, peer assessment, and group technique) to a convenience sample of EFL students (*N*=74) in the Department of English Language, College of Education for Humanities, University of Basrah. To check the results, a paired t-test analysis was

implemented showing positive impacts of the rubric and peer assessment on the performance of the participants. Secondly, it was clear that the alternative assessment practices helped creating an interactive student-centered classroom with highly motivated students. And thirdly, being tested through Cronbach's alpha reliability test, the rubric of the study achieved a high reliability of .934. To sum up, following the study findings, the methodology of the study is found to be successful in achieving the aims of the study.

LIST OF ABBREVIATIONS

AA: Alternative Assessment

CITL: Center for Innovative Teaching and Learning

Cv: Coefficient of Variation

EFL: English as a Foreign Language

ESL: English as a Second Language

PA: Peer Assessment

RDG: Rubric Design Guide

TA: Traditional Assessment

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER ONE: PRELIMINARIES

1.1. Introduction

Following the development occurring in the world's tremendous fields of education, the field of assessment and evaluation found its way similarly. New terms and practices have started to appear, and new trends of interest have followed. Traditional assessment is replaced by authentic and alternative assessment that consider assessment as a way of rectifying the teaching process and directing the learning process to the targeted goal/s more than mere grading and firm testing. Yet, while some specialists worked on establishing the base line for their new trends, other educators found themselves free from following them and some others were more strict in opposing them. Nevertheless, neither following nor opposing the new trends is a big issue since experiment is the best proof.

Since the mid 1990s, new practices of assessment flourished under the titles of alternative assessment and authentic assessment. Rubric and peer assessment are the two practices being investigated here. Not being cited in a common dictionary, the definition of rubric is introduced by a number of educators and researchers as an assessment tool of students' levels of proficiency in a certain skill (Popham, 1997; Brualdi 1998; Andrade 2001; Reddy 2007). Furthermore, rubric's advocates introduce rubric as an effective approach of teaching that goes side by side with assessment (Andrade, 2000; Reddy and Andrade, 2010). As a main type of formative assessment, rubrics became an essential assessment tool used by teachers worldwide. Yet, in order to make the best of a rubric, the practice of peer (and/or self) assessment is highly advised to be integrated simultaneously.

Peer assessment is another primary practice of formative assessment. It is defined as "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). In view of what is stated above, having students participate in their assessment offer valuable outcomes to students' performance since peer assessment proved to be a successful motivating tool as well as assessment and feedback one (White, 2009; Reddy and Andrade, 2010; Kutlu *et al*, 2010; Topping, 2012). Adding to that, it gives behavioral gains since students are indulged in practices of communication and cooperation (Topping, 2012).

The current study is an experimental research of two new trends in assessment, namely: Rubric and Peer Assessment, being implemented along with a third practice, group discussion technique, in EFL conversation classes. The participants represent a convenience sample of the target population, i.e. Iraqi EFL learners at the university level, who experience the alternative assessment practices for the first time. The main goal behind this study is to examine the impact of using these practices on the students' performance in conversation classes. Additionally, some extra goals have been integrated along the study, like examining the students' perspective of the application of rubrics, peer assessment, and group discussion technique, testing the reliability of the use of the rubric in the assessment of the speaking skill, and comparing the traditional and rubric assessment given by the same raters.

1.2.   The Problem of the Study

Speaking is one of the most difficult skills to assess objectively (Harris, 1969: 81). Teachers usually have their own assessment values (or scales) according to which they assess the students' performance inside the classroom. However, such values are vague to students and usually not publically

announced, and in many cases, personal intuitions about the students are involved.

As a result, EFL learners complain from not understanding the basis upon which they get their assessment of the speaking skill. The final grades they get at the end of the semester are not accompanied usually with justifications of why they are as such. Moreover, neither active students nor passive ones are aware of their strengths or weaknesses in speaking, and how to work on improving the latter. In this case, and regardless of their performance in the classroom, failing students always find themselves the victims of the teaching process. In simple words, the problem being stated here is that traditional assessment of the speaking skill is a subjective assessment that merely grade the students into levels of proficiency with no clear justification.

1.3.   Aims of the Study

The study aims at introducing the new practices of assessment to both EFL teachers and students. Through implementing these practices in conversation classes, it aims at:

1. Investigating the impact of using scoring rubrics in promoting the EFL students' speaking performance in conversation classes,

2. Raising the students' awareness of their weaknesses and strengths through the use of rubrics in peer assessment,

3. Raising the students' motivation and creating student-centered classrooms through the application of rubrics, peer assessment, and group technique,

4. Examining the students' perspectives of the use of rubrics and peer assessment in conversation classes, and

5. Diminishing the teachers' subjectivity of speaking assessment through out the use of a reliable rubric.

1.4.   Hypotheses of the Study

Generally speaking, the study hypothesizes that the practices of alternative assessment are successful tools of teaching, learning, and assessment. However, specific hypotheses are investigated in the study. They are as follows:

1. A well-formed rubric can be an effective tool in promoting students' speaking skill.
2. Rubrics, when used in peer assessment and group work technique, can create a highly motivated student- centered classroom.

3. A reliable rubric can enhance the objectivity of speaking assessment.

4. Rubric assessment can be more reliable than traditional assessment.

1.5.   Procedure of the Study

In order to check the truthfulness of the hypotheses of the study, the procedure given below is followed:

1. *A questionnaire implemented on the teachers of conversation*: In order to establish the foundation of the study, a questionnaire implemented on the teachers of conversation in University of Basrah and Shatt Al-Arab Private University College is designed. The main aim behind it is to gather the related information to the study, concerning the teachers of conversation and the students' performance in conversation classes.

2. *A pilot study:* Since the application of rubrics and peer assessment is the first of its kind in the Iraqi universities during the period of the study, a pilot study was conducted as a training course for the researcher in the use of rubrics and peer assessment in the formative assessment of the students.

3. *Pretest- posttest procedure to the students participating in the study*: Prior to engaging the participants in the alternative assessment practices, a pretest was made and considered as a basis for the students' levels in

conversation classes. Then, the researcher applied the use of rubrics and peer assessment for a period of ten lectures, after which a posttest followed to test the progress in the students' performance.

4. *A students' questionnaire*: Since the participants in the empirical study are the first to experience the use of the alternative assessment practices, then their opinions are important to the study. Accordingly, a questionnaire (Appendix VIII) was carried out to examine their perspectives in that concern.

5. *An interrater reliability test*. In order to test the consistency of the rubric applied in the empirical study, an interrater reliability test was made with the help of a number of the faculty members in the Department of English, College of Education for Humanities. Besides, the test was supported with a traditional assessment that was performed to check the differences between the two types of assessment.

1.6. Scope of the Study

The scope of the study extends in investigating the applications of three alternative practices, namely: scoring rubric, peer assessment, and group work, in the assessment of the Iraqi EFL learners' speaking skill.

1.7. Limits of the study

The empirical study is limited to the application of rubrics, peer assessment and group work technique in conversation classes with a convenience sample of the Iraqi EFL learners (third stage students) from the Department of English, College of Education for Humanities, University of Basrah, during the academic year 2011-2012.

1.8. Significance of the Study

The study is significant due to the novelty of the topic. No thesis implemented in the assessment of speaking in the Iraqi universities could be

located, let alone alternative assessment of speaking. Thus, in tackling the alterative assessment of the speaking skill, the current study is the first of its kind in the assessment of the speaking skill at the university level.

1.9. Thesis layout

The thesis consists of six chapters summarized as follows:

Chapter One: introduces the preliminaries of the study and builds the bases upon which the study is held.

Chapter Two: surveys the theoretical framework of the study, starting from primary information differentiating assessment from testing, moving to the introduction of alternative assessment, the shift from traditional to alternative assessment, the advantages and the disadvantages of alternative assessment, alternative assessment as a link between learning, teaching, and assessment, principles of alternative assessment, and finally a theoretical overview of speaking assessment and teaching.

Chapter Three: tackles the related details of rubrics and peer assessment and their advantages and disadvantages. The chapter ends with a literature reviewing the available related studies, focusing on the use of the investigated alternative assessment practices in assessing the speaking skill.

Chapter Four: outlines the materials and methods used and followed by the researcher in the empirical part of the study. It gives details on how the study was implemented, the participants, the materials used, and all the related aspects.

Chapter Five: discusses the results gained out of the empirical study and the statistics analyses employed with their detailed discussions.

Chapter Six: draws some conclusions, recommendations, and suggestions for future works.

1.10. Keywords

Assessment, testing, alternative assessment, authentic assessment, formative assessment, criterion-referenced assessment, rubric, self-assessment, peer assessment, reliability, validity, speaking, Cronbach's alpha, paired t-test.

CHAPTER TWO: THEORETICAL OVERVIEW

2.1. Introduction

In order to establish a clear understanding of the major aspects of the research, this chapter introduces the basic principles of testing and assessment, alternative assessment, the reason that led to the shift from traditional assessment to alternative assessment and differences between the two trends, advantages and disadvantages of alternative assessment, its main practices, and some other principles related to the field.

2.2. Testing vs. Assessment

Assessment and testing are sometimes misunderstood to refer to the same meaning. In researching the topic, several differences between the two terms are clarified. While testing is seen as a tool of accountability, classroom assessment is more about reviewing and promoting students' performance. A quick view about some definitions may help to clarify the differences, and to help recognizing the meaning of assessment intended in the current study.

In Longman Dictionary of Language Teaching and Applied Linguistics (Richards and Schmidt, 2002: 35), the term "assessment" is defined as

> ... a systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence. ... The term "testing" is often associated with large-scale standardized tests, whereas the term "assessment" is used in a much wider sense to mean a variety of approaches in testing and assessment.

Brown (2003: 3-5) defines a 'well-constructed test' as "an accurate measure of the test-taker's ability within a particular domain." On the other hand, he

defines assessment as "an ongoing process that encompasses a much wider domain." He goes on explaining the relationship among the three processes (tests, assessment and teaching) involved in language classroom by drawing a figure (Figure 2.1) of embodied focused circles reflecting the connectedness and dependency of each component on the other.
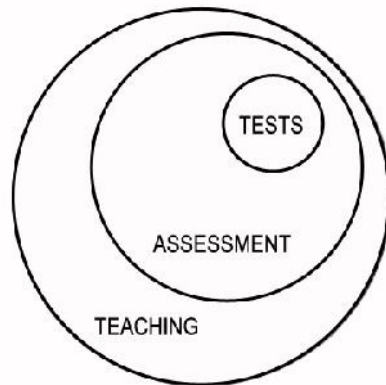
Figure (2. 1) Tests, Assessment, and Teaching (Brown, 2003: 5)

From Figure (2.1), it is clear that the process of teaching covers the widest range among the whole processes. This reflects that through teaching, the teacher has the responsibility for teaching and observing students' performance along the duration of the learning process. Assessment appears almost interacting with teaching which assures the idea that assessment occurs along the whole process of teaching, for the sake of giving feedback and redirecting the whole teaching process for achieving the learning objectives. In this sense, teaching and assessment cooperate in reaching the ultimate goals of the learning process.

At the end of the learning process comes the last item that seems to be inevitable, which is testing. Although it has a rare achievement in the learning process, testing is inevitable for the sake of accountability. Primarily, its main importance lies in moving students from one level to another higher one.

2.3.   Alternative Assessment and Authentic Assessment

Alternative assessment consists of any method of examining what students know or can do that is intended to show growth and inform teaching. 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 overlapping 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." Consequently, the current study uses the two terms as referring to one aspect of assessment that involves *integrating assessment practices into the classroom as essential ingredients of teaching for the sake of achieving the learning goals.*

2.4.   The Shift from Traditional Assessment to Alternative Assessment

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

In 1998, Anderson studied the shift in assessment from traditional assessment towards alternative assessment. She tried to figure out the differences between the two trends of assessment by comparing the philosophical beliefs and theoretical assumptions associated with each trend. The following points illustrate the whole comparison (using TA for Traditional Assessment and AA for Alternative Assessment): (Anderson, 1998: 8-11)

1.   Knowledge is assumed to have universal meaning in TA and multiple meanings in AA. In other words, "it is possible for everyone to reach a

consensus about meaning because knowledge has 'the same meaning for all individuals everywhere" (Berlak, 1992, p. 13, cited in Anderson, 1998: 8) while in AA "it is impossible for everyone to reach a consensus about meaning because each individual brings his or her own diverse interpretation to an ever-changing situation."

2. TA "Treats Learning as a Passive Process" whereas AA "Treats Learning as an Active Process." The old metaphor of "empty vessel" referring to the mind of a student to be "filled" with the knowledge introduced by the teacher turns the focus of learning process on "learning about something rather than learning how to do something." This will end with a passive learning process dealing with passive learners, "novices", memorizing the knowledge they receive from their teacher, the "expert". On the contrary, AA looks at learning as a "natural, integral, and ubiquitous part of living" (Bintz, 1991: 309, cited in Anderson, 1998: 9). Students are seen as active learners, and learning involves "producing, rather than reproducing knowledge" (Newmann and Archbald, 1992: 72, cited in Anderson, 1998: 10).

3. TA "Separates Process from Product" while AA "Emphasizes Process and Product." In TA, tests are used to evaluate students' final products. The outcome of the test is taken to reflect students' learning, regardless of whether or not the learning process happened along the period of teaching. In AA, the learning process and students' products are integrated.

4. TA "Focuses on Mastering Discrete, Isolated Bits of Information" whereas AA "Focuses on Inquiry." TA deals with a hierarchy of bits of information that represent "lower-level thinking skills". It expects students to "master and demonstrate specific skills at one level before

moving on to the next." Conversely, AA focuses on "developing real-world problem-solving skills that will lead people to observe, think, question, and test their ideas" (Herman, Aschbacher, and Winters, 1992, cited in Anderson, 1998: 10).

5. TA "Assumes the Purpose of Assessment is to Document Learning" while AA "Assumes the Purpose of Assessment is to Facilitate Learning." "Typically, traditional assessment is used **only** [Bold is the researcher's] to monitor students' learning." Moreover, it ranks students according to their test outcomes into levels. Differently, the purpose behind AA is "to enhance students' learning" (Johnston, 1989; Short and Burke, 1991; Wolf, 1990, cited in Anderson, 1998: 11). The feedback received by students helps in redirecting the learning process for a better performance. Furthermore, students are not meant to be sorted or classified by AA.

6. TA "Believes that Cognitive Abilities Are Separated from Affective and Conative Abilities" while AA "Recognizes a Connection between Cognitive, Affective, and Conative Abilities." The focus of traditional assessment is primarily on cognitive abilities. No attention is given to students' interest in performing activities. That is in contrary to AA where students' care about an activity and its goals are engaged. This encourages them "to invest their time and effort in it, and, as a result, they learn more from it." (Anderson, 1998: 10)

7. TA "Views Assessment as Objective, Value-Free, and Neutral" whereas AA "Views Assessment as Subjective and Value-Laden." Traditional assessment assumes that facts and values are distinct and separable entities that can be measured objectively (Berlak, 1992, cited in Anderson, 1998: 8). Decisions about what to teach and test are not considered value-laden. Advocates of alternative assessment practices

believe that decisions about what to teach and assess are subjective and value-laden (Bintz and Harste, 1994, cited in Anderson, 1998: 10). "Indeed, value systems not only influence decisions about what assessment questions get answered, but also about what assessment questions get asked in the first place" (Bintz,1991, p. 309, cited in Anderson, 1998: 11).

8. TA "Embraces a Hierarchical Model of Power and Control" while AA "Embraces a Shared Model of Power and Control." InTA, generally the teacher alone has the power to make decisions about what is learned and how it is assessed while in AA teachers are advised to share the power with students to make decisions about what they learn and to determine how well they are learning.

9. TA "Perceives Learning as an Individual Enterprise" whereas AA "Perceives Learning as a Collaborative Process." TA focuses on individual students' performance. While TA asserts competition among students, AA raises a high collaboration among students in the learning process. Cooperation is found also between students and teachers who will share the responsibility for creating "a substantive curriculum in the classroom".

Anderson (1998: 9) draws the following figure to summarize the above comparison.

Traditional
Assessment

Alternative
Assessment

Universal assessment	Knowledge	Multiple meanings
Passive process	Learning	Active process
Separates process from product	Process	Emphasizes process and product
Discrete, isolated bits of information	Focus	Focuses on inquiry
To document learning	Purpose	To facilitate learning
Cognitive abilities as separate from affective and conative abilities	Abilities	Connects between cognitive, affective and conative abilities
Views assessment as objective, value-free and neutral	Assessment	Views assessment as subjective and value-laden
Hierarchical model	Power and Control	Shared model
Learning as an individual process	Individual vs Collaborative Process	Learning as a collaborative process

Figure (2.2) Traditional Vs. Alternative Assessment (Anderson, 1998: 9)

The need to shift from traditional assessment towards alternative assessments is being highly appreciated in the 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). Yet, Lombardi (2008: 3)

states that such a shift occurred due to several factors, among which he cites "economic conditions, new scholarship on learning, and a student population with new expectations of educational institutions." Moreover, he proposes the following table that shows the comparison between traditional and alternative assessment. (Lombardi, 2008: 5)

Table (2.1) 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
		(Lombardi, 2008: 5)

2.5.   Alternative Assessment as an Integral Part of Teaching

Traditional assessment is seen as focusing on the outcomes of the learning process more than the process itself. Usually, students are busy studying the materials "which are assessed in a test" and teachers are busy "teaching to the test" (Anderson, 1998: 5). Furthermore, the kind of feedback given in traditional assessment appears to be worthless since it is only given at the end of a learning process. Black and Wiliam (1998: 8) point out that the feedback given to students in the type of marks or grades does not benefit them. Students need to

be aware of their achievements along the learning process, to be informed of their strengths and weaknesses, and suitable means for improvement. It is very important to know one's weaknesses in advance to start curing them before the whole process is over. It is obvious that a test at the end of the process of teaching/learning is pointless since it "is too late" to make any reformation in the results (Black and Wiliam, 1998: 8; Gusky, 2003: 6). Vu & Alba (2008: 7) put it clearly; "[a]ssessment is not an end in itself but, rather, an opportunity for students to learn and to reflect on their learning in a way that enhances future learning and professional development."

Unlike traditional assessment, the practices of alterative 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 is not to say 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 precious information of students' levels (Black and Wiliam, 1998: 8). Thus, teachers should be aware of the importance of linking feedback with teaching and redirecting their teaching 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. Guskey (2003: 9) describes the integration between teaching and assessment in the following way:

> To become an integral part of the instructional process, assessments cannot be a one-shot, do-or-die experience for students. Instead, assessments must be part of an ongoing effort to help students learn. And if teachers follow assessments with helpful corrective instruction,

then students should have a second chance to demonstrate their new level of competence and understanding. This second chance helps determine the effectiveness of the corrective instruction and offers students another opportunity to experience success in learning.

For the purpose of illustrating how assessments "influence and inform" teaching, Brown (2004:105-6) mentions three major points associated with students, curriculum, and teachers. First, assessment helps teachers to conduct a comprehensive evaluation. Second, assessment provides "good pedagogical templates" in support to teaching and curriculum. And third, assessment helps educators "better assess students' understanding of procedural knowledge, which is not so easily judged through traditional assessment methods." Since assessment helps to investigate students' weaknesses (as well as strengths), then it enables teachers to redirect their methods of teaching in a way that handle those weaknesses (Stiggins and Chappuis, 2005: 12). However, the effectiveness of assessment practices depends not only on the teacher's realization of students' weaknesses and strengths, but also on the teachers' effort on helping the students to understand the goals of the learning process and how to achieve them. (Brookhart *et al* 2009: 53)

By evidence, studies found that when the practices of classroom assessment are integrated with the teaching/learning process, the results to both teachers and students will be positive. (cf. 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.*

For the sake of raising standards, Black and Wiliam (1998) conducted an extensive survey by investigating what happens inside the classroom. Their survey focused on one aspect of teaching that is known as "formative assessment*"*. The argument they developed is that *formative assessment is "at the heart of effective teaching."* [Italic is the researcher's]. In their survey, Black and Wiliam studied 580 articles and chapters out of which they used 250 as sources. In addition to that, they included comments on their work by leading educational experts from Australia, France, Hong Kong, Southern Africa and the USA and later to the study, they published a summarized draft entitled 'Inside the Black Box', a metaphor they used in reference to the unsupervised classroom assessment processes performed by teachers of students' daily performance. They concluded that if it is communicated in the right way, formative assessment is powerful in enhancing students' achievements, particularly "low achievers". This is because formative assessment focuses on diagnosing the students' weaknesses, the problems they confront, and setting the goals of treating them (Black and Wiliam, 1998: 4).

2.6.   Authentic Assessment in Higher Education

Though started its application in kindergarten, authentic assessment gained a wide acceptance among educators and students in higher education. Different practices were used since the evolvement of authentic assessment that proved a 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 & Alba,

2008: 7). In their exploration for an authentic approach of assessment to enhance students' learning, Vu & Alba (2008: 7-8) highlight five features that make successful the use of authentic assessment in higher education. In brief, the points suggest the following:

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

2   Experiencing assessment tasks "can also provide space for students to challenge outdated ideas, routinized 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.

2.7. Typology of Alternative Assessment

Classroom assessment can be either **formal**, referring to "exercises or procedures specifically designed to tap into a storehouse of skills and knowledge", or **informal**, which can be of different forms, like unplanned comments and spontaneous feedbacks to students. (Brown, 2003: 5). Besides, alternative assessment can be either **formative** or **summative** or both (Topping,

2012: 3[1]). Brown (2003: 6) defines formative assessment as "evaluating students in the process of 'forming' their competencies and skills with the goal of helping them to continue that growth process." The aim behind formative assessment is to "improve learning while it is happening in order to maximize success, rather than merely determine success or failure only after the event." (Topping, 1998: 249) On the other hand, the aim behind summative assessment is to "measure, or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction." (Brown, 2003: 6)

Formative assessment (the one adopted in this study) is an essential tool in enhancing the learning/teaching process. Through implementing its practices into classroom assessment, teachers can have an ongoing awareness of students' weaknesses and strengths, and can work on improving the weaknesses and encouraging the strengths as well. What is more, formative assessment can lead to a final summative assessment. Teachers, if conducting formative assessments in their classrooms, can gather continuous information about the progress of each student's performance, and in turn can give a fair summative assessment to each student (William and Black, 1996, cited in Black and Wiliam, 1998: 12). It is for this reason summative assessment is usually referred to as 'assessment *of* learning' (given that it gives a summery to the students' achievements at the end of the learning period) while formative assessment as 'assessment *for* learning' (because it informs about the students' progress in a continuous feedback along the learning process that help in directing both the learning and teaching processes towards the preset goals (White, 2009: 3).

Furthermore, there are two other types of assessment that are important to consider, they are **norm-reference assessment** and **criteria-reference**

---

[1] This is a chapter in an unpublished book, received via email in a separate document from the author himself (i.e. Prof. Keith Topping) in Dec. 2011. The paginations used here are the ones found in the document received.

assessment. In norm-reference assessment, Brown (2003:7) proposes that "each test-taker's score is interpreted in relation to a mean (average score), median (middle score), standard deviation (extent of variation in scores), and/or percentile rank." On the other hand, criteria- referenced assessments are designed "to give test-takers feedback, usually in the form of grades, on specific course or lesson objectives." (Brown, 2003: 7) Thus, criteria- referenced assessment may be highly appreciated in alternative assessment since it provides feedback to students that may be invested to redirect the teaching and the learning processes.

2.8. Practices of Performance Assessment

Different kinds of performance assessment practices are widely spread and applied to numerous fields of knowledge in a way that meets the teaching objectives. Performance assessment (the type of assessment that focuses on the students' performance) and criterion-referenced assessment are encouraged in highly subjective assessment tasks. Tasks such as multiple-choice and true-false are typically easy to be assessed with high objectivity, since they provide accurate answers. However, other topics that involve critical thinking, communication, and problem-solving skills may not be well assessed with such tests like the multiple-choice test, and for those, performance assessment practices are encouraged. The main idea behind such practices is to decompose a specific skill into its constituent criteria and assess students' achievement on whether they meet a certain level of that criteria or another. As a result, this will help in decreasing the subjectivity in assessment (Perlman, 2003:497).

Among the different practices of performance assessment, Lombardi, (2008: 6) introduces four basic ones. They are:

1.   **Rubric:** a rating scale that is shared with students and mostly preferred to be designed in cooperation with them. A well-

constructed rubric analyzes the criteria of the work being assessed and describes clearly "the difference between excellent and weaker work." (Lombardi, 2008: 6)

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." (ibid)

In addition to the mentioned practices, there is the use of self-assessment in which students are engaged in assessing their own performance similarly to peer assessment. "Far from being a luxury", Black and Wiliam (1998: 7) consider self-assessment as an essential component of formative assessment because it enable students to "understand the main purposes of their learning and thereby grasp what they need to do to achieve."

The present study, as clarified in the methodology, adopts the use of the three practices of alternative assessment, namely: rubrics, peer assessment, and group work technique, in order to help enhance students' performance and decrease the subjectivity of assessing students' speaking skill. Such a skill usually does not lend itself to an objective assessment. This is partly based on Perlman's (2003:497) statement that

> Because performance assessment does not have an answer key of the
> type that a multiple-choice test does, scoring a performance assessment
> necessarily involves making some subjective judgments about the

quality of a student's work. A good set of scoring guidelines or rubrics provides a way to make fair and sound judgments by setting forth a uniform set of precisely defined criteria or guidelines for judging students work.

2.9. Advantages and Disadvantages of Authentic Assessment

In spite of the numerous advantages of authentic assessment that proved to be helpful in enhancing students' achievements as well as improving the whole teaching process, authentic assessment still suffers from several disadvantages. Although being encouraged by different researchers, authentic assessment is seen as difficult to apply and loading more responsibilities on teachers. Lombardi (2008: 5) proposes that teachers, loaded with nonteaching responsibilities, will be overloaded if they were to prepare and assess authentic tasks.

Yet, it is important to have a look at both the advantages and disadvantages of performance assessment. Table (2.2) explains them as cited in Perlman (2003:504).

Table (2.2) Advantages and Disadvantages of Performance Assessment

Advantages	Disadvantages
1. Provide rich learning experiences 2. Simulate real-world problem solving 3. Encourage students to critically evaluate their own work 4. Provide teachers with insight into their students' cognitive processes 5. Foster good instruction 6. Can be an excellent measure of students' abilities to synthesize, evaluate, and solve problems	1. Can be expensive and time-consuming to administer and score 2. a good result on one performance task may not generalize well to similar tasks 3. the subjectivity inherent in scoring a performance assessment may make some people uncomfortable 4. certain kinds of knowledge and skills are more efficiently assessed using other assessment formats, such as multiple-choice tests
	Perlman (2003:504)

However, Brown (2004: 105-8) cites the advantages of performance assessment stated by ten previous studies extending from 1992 until 2002. They are: Moss (1992, 229-230), Herman, Aschbacher, and Winters (1992, 48), Miller and Legg (1993), Khattri, Reeve, and Kane (1998, 26-27), Jones, R. L. (1985), Shohamy (1992, 517-518), Short (1993), Hudson and Yoshioka (1998, 15-16), Brown, Hudson, Norris, and Bonk (2002, 6), Brown and Hudson (2002, 74-78). After explaining each researcher's proposed benefits of performance assessment in detail, Brown (2004: 109) summarizes them in one brief comprehensive list as shown in Table (2.3).

Table (2.3) Summary of Benefits of Using Performance Assessment (Brown, 2004: 109)

Content
Assess only relevant content
Measure productive language use
Can measure the interaction of receptive and productive skills
Add a personal aspect to assessment
Measure abilities to respond to real-life language tasks
Assess language ranging from achievement to proficiency
Test contextualized and complex language
Test more than multiple-choice recognition
Test higher order thinking skills
Assess learning processes
Assess students' understanding of procedural knowledge

Scoring
Use only real-world criteria for selection and scoring
Help teachers or other raters be accurate, unbiased, and consistent in scoring
Mediate rater bias effects in testing

Score Interpretations
Minimize guessing as a major factor
Provide diagnostic information in functional or task-based curriculums
Supply achievement information in functional, or task based curriculums
Assess students' knowledge and abilities better than traditional multiple-choice tests do
Encourage and document critical thought, creativity, and self-reflection
Demonstrate students' weaknesses *and* strengths in detailed and real-world terms
More accurately predict students' abilities to use language in future real-life situations
Encourage control of score interpretations at the local classroom and school levels

Curriculum Development
Can be integrated into and become a part of the curriculum
Align assessment and instructional activities with authentic, real-life activities
Test in harmony with curriculum goals and objectives
Help teachers define excellence
Help teachers plan how to help students achieve excellence
Support instruction and curriculum by providing good pedagogical templates
Help teachers conduct *comprehensive* evaluation of students' achievement
Create positive washback effects on instruction
Avoid factors leading to Lake Wobegon effects

Decision Making
Counterbalance the negative effects of washback from standardized tests
Document the procedures used in making important judgments about students
Help promote multi-faceted approaches to information gathering for decision making
Support drawing of conclusions at the local classroom and school levels
Encourage control of decision making at the local classroom and school levels
Assess continuously and repeatedly so that change can be monitored over time

Communication
Involve faculty and assessment team in collaborative activities
Establish standards that are clear to students and authentic
Help teachers communicate to students what constitutes excellence
Help teachers communicate to students how to evaluate their own work
Help teachers communicate goals and results to parents and others

2.10. Principles of Language Assessment

In designing any tool of assessment, certain principles are inevitable to be considered. Concerning language assessment, Brown (2003:19) suggests five "cardinal criteria" to consider in "testing a test", with no priority in ordering a certain criterion over another. They are practicality, reliability, validity, authenticity, and washback. The following points cover them in detail.

2.10.1. Practicality

An effective test is practical if it:

- Is not excessively expensive,
- Stays within appropriate time constraints,
- Is relatively easy to administer, and
- Has a scoring evaluation procedure that is specific and time-efficient. (Brown, 2003: 19)

2.10.2. Reliability

"Reliability refers to the consistency of assessment scores." (Moskal and Leydens, 2000: n.p.) Reliability of a test can be considered from several basic aspects. Mousavi (2002, cited in Brown, 2003: 21) lists four points: "fluctuations in the student, in scoring, in test administration, and in the test itself."

2.10.2.1. Student-Related Reliability

This issue is related to the students themselves. Some factors like anxiety, illness, fatigue, and other physical or psychological factors, may affect the achievement of any test-taker.

2.10.2.2. Rater Reliability

There are two forms of rater reliability that are important in classroom assessment, they are

a) **Inter-rater reliability**, and

b) **Intra-rater reliability**.

Unlike multiple-choice tests, authentic assessments need the personal judgement of a teacher to score the students' achievements that may lead to subjectivity and lack of consistency with other teachers (O'Malley and Pierce, 1996: 19). Therefore, **inter-rater reliability** occurs when two independent raters reach a consensus on a particular student's performance under the same conditions. To achieve such a goal, the criteria that guide the rating process should be clear enough for both raters (Brown, 2003: 21; Moskal and Leydens, 2000: n.p.).

On the other hand, **intra-rater reliability** is a lack of internal consistency which happens frequently to classroom teachers when they have a large number of tests to correct. Brown (2003: 21) explains that a teacher's assessment of the first few students' papers among 40 papers, for instance, would definitely be different from that of the last few ones. In this case, the teacher may lack intra-rater reliability due to "unclear scoring criteria, fatigue, bias toward particular "good" and "bad" students, or simple carelessness." (Brown, 2003: 21)

2.10.2.3. Test Administration Reliability

Brown (2003: 22) suggests that the test administration is a main factor affecting an assessment. An aural test that is given in a noisy place, for example, will be unreliable since there will be several factors that affect the student's comprehension.

2.10.2.4. Test Reliability

The reliability of the test can be affected by other factors like time limit. If the test is long, students may get tired at the end of the test and may perform

less in comparison with their performance at the beginning of the test. (Brown, 2003: 22)

2.10.3. Validity

Validity is defined as "the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment." Gronlund (1998: 226, cited in Brwon, 2003: 22) In simple words, it is the extent to which the results of the assessment reflect what is being assessed. (Genesee and Upshur 1996: 62)

There are three types of validity that are important in the design of any assessment tool. They are:

2.10.3.1. Content Validity, or content-related evidence, which refers to the extent to which the response of a student reflects the subject matter being assessed and whether or not the tool of assessment adequately samples that content. (Brown, 2003: 23; Moskal and Leydens, 2000: n.p.)

2.10.3.2. Construct Validity: Brown (2003: 25) defines a construct as "any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions." To simplify, he gives the following example. A teacher has to manage an oral interview with a scoring analysis based on pronunciation, fluency, grammatical accuracy, vocabulary use, and socio-linguistic appropriateness. These five factors are justified by a theoretical construct to be major components of oral proficiency. So, if the teacher conducts an oral proficiency interview that evaluates only two of those factors, then the test is suspicious about construct validity.

2.10.3.3. Criterion Validity, or criterion-related evidence, which refers to "the extent to which the result of an assessment correlates with a current or future event... [and] the extent to which the student's performance on

the given task may be generalized to other, more relevant activities" (Rafilson, 1991, cited in Moskal and Leydens, 2000: n.p.) In this sense, there are several tests that are designed to assess specific performance and cannot be generalized to measure other performances.

The relation between validity and reliability is very important and close. This is confirmed by Genesee and Upshur (1996: 63) who state that an "inconsistency in a measurement procedure reduces validity." Also, they assert that validity is the most important principle among practicality and reliability.

2.10.4. Authenticity

Benchman and Palmer (1996:23, cited in Brown, 2003: 28) define authenticity as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task". To be authentic, learning tasks should simulate real-world tasks. Brown (2003:28) adds that items sequenced with no relationship to each other lack authenticity.

2.10.5. Washback (or Feedback)

Washback is one facet of validity that refers to the outcomes gained from testing and its effect on the learning and teaching processes (Hughes, 2003:1, cited in Brown, 2003: 28). The term 'washback' is coined in such a way since the information given by students assessment "'washes- back' to students in the form of useful diagnosis of strengths and weaknesses." (Brown, 2003: 29)

For the feedback to be valuable, it should be given continuously to students within the process of learning. Brookhart *et al* (2009:53) emphasize that whether positive or negative, feedback can be beneficial to students if it is given in a timely manner. However, although giving feedback is a main point in alternative assessment, it may be criticized as taking time, particularly by

teachers under the pressure of a standard curriculum to cover (Black and Wiliam, 1998: 12).

2.10.6. Assessment Goal

In addition to the previous five principles of assessment, still there is another aspect that is considered as essential in assessment, which is the purpose of the assessment. Stiggins (1992: 212) points out three purposes for teachers setting, administrating, and using assessment, they are "to inform specific decisions, to instruct, and to control student behaviour". In more practical terms, Brookhart *et al* (2009:58) describe the process of setting a goal of the assessment as making a promise to oneself.

2.11. Assessment of Speaking

Speaking is one of the most (if not the most) challenging skills to assess and "[n]o language skill is so difficult to assess with precision as speaking ability" (Harris, 1969: 81). Despite that difficulty, still there is a need for special attention and efforts from educators to reach a consensus on what are the most effective ways to follow in the assessment of speaking. Some ways are being investigated in the present study, for the purpose of which a question like '*what to assess in a speaking skill'* is to be answered in the following sections.

2.11.1. What is Speaking?

Speaking is defined as an "activity requiring the integration of many subsystems… [that] combine to make speaking a second or foreign language a formidable task for language learners..." (Baily and Savage, 1994, cited in Lazarton, 2001: 103). Understanding those 'subsystems' is essential in identifying the criteria used in the assessment of the speaking skill.

2.11.2. Nature of Speaking

According to the definition of speaking mentioned earlier, speaking is a complex of components that develops along with the learning process. Accordingly, in designing assessments of the speaking skill, it is important to identify those components as the criteria of the performance assessment.

The basic components of the speaking skill, as listed by Harris (1969:81- 2), are:

1. *Pronunciation*: including segmental (vowel and consonant sounds) and suprasegmental features (stress and intonation patterns)
2. *Grammar*
3. *Vocabulary*
4. *Fluency* ("the ease and speed of the flow of speech"), and
5. *Comprehension*: Harris states that comprehension is a component of the speaking skill "for oral communication certainly requires a subject to respond to speech as well as to initiate it."

2.11.3. Speaking Components: the Criteria of Assessment

As presented earlier, defining the type of assessment used for the speaking skill depends upon the components of speech. Though some of such components are found within the writing skill, their assessment is quite different in speaking assessment as speaking differs in more than one aspect from writing. In this respect, Luoma (2004) details the components of speaking and how to treat each one in the assessment.

1. **"The Sound of Speech":**

    In this section, Luoma defines pronunciation as "many features of the speech stream, such as individual sounds, pitch, volume, speed, pausing, stress and intonation." (Luoma, 2004:9). However, he raises two essential questions regarding pronunciation as an assessment criterion of

speaking. The first question is whether all of the previously mentioned features can be covered under one rating criterion, and the second is "should the focus be on accuracy of pronunciation or expressiveness of the speaker's voice, or both?" (Luoma, 2004:11)

To solve such a difficulty, Luoma proposes that if the speaking rating scale includes many other criteria beside pronunciation, then the only option is to fit both "accuracy and effectiveness" under one criterion like, for instance, "naturalness of pronunciation". Moreover, Luoma asserts that assessing speaking depends on the context and the purpose behind the assessment. Thus, according to the purpose, the focus of assessing pronunciation may be on 'accuracy', 'comprehensibility', 'interactional efficiency' (that tests the ability to create meaning in discourse), or 'expressiveness' (which covers 'general texture of the talk, speed and pausing, and variety in pitch, tone and volume). (ibid)

2. **"Spoken Grammar"**: taking speech as a social activity, Luoma proposes "... the grammar that is evaluated in assessing speaking should be specifically related to the grammar of speech." (ibid:12). Consequently, he lists the following points to be considered in a speaking assessment rating scale:

a) *"Written sentences, spoken idea units"*

While written form is based on sentences, speech is characterized by 'idea units'. These "are short phrases and clauses connected with *and, or, but, or that*, or not joined by conjunctions at all but simply spoken next to each other, with possibly a short pause between them." (ibid).

b) *"Grammar in planned and unplanned speech"*

A consideration of whether the speech is planned (as in "speeches, lectures, conference presentations, and expert discussions") or unplanned (as spoken at the moment in a reaction to an interlocutor) is

important in assessment. This supports the fact that in planned speech, which tends to be relatively formal, sentences are usually complete in a written-like format while in unplanned speech, which may be formal or informal, the possibility of uttering short ideas and incomplete sentences occurs commonly. (ibid:13)

c) *"The internal structure of idea units"*

Certain structures (like topicalization and tails) help making speech more natural and interpersonal. If students use such structures, they could be rewarded for it, however not using them does not put students to punishment since they are not obligatory (Luoma, 2004:16). While topicalization gives some emphasis on an initial element in a clause, like *"That house in the corner, is that where you live?"*, tails emphasize a point made at the beginning of a clause at its end (ibid). Examples of this are (*he's quite a comic, that fellow,* and *you know*).

   To summarize the spoken grammar section, Luoma declares, "speech is organized into short idea units, which are linked together by thematic connections and repetition as well as syntactic connectors" (ibid), like "*and, or, but, etc*.".

3. **"Words, Words, Spoken Words"**

   Usually speaking rating scales that include vocabulary use as one criterion of assessment define it in the highest level as it enables the speaker to express himself adequately and gives an evidence of the "richness of one's lexicon." (ibid) However, in authentic situations "very 'simple' and 'ordinary' words are also very common" (ibid). So, the question is, what is the basis of vocabulary assessment in speech.

   In this respect, Luoma (ibid) lists the following points under the

heading 'words':

a) *Specific and generic words*

In normal speech, speakers use many generic words, like "*this one/that one, the round thing*, ..." which help them proceed regardless of words that may be missed (ibid:17).

b) *Fixed phrases, fillers and hesitation markers*

These are certain words, phrases, and strategies that help in creating time to formulate what the speaker wants to say, like *'ah, you see, kind of, sort of,* and *you know'*, or for other purposes than creating time, including fixed responses like *'I thought you'd never ask'* or *'I'm doing all right'*, and some frames which allow various terms to fill one or two of its slots, like '*What a nice thing to say, What a horrible thing to say*'. Such expressions, if used by the speaker, may be rewarded in the assessment of his/her speaking performance (ibid:17).

c) *Word use in studies of assessing speaking*

In investigating studies that are relevant to the previously mentioned features of speaking in speaking assessment, it is found by Luoma that they were few, specifically three in number (they are Towell *et al*. (1996), Nikula (1996), and Hasselgren (1998) (Luoma, 2004:18-9)). However, one core principle was shared among the findings of all the three studies. Those fixed expressions explained above- which are referred to by Towell *et al* as "lexical phrases", by Hasselgren as "small words", and by Nikula as "pragmatic force modifiers"- are found to be an indicator in the listener's perspective about a fluent speaker. In other words, fixed expressions are to be considered as one criterion of fluency. In the third study, when Nikula studied the speech of the nonnative speakers of her language in their mother tongue, she found that such a phenomenon was not personal or

out of communication style, but it was totally related to language ability. The participants of her study couldn't use such expressions because they lack the linguistic ability in the target language.

d) *Slips and errors*

Slips and errors, such as "mispronounced words, mixed sounds, and wrong words", are normally included in the speech of native speakers. However, while native speakers are pardoned for their slips and errors because they "know", those of second and foreign learners become significant, and may be accused of "lack of knowledge". In this concern, Luoma states that raters should receive a special training to develop "a possible tendency to count each "error" that they hear." (Luoma, 2004:19)

e) *Comprehension*

Though not listed in Luoma's components of the speaking skill, still there is another component that is essential in speaking assessment (as stated earlier by Harris, 1969: 82), that is comprehension. In many contexts, speaking cannot stand alone without listening being involved. Thus, unless giving a speech or an oral presentation, the speakers' performance is highly affected by what they listen to which lead them –in a way- to interact with their interlocutors. For this reason, Brown (2003: 140) denotes that "[f]rom a pragmatic view of language performance, listening and speaking are almost always interrelated", and it is very difficult to isolate speaking tasks from aural comprehension. Therefore, it is important to include comprehension as one criterion in speaking interaction when the objective of teaching is to make students communicate interactively.

2.11.4. Teaching Speaking

Since the teaching process is interrelated with the assessment process, then it is important to tackle some main points regarding teaching speaking. The following points are devoted to such purpose.

2.11.4.1. Principles of Teaching Speaking

In order to decide what (and how) to teach in speaking classes, Lazarton (2001: 104) lists four main questions for a teacher to consider. They are:

1. *Who are the students?* (i.e. the level of the students being taught)
2. *Why are they there?* (i.e. the purpose behind the course)
3. *What do they expect to learn?* (i.e. the final goal of learning)
4. *What am I expected to teach?* (i.e. teaching activities and tasks)

The level of students and the goal of the course are essential factors in specifying what kind of methodology, techniques, activities and tasks (let alone assessment) to apply and use in the classroom. For non-academic adults, teachers concentrate on "survival English and basic communication functions" (ibid), while with the academic ones (just as the population targeted in the present study), "practice in activities such as leading and taking part in discussions and giving oral reports" should be the focus (ibid). Accordingly, the techniques and activities adopted in the empirical part of this study focus on group discussions and training to give conclusions of discussions orally.

2.11.4.2. Teaching Speaking in the Communicative Approach

Among the various approaches and methods applied in language teaching, the communicative approach seems to be the dominating one. Nevertheless, applying it inside the classrooms may vary from one teacher to another (Richards, 2006: 2). According to the communicative approach, the basic target of the teacher is to enable the students to acquire communicative competence. The communicative competence, in turn, encompasses four subcompetencies, they are:

1. ***Grammatical competence*:** covers the language forms and structure (including grammar, vocabulary, and phonology).

2. ***Sociolinguistic competence:*** rules that go beyond the linguistic forms to focus on social meanings of expressions.

3. ***Discourse competence*:** rules that deal with the "interconnectedness" of the text including its coherence (the organization and structure of a text) and the cohesion (how sentences are connected together to form larger texts).

4. ***Strategic competence:*** rules and strategies that help communication to proceed. (Lazarton, 2001: 104; Savignon, 2001: 17:18)

In more practical words, Richards (2006: 3) lists the following four aspects of language knowledge as being included within communicative competence:

- *Knowing how to use language for a range of different purposes and functions*

- *Knowing how to vary our use of language according to the setting and participants (e.g., knowing when to use formal and informal speech and when to use language appropriately for written as opposed to spoken communication)*

- *Knowing how to produce and understand different types of texts (e.g., narrative, reports, interviews, conversation)*

- *Knowing how to maintain communication despite having limitation in one's language knowledge (e.g., through using different kinds of communicative strategies)*

Accordingly, to help improving students' performance in oral communication classes, teachers of speaking courses are required to help students practice those components which they learn from other courses (such as grammar, pronunciation, comprehension, etc.) and to teach them

communication and conversational strategies. In conclusion, conversation classes may be looked at as practice courses more than being learning ones. In this case, teachers of conversation are highly encouraged to let students take part in daily discussions, that is to say, to let them speak. However, there are several difficulties that may challenge the teacher, especially when teaching speaking to EFL learners in a non-English environment. In such cases, all the students share the same mother tongue and lack practicing English except inside the classroom. Nunan (1993, cited in Lazarton, 2001: 110) clarifies that the challenges that may face an EFL teacher in such homogenous EFL classes are:

1. *Lack of motivation*
2. *Getting students to speak without being called on by name*
3. *The use of the first language*
4. *Very often large classes*
5. *Curriculum concentrating on non-speaking activities*
6. *Non native teachers*

Lazarton (2001: 110) emphasizes that in order to face such difficulties in speaking classes, the "EFL teachers need to be particularly adept at organizing class activities that are authentic, motivating, and varied".

2.11.4.3. Types of Speaking

The type of speaking chosen in a speaking course is determined according to the goal of the course and the level of the students. Following that, the activities used in the classroom will vary due to the same conditions. In this concern, Brown (2003: 141-142) sums up five main types of speaking, they are:

1. *Imitative:* repetition of words, phrases, or simple sentences with concentration only on pronunciation regardless to comprehensibility or conversational interaction.

2. ***Intensive:*** production of short structures reflecting a limited linguistic competence. Intensive speaking tasks may include sentence or dialogue completion, limited picture-cued task in simple sequences, etc.

3. ***Responsive:*** limited interactions including "... very short conversations, standard greetings and small talks, simple requests and comments, and the like."

4. ***Interactive:*** interactive speaking differs from responsive speaking in length and complexity of interaction. Moreover, interaction can be of two formats: **transaction** language, which has the purpose of exchanging specific information (like, *A: What time is it now? B: It is 10.30 am*) and **interpersonal** exchanges, which aims at maintaining social relationships. (Like, *A: How are you doing? B: I am fine, and you?*)

5. ***Extensive (monologue):*** Includes "... speeches, oral presentations, and story-telling" in which there is no interaction from the side of the listener or a very limited one. Such tasks are usually preplanned and formal.

2.11.4.4. Elements of Speaking

For learners to be fluent and effective speakers, Brown (2004: 269-71) proposes that they should have both knowledge of language features and the ability to process information and language 'on the spot'. In this regard, he lists two main categories to explain what is meant by language features and language processing. All those features can contribute in identifying the criteria of speaking performance.

1. **Language features:** This covers four main headings which are:
   a) ***Connected speech:*** fluent speakers should be able to produce connected speech forms, like 'assimilation', 'elision', 'linking r', and 'stress patterns'. Such forms should be included in speaking activities

to help students acquiring and mastering them.

b) ***Expressive devices:*** use of pitch and stress, varying volume and speed, and nonverbal (paralinguistic) expressions.

c) ***Lexis and grammar:*** common lexical phrases are important in spontaneous speech. So, teachers should supply their students with a variety of such phrases of different functions, like 'agreeing and disagreeing', 'expressing', 'surprise', 'shock', or 'approval'.

d) ***Negotiation language:*** 'Negotiation' expressions that are used to seek clarification are significant for students to learn. So, teachers should provide students with such expressions that help them interact in the learning environment, like:

- *(I'm sorry) I didn't quite catch that.*
- *(I'm sorry) I don't understand.*
- *What exactly does X mean?*
- *Could you explain that again, please?* (Brown, 2004:270)

Brown adds that students should be taught expressions that help them organize the content structure of their speech in a more 'written-like' format, such as in giving oral presentation. Certain phrases may include:

- *The important thing to grasp is that ...*
- *To begin with/ And finally ...*
- *What I am trying to say is that ...*
- *What I mean is ...*
- *The point I am trying to make is that ...*
- *... or, to put it another way ...,* (Brown, 2004: 270)

2. **Mental/Social Processing:** This category includes the following: (Brown, 2004: 270)

a) ***Language Processing:*** the ability to retrieve words and phrases from the memory and organize them appropriately.

b) ***Interacting   with   Others:*** interaction   involves   listening, comprehending others, and taking turns.

c) ***(On-the-spot) Information Processing:*** the ability of a speaker to process the information being told in the moment it is told. Brown (ibid) states that such a response is culture-specific and is not prized in many language communities.

CHAPTER THREE: RUBRIC AND PEER ASSESSMENT

3.1. Introduction

In order to give a comprehensive view concerning the new assessment practices applied in the study, this chapter is devoted to introducing rubrics and peer assessment in detail. It covers the meanings of the two practices with associated principles and the advantages and disadvantages of each.

3.2. Definition of Rubric

As the practices of alternative assessment have 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.

Stevens and Levi (2005: 3) cite the definitions of the term 'rubric' from two sources:

a. \Ru"bric\, n. ... that part of any work in the early manuscripts and typography which was coloured red, to distinguish it from other portions... (Webster's Unabridged Dictionary, 1913).

b. Rubric: n1: authoritative rule 2: an explanation or definition of an obscure word in a text [syn: gloss] 3: a heading that is printed in red or in a special type ... (WordNet, 1997)

Today's rubric has gained several related definitions in the field of assessment and evaluation which have nothing to do with the colour 'red'. Popham (1997: 2) defines the term rubric as "a scoring guide used to evaluate the quality of students' constructed responses". Brualdi (1998: 2) mentions that "a rubric is a rating scale by which teachers can determine at what level of proficiency a student is able to perform a task or display knowledge of concept." Andrade (2001: 1) states that "[a]t their very best, rubrics are also teaching tools

that support students [SIC] learning and the development of sophisticated thinking skills." In addition, Reddy (2007: 4) declares that "rubrics are assessment tools, which facilitate the process of evaluation and reporting of students' achievement by educators".

Being one of the advocates of rubric use, Andrade (2001 and 2005) differentiates between scoring rubrics and instructional rubrics. While scoring rubrics are those used by teachers merely to assign grades to 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 teaching and assessment, instructional rubrics are highly encouraged for better educational results, and they are defined as a document of one or two pages "that describes varying levels of quality, from excellent to poor, for specific assignment" (Andrade, 2000:1). This last definition is adopted in the present study to get the most of rubrics' advantages.

3.3. Composition of a Rubric

Teaching through the use of rubrics needs an experience and knowledge on how to create them 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 using method) affect and decide the value of any rubric. In this regard, Andrade (2005:27) outlines her perspective by saying: "I have found that whether they [rubrics] are good, bad, or even ugly depends on how they are created and how they are used."

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 whether holistic or analytic (to be tackled in a later section).

   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: 2006: 41).

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

   1. Determine the learning outcomes for the assignment by asking the following questions:
- *What is the intended learning that is to occur?*
- *How can such learning be measured?*
- *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 (CITL, 2007:3).

Table (3.1) illustrates the dimensions of a basic rubric format.

Table (3.1) Basic Format of a Rubric

Title of the rubric				
Description of the task being assessed				
List of criteria	Criterion Level 1	Criterion Level 2	Criterion Level 3	Criterion Level 4
Criterion 1	*Level description*	*Level description*	*Level description*	*Level description*
Criterion 2	*Level description*	*Level description*	*Level description*	*Level description*
Criterion 3	*Level description*	*Level description*	*Level description*	*Level description*

It must be mentioned that Table (3.1) 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.

The hard part in composing a rubric is setting the quality definitions, in that they should be clear enough to be understood by all raters (including students when used in self- or peer assessment). To define the quality definitions, RDG (ibid) 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.*

3.4. Types of Rubric

Scoring rubrics are of two types, holistic and analytic.

1. **Holistic rubrics:** In holistic rubrics, the criteria being evaluated are considered totally in combination and an overall judgement is made on a single descriptive scale (Popham, 1997: 72; Moskal, 2000: 4; Scott, 2006: 41). Table (3.2) shows an example of a holistic rubric (adopted from Stevens and Levi, 2005: 122).

Table (3.2) An Example of (Two Pages) Holistic Rubric

Inquiry and Critical Thinking Rubric

Students will learn various modes of inquiry through interdisciplinary curricula—problem posing, investigating, conceptualizing—in order to become active, self-motivated, and empowered learners.

6 (Highest)—Consistently does all or almost all of the following:

- Accurately interprets evidence, statements, graphics, questions, etc.
- Identifies the salient arguments (reasons and claims) pro and con.
- Thoughtfully analyzes and evaluates major alternative points of view.
- Generates alternative explanations of phenomena or event.
- Justifies key results and procedures, explains assumptions and reasons.
- Fair-mindedly follows where evidence and reasons lead.
- Makes ethical judgments.

5—Does most of the following:

- Accurately interprets evidence, statements, graphics, questions, etc.
- Thinks through issues by identifying relevant arguments (reasons and claims) pro and con.
- Offers analysis and evaluation of obvious alternative points of view.
- Generates alternative explanations of phenomena or event.
- Justifies (by using) some results or procedures, explains reasons.
- Fair-mindedly follows where evidence and reasons leads.

4—Does most of the following:

- Describes events, people, and places with some supporting details from the source.
- Make connections to sources, either personal or analytic.
- Demonstrates a basic ability to analyze, interpret, and formulate inferences.
- States or briefly includes more than one perspective in discussing literature, experiences, and points of view of others.
- Takes some risks by occassionally questioning sources or by stating interpretations and predictions.
- Demonstrates little evidence of rethinking or refinement of one's own perspective.

Continued

3—Does most or many of the following:

- Responds by retelling or graphically showing events or facts.
- Makes personal connections or identifies connections within or between sources in a limited way. Is beginning to use appropriate evidence to back ideas.
- Discusses literature, experiences, and points of view of others in terms of own experience.
- Responds to sources at factual or literal level.
- Includes little or no evidence of refinement of initial response or shift in dualistic thinking.
- Demonstrates difficulty with organization and thinking is uneven.

2—Does most or many of the following:

- Misinterprets evidence, statements, graphics, questions, etc.
- Fails to identify strong, relevant counter arguments.
- Draws unwarranted or fallacious conclusions.
- Justifies few results or procedures, seldom explains reasons.
- Regardless of the evidence or reasons, maintains or defends views based on self-interest or preconceptions.

---

1 (lowest)—Consistently does all or almost all of the following:

- Offers biased interpretations of evidence, statements, graphics, questions, information, or the points of view of others.
- Fails to identify or hastily dismisses strong, relevant counterarguments.
- Ignores or superficially evaluates obvious alternative points of view. Argues using fallacious or irrelevant reasons and unwarranted claims.
- Does not justify results or procedures, nor explains reasons.
- Exhibits close-mindedness or hostility to reason.

X—No basis for scoring (Use only for missing or malfunctioning portfolios.)

(Stevens and Levi, 2005: 122).

2. **Analytic Rubrics:** analytic rubrics focus on "more specific aspects of performance" (Scott, 2006: 41) and require the scorer to render criterion-by-criterion scores..." (Popham, 1997: 72). Table (3.3.) is an example of an analytic rubric for class participation (adopted from Anderson, 2003: 101).

Table (3.3) An Example of an Analytic Rubric

Rubric for Class Participation

Criteria	Point Values			
	4	3	2	1
Attendance/promptness	Student is always prompt and regularly attends classes.	Student is late to class no more than once every 2 weeks; regularly attends classes.	Student is late to class more than once every 2 weeks and/or skips a class or two.	Student is late to class more than once a week and/or frequently skips class.
Preparation	Student is almost always prepared for class with assignments and required materials.	Student is usually prepared for class with assignments and required materials.	Student is rarely prepared for class with assignments and required materials.	Student is constantly unprepared for class with assignments and required materials.
Behavior	Student almost never displays disruptive behavior during class.	Student rarely displays disruptive behavior during class.	Student occasionally displays disruptive behavior during class.	Student constantly displays disruptive behavior during class.
Level of engagement	Student contributes to class by offering ideas and asking questions on a regular basis.	Student occasionally contributes to class by offering ideas and asking questions.	Student rarely contributes to class by offering ideas and asking questions.	Student never contributes to class by offering ideas and asking questions.

(Anderson, 2003: 101)

3.5. Advantages and Disadvantages of Rubrics

Rubrics are not pure rating scales. Since they can merge teaching 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; those are tackled in the present section.

Andrade (2005) discusses all the aspects of rubrics, "the good", "the bad" and "the ugly" ones. For their significance to the present study, here is a summary. (Andrade, 2005: 27-30)

A. **The Good:** instructional rubrics are **good for teachers** 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.

In addition, instructional rubrics are **good for students** because they:

1. Tell them their teachers' goals, and so they can focus their efforts on meeting those goals. Applying that, Andrade states, "I never hear a student complain that she 'didn't know what I want'".
2. If used in self-and-peer assessment (that not to be counted in final grades), they can 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 might 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 their judgement might be "cruel or disorienting" of their peers or "misleading or delusional" in their self-assessment. At this point, students need training on both assessments.

C. **The Ugly:** In their worst descriptions, rubrics might 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 judgements.

3. **Fairness**: issues of "gender, race, ethnicity, or socioeconomic status" might affect the use of rubrics.

For the sake of improving a rubric, Andrade advises teachers to compare it with "published standards", consult another teacher, or ask a colleague to co-assess the performance of the same students. Added 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 [SIC] critique assessments can be the best source of information about how good evaluations really are."

Moreover, Scott (2006, 41-42) sums up the advantages and disadvantages of rubrics to students and teachers in two lists of points. Here is a summary:

a) *The advantages of rubrics*

- Assessments become more "objective and consistent".
- Raters focus on grading "the important outcomes".
- The expected outcomes are clarified with their different values.
- Students know their strengths and weaknesses with directions to improving the weaknesses.
- Clearing the demanded performance to teachers that enable them to clearly explain its criteria to students.
- Making students aware of the criteria assessed in their performance, whether by teachers, their mates, or by themselves.
- Emphasis on formative assessment that leads to summative assessment.
- They provide a scale for measuring and reporting progress
- They lower students anxiety of the expected outcome
- Ensure an objective judgement of students' performance
- They promote students' performance

b) *The disadvantages of rubrics:*

- Rubrics construction and use can be time consuming
- There is a difficulty in setting clear criteria descriptors for the different levels of each criterion

3.6. 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. Topping (ibid: 249) highlights several objectives out of his review, they are:

- to determine the extent, nature and quality of the literature to date;
- to develop a typology of peer assessment;
- to explore the theoretical underpinnings of peer assessment and elucidate the mechanisms through which it might have its effects; and
- to outline directions for future research and practice.

In order to cover the preceding objectives, Topping included 109 papers focusing upon peer assessment among students in higher education. Out of the literature he reviewed, Topping highlights some main parameters of variation in the typology of peer assessment (ibid: 251). The following table illustrates them.

Table (3.4) A Typology of Peer Assessment in Higher Education

No.	Variable	Range of Variation
1.	Curriculum Area/Subject	All
2.	Objectives	Of staff and/or students? Time saving or cognitive/affective gains?
3.	Focus	Quantitative/summative or Qualitative/formative or both?
4.	Product/Output	Tests /marks/grades or writing or oral presentations or other skilled behaviours?
5.	Relation to Staff Assessment	Substitutional or supplementary?
6.	Official Weight	Contributing to assessee [SIC] final official grade or not?
7.	Directionality	One-way, reciprocal, mutual?
8.	Privacy	Anonymous/confidential/public?
9.	Contact	Distance or face to face?
10.	Year	Same or cross year of study?
11.	Ability	Same or cross ability?
12.	Constellation Assessors	Individuals or pairs or groups?
13.	Constellation Assessed	Individuals or pairs or groups?
14.	Place	In/out of class?
15.	Time	Class time/free time/informally?
16.	Requirement	Compulsory or voluntary for assessors/ees?
17.	Reward	Course credit or other incentives or reinforcement for participation?
		Topping (1998: 252)

In conclusion to the review, several findings concerning the application of peer assessment were mentioned by Topping. Here is a summary:

1. There is positive impact of peers' feedback on improving students' grades.

2. Reliability and validity of peer assessment are achieved in various subject areas.

3. Peer assessment does not gain acceptability among all students.

4. Although peer assessment requires efforts from students, it is effective in reducing anxiety.

5. In addition to raising the learning performance of students, peer assessment raises their confidence.

6. Studies on peer assessment of writing skill proved it to be similar to (in some studies better than) teachers' assessments.

7. Studies on peer assessment of group and project work gained positive results.

Finally, 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) The latter view about the victimized students may be reflected in traditional testing that occurs at the end of the learning process. 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."

3.7. Implementation of Peer Assessment

For a successful implementation of peer assessment, different settings and qualities have been cited by Topping (1998: 265- 267) out of the extensive literature he reviewed. They are summarized below:

1. Expectations, objectives and acceptability need to be clarified to students in advance.

2. Since little literature gave interests to students' peering, students may be matched with peer assessors whom they find suitable, with their friends, or randomly.

3. Assessment criteria must be clarified to students with examples, and students are highly encouraged to participate in identifying them.

4. Practical training for students is needed.

5. Expected outcomes should be clarified with examples in advance.

6. Peer assessment, especially of inexperienced students, should be under the observation of teachers.

7. Reliability and validity of peer assessment should be checked by teachers, from time to time, even if on a "random or a targeted sample".

8. Students' improvement can be compared with their levels prior to the application of peer assessment or with a control group.

3. 8. Advantages and Disadvantages of Peer Assessment

The implementation of peer assessment in classroom has several advantages as far as both students and the learning process are concerned. For students, in addition to their educational outcomes, peer assessment proved to have behavioural gains as well. In that sense, Topping (2012: 3) emphasizes 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." In support to their advantages in enhancing students' achievements, Stiggins and Chappuis (2005: 11) propose that evidences collected in this concern over decades worldwide proved that when teachers engage their students in peer assessment practices, great outcomes in the students' attainments are gained. However, 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 making a better understanding of materials.

In spite of the advantages of peer assessment and the evidence that it can be effective in learning, several disadvantages 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 (3.5) illustrates potential advantages and disadvantages of peer assessment (adopted from White, 2009: 5).

Table (3.5) Potential Advantages and Disadvantages of Peer Assessment

Advantages
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
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).
(White, 2009: 5)

Apart from its disadvantages, peer assessment still worths its application in classroom since it prove enhancing learners' achievements, which is a typical goal of any learning process. Black *et al.* (2003, 62, cited in White, 2009: 3) point out that the goal behind peer assessment is not only to give scores to the peers, rather it is to enlighten students of the learning needs and the suitable ways of improving them. However, to avoid some problems associated with peer assessment, as mentioned above, Wheater *et al* (2005:15) recommend schemes that "require openness in dialogue, good planning, and close

monitoring in the early stages." Besides, they encourage the use of peer assessment as they have noticed students' interests in it for it helped the latter enjoy the classes and understand the assessment more actively. Still, this is not a call to let peer assessment dominate the courses, but it can be used in "all levels in an integrative assessment strategy for degree courses" (Wheater, 2005: 15)

3.9. Literature Review

A considerable amount of literature has been published on rubrics use and self-and/or- peer assessment. In this concern, the researcher made an online search through some main educational libraries and databases including IVSL 'Iraqi Virtual Scientific Library', which opens the gate to several significant databases covering humanities studies like, Science Direct, JSTOR, Project MUSE, Springer, and Citeseer, and the world largest digital library ERIC 'Education Resources Information Center'. The search revealed enormous 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. Furthermore, it was found that the use of rubrics and self and/or peer assessment is used in almost all levels of education, starting from kindergarten to higher education.

While all the studies focused on the use of rubrics, the purpose behind them varied. Some studies investigated the effects of rubrics and peer assessment on students' learning and the improvement of their 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. Yet, only works of interest to the present study are discussed in this section (focusing only on rubric use and peer assessment with no mention to self-assessment, regardless to the various literature investigating it).

To begin with, Reddy and Andrade's review (2010) can be introduced since it covers a number of the resources appeared in the above-mentioned online search. In 2009, Reddy and Andrade conducted a review of rubric's use in higher education including 20 empirical studies and doctoral theses. 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). The purpose behind the review was to explore:

1. the kind of research made on rubrics in higher education,

2. whether rubrics can promote learning rather than merely assess it, and

3. how much importance is given to rubrics in terms of validity, reliability, and fairness.

The first type of studies reviewed is associated with the use of rubrics with curriculum. Four studies (Powell 2001; Dunbar, Brooks and Kubicka-Miller 2006; Knight 2006; Song 2006; cited in Reddy and Andrade, 2010: 444) concentrated on the use of rubrics in the improvement of course delivery and design and did support the use of rubrics for such purposes. Despite the fact highlighted by Reddy and Andrade that little attention was given to the use of rubrics in program assessment, a study by Petcov and Petcova (2006, cited in Reddy and Andrade, 2010: 444) proposed that they are effective tools in this field. The second type of studies focused on students and teachers' perception of the use of rubrics. Four studies (Powell 2001; Reitmeier, Svendsen and Vrchota 2004; Andrade and Du 2005; Schneider 2006, cited in Reddy and Andrade, 2010: 444) have generally showed positive attitudes of students and teachers, while other two studies showed a resistance of teachers to using them (Bolton 2006; Parkes 2006, cited in Reddy and Andrade, 2010: 444). Nevertheless, Reddy and Andrade state that the resistance of those teachers is attributed to

that the "'overwhelming majority' of instructors have little or no preparation as teachers, and minimal access to new trends in assessment", an overview suggested by Hafner and Hafner (2003: 1510, cited in Reddy and Andrade, 2010: 444). Moreover, the latter accuse teachers of using rubrics as mere "scoring guides" and encourage using them as teaching guides to be shared with students in self and peer assessment to give the best results in improving their products, with this point Reddy and Andrade show their full agreement.

In answering the second 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.

The third issue in the review investigated the amount of research done on the quality of rubrics in higher education. Three studies (Simon and Forgette-Giroux 2001; Hafner and Hafner 2003; Dunbar, Brooks, and Kubicka-Miller 2006, cited in Reddy and Andrade, 2010: 445) concentrating on interrater reliability reflected that rubrics can help in reaching the same explanation to students' performance. However, one important point is raised that raters should be trained to get an acceptable level of reliability reaching 70% (or higher) of agreement. Similarly, three other studies (Moni, Beswick, and Moni 2005;

Green and Browser 2006; Lapsley and Moody 2007, cited in Reddy and Andrade, 2010: 445) tackled the issue of validity and proved that the clear and appropriate language of a rubric is important to ensure validity. Moreover, Reddy and Andrade found that other issues of validity- content validity and criterion validity- have not been studied at all.

Concerning the studies investigating the use of rubric and peer assessment, the following literature lists only studies related to the scope of the present study, i.e. the use of rubrics and peer assessment in speaking/oral assessment, students and teachers' perception of rubrics use, and finally studies discussing issues of interrater reliability.

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 gap, 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 the students' speaking performance; it achieved reliability between the two raters, and its descriptors were clear. Nevertheless, Naksuhara suggests using the rubric with a greater number of raters to get a more effective test of reliability.

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 to 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 were 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 of students' English abilities; however, he calls for raters training to get more reliability in assessment.

As for students and teachers' attitudes towards rubrics and peer assessment are concerned, two studies are to be reviewed. The first study was conducted by White (2009) in order to examine the students' perception of peer assessment and its impact on their learning, White conducted a study applying 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, and 1; from the 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 the 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. At the end of the course, White engaged the students in a survey to find out their opinions in practicing the peer assessment. The results of the survey revealed positive views of the students as a whole and that peer assessment indeed led to the promotion of their learning.

The second study was conducted by Kutlu *et al* (2010) in Turkey. The researchers made a survey study with 292 primary school teachers to examine the latters' positive and negative attitudes towards the use of rubrics. An attitude scale for scoring rubric use (based on a previous study by Kutlu *et al* 2009, cited in Kutlu *et al* 2010: 1567) was presented to the participant teachers. It included terms related to levels of teachers' knowledge about rubrics, sources from which they obtain rubrics, their frequency of using rubrics, their ways of preparing rubrics, the purposes behind using them in a course, and how they use rubrics in course activities.

In collecting the survey results, Kutlu *et al* divided the teachers into two types, teachers with positive attitudes and teachers with negative attitudes. Analysing the results, they found out that the teachers with positive attitudes formed 54% of the subjects while negative attitudes teachers formed 46%. The rubric and the comments that the teachers added to the survey helped in interpreting the results. Firstly, teachers of both attitudes had knowledge about rubrics though teachers with positive attitudes were more knowledgeable about them. Secondly, both teachers used rubrics in classroom activities. Thirdly, teachers with positive attitudes tend to design their own rubrics rather than depending on available ones as teachers with negative attitudes mostly do. The last finding, which goes along with Reddy and Andrade's (2010) concerned the purpose of rubrics and the way they are used. Kutlu et al revealed that while teachers with positive attitudes made use of rubrics for giving feedback for

students' performance and monitoring their learning and progress, teachers with negative attitudes used rubrics primarily as grading tools (Kutlu *et al.*, 2010: 1572).

One of the significant issues that challenge rubrics design and use is reliability and more specifically interrater reliability. In this regard, several studies have been undertaken primarily with the purpose of testing the interrater reliability of rubrics use.

A comprehensible review exploring the validity and reliability of rubrics and their impact on the improvement of students' learning and teaching was done by Jonsson and Svingby in 2007. The review covered 75 studies in various disciplines and distributed in all the levels of education, from kindergarten to higher education. The majority of the studies found to be performed in the last decade, while only seven of them were undertaken before 1997. "The distribution indicates that the rubric is a quite recent research issue" (Jonsson and Svingby, 2007: 132).

Among the studies reviewed, only a few number gave interest to the issue of intra-rater reliability ($N= 7$). In spite of that, they showed a "high internal consistency" in assessing the students' performance. On the contrary, interrater reliability studies were more than half the reviewed studies ($N= 46$) and most of them showed "sufficient" interrater reliability results. Concerning the issue of validity, one third of the studies ($N= 25$) concentrated on construct validity and revealed the need for more comprehensible works on the other issues of validity (namely, criterion validity and content validity). In addition to that, the review covered studies concentrating on whether rubrics can promote students' learning, self and peer assessment, students' perspectives of the use of rubrics, and whether or not rubrics can improve teaching. In conclusion, the following points are presented (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.

CHAPTER FOUR: MATERIALS AND METHODS

4.1. Introduction

To check the validity of the pre-defined hypotheses of the study, the following materials and methods are used and applied by the researcher.

1. *A questionnaire implemented on the teachers of conversation:* This questionnaire was important to establish background information for the study. Before introducing the questionnaire, the researcher was not quite aware of the conversation classes given in University of Basrah and Shatt Al-Arab Private University College , the general levels of the students, the teaching and assessment procedures followed by the teachers and their perspective regarding the subject matter and students as a whole. Thus, such a questionnaire was helpful in answering many questions in mind.

2. *A pilot study:* Prior to the main empirical work of the study, the researcher needed to conduct a pilot study focusing on the application of the analytic rubric inside the classroom, with a sample of the target participants in conversation classes. In fact, such a study was of primary significance and benefit to the researcher in getting training in the use of rubrics, how peer assessment works and techniques of teaching conversation, in addition to having a close observation of the students' performance in the classroom.

3. *Pretest- posttest procedure*: Those tests were applied to the students participating in the study to find out differences in their performance before and after applying the study materials and methodology.

4. *A students' questionnaire*: This questionnaire was carried out to examine the students' perspectives concerning the use of the scoring

rubric, peer assessment, and the group technique, which were applied along the period of the study in conversation classes.

5. *An interrater reliability test*. This was made with the help of a number of the faculty members in the Department of English, College of Education for Humanities, to test the interrater reliability of the scoring rubric developed by the researcher for the study. The test was supported by a traditional assessment to check which kind of assessment achieved more reliability at that moment.

4.2. Context of the Study

The study is conducted at the academic year 2011-2012 at the Department of English, College of Education for Humanities, University of Basrah . In Basrah, English language is taught as a specialization at the University of Basrah (the public university) and Shatt Al-Arab Private University College. At University of Basrah, English is taught at the Department of English in the College of Education for Humanities, and the two Department s of English and Translation in the College of Arts, while in Shatt Al-Arab Private University College, English language is a specialization only in the Department of English.

Conversation classes were selected among the other classes for the application of the study because they are the most suitable classes in which students can practice their speaking skill regardless of any difficulty of the subject materials.

4.3. Teachers' Questionnaire

In order to gather basic information concerning EFL learners' speaking skill and the teaching methodologies adopted by the teachers and their assessment of the students' performance in conversation classes, the researcher set a

questionnaire (Appendix I) engaging all the university teachers of conversation in Basrah at the academic year of the study.

4.3.1. The Jury of the Questionnaire

For the sake of ensuring the validity of the questionnaire, a jury of eight senior university teachers in the College of Education for Humanities, University of Basrah was consulted. Appendix (II) illustrates their academic qualifications and positions.

After making the necessary corrections and modifications suggested by the jury, the researcher consulted back the jury who approved the last edition of the questionnaire to be valid. After that, it was distributed among the teachers of conversation by the researcher at once.

4.3.2. Teachers and Classes of Conversation

As far as the University of Basrah is concerned, conversation classes are given to three stages at the Department of English, College of Education for Humanities (the first, the second, and the third stages), two stages at the Department of English, college of Arts (the first and the second stages), and one stage at the Department of Translation (the second stage). As for Shat Al-Arab Private University College, conversation classes are given to two stages (the first and the second stages). This schedule was for the academic year when this study was conducted.

The total number of teachers teaching conversational classes in the University of Basrah and Shat Al-Arab Private University College is nine distributed among the above-mentioned departments. Table (4.1) illustrates information concerning the number of the teachers of conversation and the stages being taught at University of Basrah   and Shatt Al-Arab Private University College during the Academic Year 2011-2012.

Table (4.1) Information of Teachers and Classes of Conversation Distributed among University of Basrah and Shatt Al-Arab Private University College during the Academic Year 2011-2012

No	Department	Number of teachers of conversation	academic qualifications	Stages having conversation classes
1.	Department of English, College of Education for Humanities, University of Basrah	5	MA	1st stage 2nd stage 3rd stage
2.	Department of English, college of Arts, University of Basrah	2	1PhD + 1 MA	1st stage 2nd stage
3.	Department of Translation, college of Arts, University of Basrah	1	MA	2nd stage
4.	Department of English, college of Arts, Shat Al-Arab University Private	1	MA	1st stage 2nd stage
	Total number of teachers	**9**		

4.3.3. Analysis and Discussion of the Teachers' Questionnaire

As presented earlier, the teachers taking part in answering the questionnaire are nine in number. Having surveyed their responses, a frequencies-analysis and a total percentage were applied for the sake of analyzing the questionnaire's findings. The following points tackle each item and discuss the purpose behind listing it in the questionnaire, and the teachers' responses.

Item 1: As being introduced in the principles of alternative assessment, the goals of the learning process are important in directing the teaching process. Hence, the first item of the questionnaire was about the objectives of the conversation classes presented by the teachers of conversation. The items listed within the objectives were taken from those prescribed by the Ministry of Higher Education and Scientific Research in the academic curricula of the Colleges of Education in Iraq (July, 2002). The results of item 1 are shown in Table (4.2) below.

Table (4.2) Frequencies of the Teachers' Questionnaire: Item 1

Objectives of Teaching	N	Percent of Cases
a) To teach students the principles of English language	2	22.2%
b) Help students understand the various relations among ideas.	2	22.2%
c) Help students understand the synonyms of words, phrases and meanings.	3	33.3%
d) Help students understand the functions of words in sentences as far as meanings and conditions are concerned.	2	22.2%
e) Help them understand some of the cultural, sociological, political and economical effects in language and literature.	2	22.2%
f) Help them understand the harmony between ideas and their arrangements, logic and classifications.	2	22.2%
g) Help them apply the new ideas they have acquired.	2	22.2%
h) Help them explain and interpret what they read or hear and the like.	3	33.3%
i) Help them choose the suitable meaning from a context.	1	11.1%
j) All the above.	5	55.6%
k) None of the above.	1	11.1%
l) Others.	1	11.1%
Total	26	288.9%

From the results shown in Table (4.2), it can be noted clearly that 55.6% of the teachers (*N*= 5) agree to have all the listed objectives as targets in their teaching. The responses of the other teachers varied among the twelve objectives. While objectives C and H gained a percentage of 33.3% (3 out of 9 teachers), objectives A, B, D, E, F, and G gained a percentage of 22.2% (2 out of 9 teachers), and objective I gained only 11.1% (only 1 teacher).

It seems that the teachers are concentrating in their objectives of teaching primarily on meaning. Objectives C and H direct the teaching process to

explanation of meaning and vocabulary, which may indicate that the majority of the teachers of conversation (the total percentage of teachers selecting these two objectives is 33.3%) concentrate in their teaching on meaning and vocabulary. The other objectives, since getting the same percentage of agreement among all the teachers, gained an equal preference by the teachers. Nevertheless, the only teacher who chose none of the above (i.e. option K) listed some objectives instead; they are given below (verbatim).

    a. *Using the English language in communicative situations.*
    b. *Learn to use different expressions used in real life situations.*
    c. *Know the different communicative sentences and expressions used by native speakers of English in different situations. "Listen to native people talking about and discussing different topics".*

Items 2 and 3: Items 2 and 3 tackle the issue of the stage variable. It was important to know if the teachers of conversation consider the stage level as a variable in their teaching and assessments of students' performance. Table (4.3) shows that 66.7% of the teachers (*N*= 6) consider the stage level of students as a variable in assessing the students' performance, while 22.2% of the teachers (2 out of 9 teachers) disagree with them.

Table (4.3) Frequencies of the Teachers' Questionnaire: Item 2

Item 2: If you teach conversation classes for more than one stage, do you consider the stage-level variable in your assessment of students' performance?		
Responses	Frequency	Percent
Valid Yes	6	66.7
No	2	22.2
Total	8	88.9
Missing 9.00	1	11.1
Total	9	100.0

As shown in Table (4.4) given below, 55.6% of the teachers (*No= 5*) agreed upon considering the level of students as a variable in setting the objectives of teaching, while 33.3% (*No= 9*) disagreed with them. However, only one teacher did not answer items 2 and 3, which may be due to teaching only one stage along the whole period in her/his teaching process of conversation.

Table (4.4) Frequencies of the Teachers' Questionnaire: Item 3

Item 3: If you give conversation classes to more than one stage, do your objectives differ from one stage to another?		
Responses	Frequency	Percent
Valid Yes	5	55.6
No	3	33.3
Total	8	88.9
Missing 9.00	1	11.1
Total	9	100.0

Item 4: In considering the stage level as a variable, the teachers were asked in item 4 if their objectives will vary accordingly. Consequently, though paradoxically, only the teacher who was responsible for teaching two stages at the time of the current study did not consider the stage level as a variable in her/his teaching and assessments. On the contrary, the teachers who teach only one stage showed a consideration of the stage variable in teaching and assessment.

Responses to item 4 varied in defining the objectives of each stage the teachers teach. Table (4.5) illustrates the teachers' responses of item 4.

Table (4.5) Frequencies of the Teachers' Questionnaire: Item 4

Item 4: The items numbers of the objectives the teacher decide for each stage			
Objectives	First Stage	Second Stage	Third Stage
1	22.2% (2)	11.1% (1)	
2		11.1% (1)	
3	22.2% (2)		
4	11.1% (1)	11.1% (1)	
5	11.1% (1)		
6	11.1% (1)		
7	33.3% (1)		Focus on fluency
8	11.1% (1)		
9	11.1% (1)		
10			
11			
12			
All the items		22.2% (2)	

What is worth noting here is that the objectives of teaching listed to students of the third stage focus upon fluency. The researcher, out of this question, needed to have an idea about the objectives in teaching conversation in particular to this stage.

**Item 5**: Having to design an analytic rubric focusing upon different criteria of the speaking skill, the researcher needed the teachers' views regarding the criteria to consider in the assessment of this skill. Thus, item 5 responses were important in this concern.

As shown in Table (4.6), all the teachers take into consideration fluency and interaction in assessing students' performance. Pronunciation and vocabulary also gained significant percentages, while the other criteria got the same number of responses. Additionally, the majority of the teachers take into consideration all the listed criteria in the assessment of the students' speaking skills. Still, there is the general impression that affects the teachers' assessments. 77.8% of the teachers (*N*= 7) chose the general impression as one criterion of assessment

which adds to the problem of subjectivity in assessment that characterizes the traditional assessment.

Table (4.6) Frequencies of the Teachers' Questionnaire: Item 5

Item 5: What are the criteria you consider in assessing the students' speaking skill?		
Criteria of Assessment	N	Percent of Cases
Pronunciation	8	88.9%
Vocabulary	8	88.9%
Fluency	9	100.0%
Comprehension	7	77.8%
Grammar	7	77.8%
Interaction	9	100.0%
all the above	7	77.8%
general impression	7	77.8%
Others	1	22.2%
Total	63	711.1%

In addition to the list of the criteria mentioned in item 5, one teacher added; "*expressing themselves (ideas) via different communicative skills*."

***Item 6*:** In scoring the criteria of the rubric, it is common to give one (or more) criterion a higher score than the other ones. In order to decide which criterion deserves a higher score, item 6 responses were devoted to this purpose. By asking the teachers to rank the criteria which they chose in item 5 according to their importance, the researcher can define which criteria are more valuable in the speaking skill (as far as the teachers' perspective is concerned).

Table (4.7) Frequencies of the Teachers' Questionnaire: Item 6

Item 6 :Please, list the criteria you mentioned in item no.5 according to their importance in the following grid (from the most to the least)

Criteria of Assessment	First Ranking		Second Ranking		Third Ranking		Fourth Ranking		Fifth Ranking		Sixth Ranking		Seventh Ranking	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Pronunciation	2	22.2%	2	22.2%	2	22.2%	1	11.1%	1	11.1%				
Vocabulary	1	11.1%	2	22.2%	3	33.3%	1	11.1%	1	11.1%				
Fluency	4	44.4%	2	22.2%					1	11.1%	2	22.2%		
Comprehension			1	11.1%	1	11.1%	3	33.3%	2	22.2%				
Grammar	1	11.1%	1	11.1%	1	11.1%	1	11.1%	1	11.1%	1	11.1%		
Interaction	1	11.1%	1	11.1%	1	11.1%	1	11.1%	2	22.2%	3	33.3%		
General Impression							2	22.2%					1	11.1%

___Item 7___**:** Item 7 deals with the period of assessment and how often the teachers assess the students' performance. Table (4.8) shows the results of item 7.

Table (4.8) Frequencies of the Teachers' Questionnaire: Item 7

Item 7: Is your assessment to students' performance daily, weekly, monthly, or semesterly?		
Period of Assessment	N*	Percent of Cases
daily assessment	4	44.4%
weekly assessment	3	33.3%
monthly assessment	1	11.1%
semesterly assessment	3	33.3%
Total	11	122.2%

a.   some teachers chose more than one way

Although the periods of assessment varied among the teachers, the manner of feedback seems to be the same among the teachers. The students receive their semesterly grades as a mere official feedback of their performance. When talking to the teachers and the students (in the Department of English, College of Education for Humanities) about this point, they explained giving feedback for the students' mistakes. Yet, no clear continuous feedback is given to the students along the period of the semester except the final grade at the end of the semester as a final evaluation of their performance. Besides, some students complained from not being able to know why they get specific grades while their friends get some higher ones. There were no clear reasons, as they told, for their evaluations and little guidance to how to improve their performance.

___Item 8___**:** Item 8 tackles the type of assessment adopted by the teachers of conversation. The two types of assessment, i.e. criterion-referenced assessment and norm-referenced assessment, which are introduced earlier in Chapter Two, are given as options in item 8. From Table (4.9), it is clear that the majority of the teachers adopt the criteria- referenced assessment with a percentage of 77.8% (7 teachers). In this regard, it is evident that the teachers concentrate on

building up the criteria of the speaking skill within the students' performance. Accordingly, the responses of item 8 support the responses of items 5 and 6.

Table (4.9) Frequencies of the Teachers' Questionnaire: Item 8

Item 8: What type of assessment you follow in your class?		
Type of Assessment	N	Percent of Cases
Criterion-referenced assessment	7	87.5%
Norm-referenced assessment	2	25.0%
Total	9	112.5%

***Item 9*:** Item 9 was listed to get some knowledge about the teachers' preferences in the use of any textbooks. Browsing several online language platforms and discussing with a number of teachers online show that the majority of EFL/ESL teachers are interested in online websites that offer extensive materials for teaching conversation and little interest in using textbooks. However, it was important to know how the teachers of conversation in University of Basrah and Shatt Al-Arab Private University College choose their teaching materials. Hence, as shown in Table (4.10), 66.7% of the teachers (6 teachers) showed interest in using textbooks (though no definite textbooks were defined by the authorities in the university) and 33.7% of them (3 teachers) did not seem to use any textbooks.

Table (4.10) Frequencies of the Teachers' Questionnaire: Item 9

Item 9: Do you make use of any textbooks?		
Responses	Frequency	Percent
Valid Yes	6	66.7
No	3	33.3
Total	9	100.0

***Item 10*:** As a support to item 9, item 10 elicits information about the different ways the teachers follow in choosing the topics of their classes. This is to give the researcher an idea about how the teachers organize their syllabi. Form Table (4.11), the teachers chose more than one way, the most common among which

is to choose the topics of their classes by themselves. Giving the students an opportunity in deciding the topics and following the course objectives came in second in the preference of the teachers. Moreover, although there is no defined syllabus of conversation in the departments involved, two teachers chose to follow the course syllabus in their teaching. However, one teacher added "*creativity*" to the four listed ways.

Table (4.11) Frequencies of the Teachers' Questionnaire: Item 10

Item 10: How do you choose the topics of the lectures? (you may choose more than one option)		
Choice of Topics	N	Percent of Cases
I choose them myself	8	88.9%
I ask the students what they like to talk about	6	66.7%
I follow the course syllabus	2	22.2%
I choose those which match the course objectives	6	66.7%
I have another way	1	11.1%
Total	23	255.6%

Item 11: To get a close view of the nature of conversation classes and how active the students are in those classes, items 11 and 12 were designed. Table (4.12) shows the results of item 11.

When asked about the nature of their classes, 66.7% of the teachers (6 teachers) answered that their classes are teacher-students centered. Only 22.2% (2 teachers) responded to have students-centered classes, while one teacher admitted to have a teacher-centered class. Conversely, talking to the teachers (and attending some classes) showed another picture. The whole efforts of teaching seem to rely on the teacher per se.

Table (4.12) Frequencies of the Teachers' Questionnaire: Item 11

Item 11: Is your class student-centered, teacher-centered, or teacher-student-centered?		
Responses	Frequency	Percent
Valid students-centered	2	22.2
teacher-centered	1	11.1
teacher-student – centered	6	66.7
Total	9	100.0

***Item 12*:** Item 12 responses reflect the teachers' perspectives of the majority of their students in conversation classes. As Table (4.13) shows, more than half of the teachers considered the majority of their students to be active learners, while two teachers chose to give equal judgement of their students and only one teacher admitted to have the majority of her/his students in conversation classes to be passive learners. Furthermore, one teacher missed to respond to this item that may reflect no real interest in giving a judgement of her/his students' performance publically.

Table (4.13) Frequencies of the Teachers' Questionnaire: Item 12

Item 12: How do you describe the majority of students in conversation classes?		
	Frequency	Percent
active L	5	55.6
passive L	1	11.1
neither active-nor-passive	2	22.2
9.00	1	11.1
Total	9	100.0

Away from the previous two items, i.e. items 11 and 12, when asking the teachers of conversation directly how well their students perform inside the classrooms, their answers never differ from one opinion, that is, most of the classroom activities are performed by some specific members in the class while the majority of the classes seem passive along the whole course.

***Item 13*:** Item 13 deals with the various teaching approaches and methods adopted by the teachers in their conversation classes. While the majority of the teachers follow the communicative approach in their classes, three teachers (33.3%) selected the eclectic approach, which reflects a more control over the teaching approaches and methods. However, one teacher chose all the approaches and methods listed as an approach he/she follows in the classroom. Table (4.14) given below shows the results of item 13.

Table (4. 14) Frequencies of the Teachers' Questionnaire: Item 13

Item 13: Which approach do you follow in the course?		
	Frequency	Percent
Valid ComApp	5	55.6
Eclectic App	3	33.3
All the above (all the approaches and methods listed in Item 13)	1	11.1
Total	9	100.0

***Item 14*:** To see how the teachers deal with the mistakes of their students, item 14 asks if they correct, do not correct, or over correct them. All the teachers reached a consensus of correcting the mistakes, with an additional comment from one of them stating, "*It depends*". However, correcting mistakes is one type of giving feedback to the students; no matter it does not go along with the principles of the communicative approach (which is adopted by the majority of the teachers). Table (4.15) shows the results of item 14.

Table (4.15) Frequencies of the Teachers' Questionnaire: Item 14

Item 14: What do you do concerning the students' mistakes?		
	Frequency	Percent
Valid Correct	9	100.0
Over Correct	0	0
Do not correct	0	0
Total	9	100.0

**Item 15****:** To see if the teachers have studied something about assessment methods and techniques, item 15 revealed that six teachers (66.7%) have studied assessment as one topic in the curriculum in their academic learning at college. In this regard, training teachers at assessments practices is a very significant issue to enable them give a valid judgement of the students' performance. Table (4.16) shows the results of item 15.

Table (4.16) Frequencies of the Teachers' Questionnaire: Item 15

Item 15: In your academic learning, have you studied assessment as one topic in the syllabus?		
	Frequency	Percent
Valid Yes	6	66.7
No	3	33.3
Total	9	100.0

**Items 16****:** The last two items in the teachers' questionnaire, i.e. 16(a) and 17(a), tackle the teachers' perspectives regarding teaching conversation, how interested they are in doing this task, and whether or not their students are interested in the classes. On the other hand, items 16(b) and 17(b) ask for their comments in case they responded with "No". The responses of item 16(a) are shown in Table (4.17).

Table (4.17) Frequencies of the Teachers' Questionnaire: Item 16(a)

Item 16(a): Are you interested in teaching conversation?		
	Frequency	Percent
Valid Yes	8	88.9
No	1	11.1
Total	9	100.0

When asking the teachers if they are interested in teaching conversation, eight teachers (88.9%) showed interest while only one teacher showed disinterest in teaching conversation; see Table (4.17). Although item 16(b) was specified to ask the disinterested teachers to justify their disinterest, some interested teachers found it a room to add their comments reflecting their needs

and different opinions concerning teaching conversation. For their importance, the following paragraphs list the teachers' comments (verbatim), starting with the comments of the disinterested teacher who was frank in reflecting the actual situation inside the classroom.

> "*I don't like pushing students to talk. Students are not interested in talking.*"

However, the comments added by the interested teachers were three. Whereas the first one showed an extra interest by adding the comment

> *"Very much",*

the other two teachers gave further details reflecting a message to deliver through this project (apparently to the authorities of the university). Their comments were:

> *"It is a very important course that unfortunately be underestimated for a long time!"*

> *"It is not just a matter of like or dislike. May be you like to do something, but there are no facilities to do it in a good way. For example, is it possible to apply different techniques in conversation without having for example a lab in the department?"*

Moreover, some other comments were told directly to the researcher focusing upon the facilities that a teacher of conversation usually needs inside the classroom. Some mentioned lack of the audio-visual aids, activities textbooks, flash cards, and the like. While others called for a special room supplied with all the facilities dedicated only for conversation classes.

Item 17: Whereas item 16 tackles the teachers' interest in the teaching of conversation, item 17(a) is set to examine the teachers' perspectives concerning the students' interest in the classes. As can be seen in Table (4.18), the responses

of the teachers were exactly the same as their responses to item 16, and the same disinterested teacher found her/his students to be disinterested too.

Table (4.18) Frequencies of the Teachers' Questionnaire: Item 17(a)

Item 17(a): Do you feel your students are interested in the conversation lessons? ☐ Yes ☐ No ☐ Not much		
	Frequency	Percent
Valid Yes	8	88.9
No	1	11.1
Total	9	100.0

For answering item 17(b), which asks for a justification for the disinterest, the disinterested teacher, commented as follows:

"They like to get rid of extra duties."

On the other hand, one of the interested teachers commented:

"Very interested"

4.4.   The Pilot Study

Since the use of rubrics is the first of its type during the period of the study, the researcher had to raise many queries on how to start using a rubric in formative assessment inside the classroom. Several professors, researchers, and EFL teachers were contacted for that purpose via emails and discussions in the academic platform "Academia.edu" and the world's largest professional network "LinkedIn.com". At this point, the most guiding suggestion was given by one of the advocates of rubrics, Dr. Heidi G. Andrade, the Assistant Professor at the University of Albany/ New York, who recommended conducting a pilot study for practicing the application of the rubric with students.

4.4.1. Participants of the Pilot Study

Prior to the pilot study, the researcher attended three lectures as an observer of students' performance in conversation classes with one teacher at the

Department of English, College of Education for Humanities. The participants were in the third stage which was divided into three groups (A, B, and C), each ranging between 24-26 students. Group B (consisting of twenty-five female students) was chosen arbitrarily, regardless of the students' gender or level of proficiency.

4.4.2. The Researcher as a Teacher

After being introduced by the teacher of conversation to group B, the researcher started instructing the group along the next ten lectures for the purpose of the pilot study. In her first lecture as their teacher, the researcher video-recorded the students while they were introducing themselves, giving some information about their families, and talking about their likes and dislikes. This was done for the sake of memorizing the students' names and to listen to each student's performance closely. The researcher explained to the students and promised them that the recordings will be kept only with her and never to be shared publically. This was important for the students to ensure some sort of comfort and trust with the researcher whom they deal with for the first time.

Finishing the recording, the researcher told the students some primary remarks about the method she wanted to involve them in, and why she would replace their teacher for the next ten lectures. At the end of the lecture, she asked the students to write her lists of the topics they like to discuss inside the classroom so as to make them suitable handouts.

4.4.3. Developing the Rubric

Playing the recordings of the students repeatedly helped the researcher to have an idea about their speaking performance. More significantly, they provided the baseline in developing the first rubric, which was used in the pilot study. For the sake of developing the rubric, the researcher surveyed the available speaking rubrics, those used in previous studies, found in books of

assessment, or published online. The primary components of speech available in the majority of those rubrics were fluency, grammar, pronunciation, vocabulary, and comprehension, but with different rankings.

Having the purpose of practicing the use of rubrics in classroom assessment in mind, the researcher developed a rubric containing only the previous five speaking criteria, hoping to establish the basis for the rubric that will be used later in the project study. The available rubrics were surveyed to have an idea of the speaking criteria involved, how to write the descriptors clearly and how to avoid vague words. Apparently, even those official scales which were used for national certificates like the Finnish National Certificate, the American Council for the Teaching of Foreign Languages (ACTFL), the Test of Spoken English (ETS), and the Common European Framework (CEF) included quantifiers like "few, little, some, somewhat, wide range" and qualifiers like "adequately, well, generally appropriate, sufficient, good" (cited in Luoma, 2004: 61- 74), which are vague and might gain a variety of interpretations among raters. After having an understanding about those rubrics, the researcher developed the first rubric (Appendix III) which was refined later in a valuable discussion with her supervisor. Yet, admitting the fact that the best rubric is the one developed in collaboration with students, the researcher could not achieve this point since the whole issue of formative assessment- let alone the term 'rubric'- was new to the participants, and they needed time to understand their meanings and how all related things work.

In the second lecture, the researcher introduced the rubric to the students and explained its three dimensions. She used to give examples for each criterion and levels with different performances taken from their recordings. When she was certain enough they had understood the rubric, she told them that it would be their tool of assessment along her presence and she would be using it in a daily assessment of their performance.

Concerning the scoring of the rubric, all the criteria were given two scores to result in a total score of 20 for the whole rubric.

4.4.4. Teaching Technique and Materials

Noting that the students were accustomed with the interview technique, the researcher adopted that technique as a first step in her instruction. Moreover, since there was no particular curriculum for conversation classes, the researcher set some handouts (some were topics suggested by the students in their lists) to establish materials for the classroom discussions. The total number of handouts used along the pilot study was six (attached in Appendix IV). At the end of each lecture, the researcher used to give all the students copies of the handout to be covered in the next lecture. Also, she prepared several games to start the lectures with, which was a motivating factor inside the classroom. The students-including the passive ones- liked the games a lot, and were so amused while practicing it. However, once a game is ended, the students get back to their own status.

4.4.5. Difficulty in Applying the Rubric

A big problem appeared at the first day for the researcher when she started her formative assessment with the use of the first designed rubric. Being used to questions and answers format, students did not participate in class discussions without being asked by name. Therefore, it was her responsibility to raise questions, engage students in discussions, make comments, and keep conversation going along the lecture time. In other words, it was a plain teacher-centered classroom. Moreover, she had to tick the rubric of each student while they were speaking which was a confusing issue at the first time. Doing all that together seemed challenging from the first lecture, so the researcher did not use the rubric from the first day, aiming at establishing a good relation with the students and getting some informal practice with the rubric.

Facing that difficulty, the researcher found that the descriptors were not clear enough to be followed along the classroom discussions. For that reason, she made certain modifications according to the students' weaknesses that became apparent to her from the first lecture and through repeating watching its videos. The most effective modification that was added to the rubric was the no use of the native language (i.e. Arabic) in the classroom discussions as one criterion and giving it a higher mark than the other criteria. This was very important as some students used to frequently shift to the use of Arabic repeatedly in the discussions. Since then, all the students stopped using their native language along classroom discussions, which was encouraging for the researcher at that time. Even the most common word "yaani" was replaced by its English translation "I mean". In addition, the researcher included the scoring marks in the rubric sheet, so that students would be aware of the scorings and accountability of the rubric. This particular point made students more thoughtful of the use of the rubric and started to take it more seriously. Earlier to that, they did not give much importance to the feedback they receive when the researcher started its application in the classroom. Nevertheless, scoring their rubrics with a final mark and handing them back to the students raised actual attention to the researcher's feedback, which she used to give to the students at the beginning of each next lecture.

Simultaneously, the researcher started several discussions online with the researchers and EFL teachers worldwide about her procedure and the specific difficulty she faced in the first lecture. For solving that, Dr. Heidi G. Andrade has recommended the use of peer assessment and not to depend solely on the teacher in classroom assessment. That was the same opinion of some EFL and ESL teachers experienced in the use of rubrics and peer assessment.

4.4.6. Introducing Peer Assessment to the Students

After reviewing several studies and articles in the field of self and peer assessment and discussing the matter with her supervisor, the researcher determined to introduce the practice of peer assessment to the students. In her third lecture- the second in using the scoring rubric- the researcher explained the peer assessment to the students and put them in pairs according to their closeness in seats. And she asked them to keep the peering along the study.

4.4.7. Students' First Reaction to Peer Assessments

Since being presented with no marks on its sheet, the scoring rubric did not gain the students' interest in the first two lectures. They did not seem to take it seriously, which made it a new challenge for the researcher on how to gain their interest. So, she presented the scoring marks for each criterion in a small table just below the rubric scheme (Appendix V). In the first lecture of using peer assessment with the marked rubric, the researcher gave a Zero score to each unassessed criterion by a peer. Receiving their rubrics, the students started to complain from getting zero though they have participated in classroom discussions. The researcher answered that it was their peers' assessment not hers. Since then, the assessees asked their assessors to assess well, and those who did not use to participate in classroom discussions started to raise their hands to take part in discussions once they got back their rubrics empty of ticking and getting Zero as a final assessment. This was done deliberately by the researcher to evoke passive students, who began to be aware of the importance of participation (essentially to be peer-assessed), to start participating in classroom discussions.

When asking students what they thought about the use of peer assessment, many of them said it helped them recognize their strengths and weaknesses in speaking and enabled them to try to overcome their weaknesses. However,

when the researcher asked the students if they found the rubric as a fair tool of assessment, one student, who had difficulties in pronunciation and grammar, made a passive remark in that it shows her weaknesses which she does not like to be noticeable by others.

4.4.8. The Problem of Time and Large Class Members

When students started to take the rubric and peer assessment seriously, all of them showed interest in participating in classroom discussions. However, the new challenge was the large number of students (twenty-five) and the limited time allotted to classes (fifty minutes). Students who used to talk at the beginning of the class discussion used to take a longer time than those talking at the end of the class time. The researcher and the assessee peers found it unfair to assess those students who could hardly say two or three statements at the end of the class time in comparison with those discussing and responding to the researcher's comments and questions at the beginning of the lecture.

Though this method seemed to motivate passive students to take part in each class, it started to passivize those active ones in a certain way. Because of time limitation, active students could not have the chance to participate more than once in classroom discussions because in doing so, they will take much time in interacting with the researcher and this will decrease the opportunity for other students to participate (and be assessed by their peers). This passive aspect urged the researcher to repeatedly search for and contact different researchers online to find out a suitable solution. As a result, the pilot study ended since the researcher got a fair practice with the use of the rubric and gained a good understanding of the different aspects that were obscure previously. Meanwhile, the researcher had applied the use of peer assessment with the rubric along four lectures in which all the present students used to take part though not in fair timing.

4.5.   The Methodology of the Empirical Study

The following sections detail the procedure followed in the empirical part of the study.

4.5.1. Solving the Problem of Time and Large Class Members

The main problem that was facing the researcher at this point was how to help all the students practice and improve their speaking skill while they have a very limited time to take part in classroom discussions. Moreover, watching the videos of the pilot study, the researcher felt unsatisfied to see more than twenty students waiting silently while she was busy trying to make one student to talk. Hence, the interview method in conversation classes, with such a large number of students, proved to be a passive method of interaction inside the conversation classroom.

At that time, the need for creating a student-centered classroom emerged seriously to offer all students fair opportunities to take part in classroom discussions and create a more interactive session in each conversation class. For that purpose, new online discussions started by the researcher with some researchers referred to in this study, particularly Professor Keith Topping[2] who suggested implementing subgroup discussions inside the classroom for the sake of creating a student-centered class that engages all the students in interactive sessions in the limited period of class. The procedure to be followed and the group discussion rubric to be used were discussed in detail and many precious comments were added by him.

4.5.2. The Participants of the Empirical Study

All the students of the third stage participated in the study that has extended along two months. The whole number of the students was seventy-four, sixty-

---

[2] The Director of the Centre for Paired Learning at the University of Dundee in Scotland, U.K.

eight females and 6 males. Since the number of males was very few in comparison to the number of females, the gender factor has not been taken into consideration 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 for Humanities. For example, Table (4.19) and Figure (4.1) show the distribution of males and females for the four stages in the above-mentioned department at the academic year 2011-2012. Thus, though they may not be an adequate representation concerning the gender factor, the participants are the convenience sample of the targeted population.

Table (4.19) Distribution of the students according to gender in the Department of English, College of Education for Humanities, during the academic year 2011-2012

Stage	Females	Males	Total
First stage	88	30	118
Second stage	77	13	90
Third stage	68	6	74
Fourth stage	61	13	74
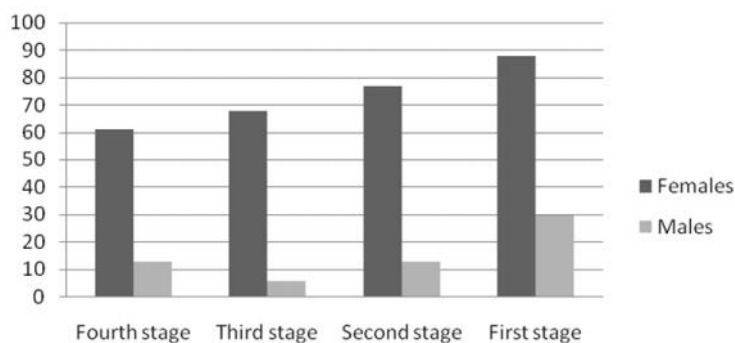
Figure (4.1) Distribution of the students according to gender in the Department of English, College of Education for Humanities, during the academic year 2011-2012

4.5.3. Period of the Empirical Study

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 to each group of the third stage. So, the total number of the lectures of the study was 12 for each group and the researcher had to attend six lectures weekly to cover the whole stage.

4.5.4. The Researcher as a Teacher

Following the procedure given below, the researcher taught the three groups without the presence of the teachers of conversation. (Two teachers were responsible for teaching conversation classes to the third stage).

4.5.5. Procedure

Since the goal of the teaching process was changed from merely concentrating on the linguistic criteria of the speaking skill to enhancing the motivation inside classroom and creating a student-centered class, the study methodology was formed in a way that is suitable for the new goals. Primarily, each group in the third 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; CA, CB, CC, CD). A group technique replaced the interview technique to achieve the goals of the study. The following sections illustrate the procedure applied in detail.

4.5.5.1. The First Lecture

The first lecture was an introductory one, especially with the new groups: A and C. The researcher asked the students their permission to video-record them while introducing themselves and talking about their likes and dislikes. Some of the students queried about the purpose of the video recording, so the researcher comforted them that it is for the study purposes, it is confidential and will only be shown to some faculty members inside the department (who were already

their teachers). She, also, clarified the special necessity of the first recording, which was primarily to help her memorize their names quickly, since she is going to lecture them in the next session.

Interviewing the students while talking about their personal backgrounds was very helpful, especially when they talked about their likes and dislikes that offered a good opportunity to the researcher to know them closely and helped in creating an interesting atmosphere inside the classroom. At the end of the recording process, the researcher told the students about the group technique that she would follow in the study. She asked them to arrange their groups of six as they wish and prepare topics to talk about in the following lecture. She also told them that only the following and the last lectures (of the study period) would be recorded and would be considered as the pre- and post-tests of the study, while the other middle lectures are not to be recorded. Finishing the first lectures with groups A and C, the researcher asked the students to write the topics they like to talk about inside the classroom. (Group B had already done that in the pilot study.)

Leaving the class, several students asked some questions about the study, while others raised more practical queries on how to improve their speaking performance. One of the students remarked:

> *"We know that language skills are four, but we practice only two of them; reading and writing".*

Hearing from the participants of the pilot study, the students showed their interest to take part in the study, hoping primarily to improve their speaking skill. It was clear that they were seeking more chances of practice inside the classroom.

4.5.5.2. The Pre-test

In the second lecture, the students sat in groups of six, four and/or five (where there are absentees). Some of them were ready for the pre-test while some others claimed that they forgot to prepare a topic. To give them a fair chance, the researcher told them that the lecture would be a try for the group discussion and the pre-test would take place in the following lecture. This was also helpful to give them a practice on talking in front of the camera. As a result, all the three groups were given two chances to prepare for a topic and were told that the second recording would be regarded as their pre-tests.

In the days of the pre-tests, the students were asked to start their discussions with the topics they had prepared. The researcher gave each group ten minutes of discussion so that she could cover all the class in one lecture. Prior to the discussions, each subgroup was asked to nominate a leader to be responsible for managing the subgroup discussion, questioning, commenting, etc. One group after the other was video-recorded by the researcher while the other groups were either listening or planning for their own discussions. The latter were asked to keep quiet so that the recording could be clear enough.

Having 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. Moreover, the repetition of the recordings helped to establish a basis for the students' performance and develop the suitable rubric.

4.5.5.3. The First Use of the Pre-test Videos

Giving the students authentic products of their own helped in enhancing their awareness to their strengths and weaknesses. For this reason, the researcher made use of the students' pre-test videos and presented them to the students in the third lecture as examples of different performances. 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.

Watching their own videos, each sub-group was asked to evaluate its own performance. The researcher recommended the students not to make fun of each other's speech and to give constructive criticism instead. Some students who were confident in speech used to criticize themselves while others used to watch eagerly how they performed in the pre-test. On the other hand, the other sub-groups were asked to criticize their mates in the videos and the whole class shared their comments, which were intended for the improvement of their own and mates' performance. It is worth mentioning that the sub-groups videos were showed only to their main groups and not shared with the other groups in the stage.

4.5.5.4. 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 chose their sub-groupings in their own way. So, the researcher asked them to keep those sub-groupings along the period of her teaching. 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 researcher introduces to the students some conversational strategies, which aimed at helping them manage the subgroup discussions and improve their fluency (following Brown, 2004: 270). In each lecture, a list of strategies is written on the board to the students 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 does not participate in these discussions and her role is 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 finishing 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 used to peer assess each other within their sub-groups and collect their rubrics with the leaders. The peer assessment takes about 5 minutes to finish.

With 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.

4.5.5.5. 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, which aimed at engaging the students in interactive discussions. In all 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 and others were set by the researcher. The tasks are listed below according to their sequence in the study. This sequence was followed due to the availability of the tasks to the researcher.

1. *Ranking*: Adopted from Gammidge (2004:69) with certain modifications. This task concentrates upon ranking and friendship discussion. Prior to the task, the students were asked what the qualities of a good friend are. From their answers, several characteristics were written by the researcher on the board. To make the list short, only eight characteristics were left on the board and the students were asked to do the task after explaining the meaning of each one. The characteristics were honesty, loyalty, confidentiality, amusement, temperament, generousity, intelligence, and like-mindedness.

   The students were asked to discuss that list according to their own perspectives covering questions like what kind of characteristics they liked to have in their friends, which were important and which were not, etc. At the end of the discussions, they had to rank the characteristics in a new list according to their importance, from the most to the least important. An open discussion followed and the students were interested to participate once they had a goal in the discussion unlike if it was a common discussion about friendship, which they described earlier as being boring.

2. *Problem Solving*: Adopted from Ur (1996: 127). This task concentrates on problem solving and how students can work together in a cooperative team to reach a consensus. In this task, the students were told that they

were going to play the role of an educational advisory committee who were responsible for solving the problems of students. Their immediate task was to advise a school principal to solve the problem of one of his students. The problem was written in a sheet of paper and given to the discussion leader. Prior to that, the researcher gave the instruction and made sure that the students understood their mission in the task. The problem was the following.

Banny, the only child of rich parents, is in the 7th Grade (aged 13). He is unpopular with both children and teachers. He likes to attach himself to other members of the class, looking for attention, and doesn't seem to realize they don't want him. He likes to express his opinions, in class and out of it, but his ideas are often silly, and laughed at. He has bad breath.

Last Thursday, his classmates got annoyed and told him straight that they didn't want him around; next lesson a teacher scolded him sharply in front of the class. Later, he was found crying in the toilet saying he wanted to die. He was taken home and has not been back to school since.

At the end of the task, the leaders presented the solutions that were discussed openly with the researcher and other groups' members. Then, ideas were exchanged among the groups in defining the best solution.

3. *Shopping list***:** Adopted from Ur (1996: 126-127). This task concentrates upon convincing. The students had to imagine that they were going to a miracle store which sells the commodities shown in Table (4.20) given below.

Table (4.20) Shopping List (a classroom activity)

No	Name of commodity	1st buyer	2nd buyer	3rd buyer	4th buyer
1.	More free time				
2.	An automatic house-cleaning robot				
3.	Popularity				
4.	A job that involves travel abroad				
5.	Fame				
6.	More patience				
7.	A perfect figure				
8.	More excitement in my life				
9.	Perfect health				
10.	A talent for making money				

The leaders of the sub-groups were the owners of the stores. They had to find each commodity four buyers (if the sub-group consisted of six students) and two buyers (if the sub-group consisted of four students) to stock the items. Sub-groups members should not buy things that they do not need, and the storekeepers should try their best to convince them of buying their items. The members were also encouraged to assist the storekeepers in persuading each other to buy their goods. The storekeepers had to write the names of the buyers next to each item in the table.

At the end of the time allotted to the discussion, there was a comparison between the lists of the storekeepers to see who sold all their items. The leaders talked about how difficult their job was in convincing others of buying things and members explained why- and why not- they bought certain items.

In this lecture, the class atmosphere was very active and storekeepers were very excited to sell their goods. The class contained lots of shouts and it gave the impression of having the lecture at the market!

4. ***Technology***: This task was designed by the researcher. The students were told that they were going to get back to the date before 2003 (the date of ending the sanction). They had to imagine their lives without the technology they got after that date. They were asked to decide what were the most important technologies they got after 2003 and wanted to keep if they were to get nine years backward. The condition was that each subgroup had the right to keep only three devices and should nominate the names of the owners of the devices who should be ready to share their possessions. For instance, if they choose to keep a cell phone, they should find a cooperative friend who is willing to lend them her/his device whenever they need it.

In this task, the students shared their opinions upon the available devices which were not found before 2003 and which were the most beneficial ones. The leaders had to ask the members their opinions and decide with them what devices to keep. Since the number was limited, the students had the task of agreeing and disagreeing within the sub-group in their decisions. Moreover, they were asked to practice agreeing and disagreeing statements that were introduced at that lecture to reach their conclusions.

5. ***Balloon Debate***: Adopted from Harmer (2001: 273)*.* In this task, the students play the role of people of occupations travelling in the basket of a balloon. Unfortunately, the balloon had a leak, so it could not bear their weight anymore. Therefore, unless some passengers leave the basket, they will all die. The students should have arguments upon whom should be the survivors and why. Students should defend their positions and convince the others of the importance of their occupations. After having the first argument, the students should vote on whom would jump first from the balloon. As more air was escaping, the students should decide

on the other, and so on until leaving only one member in the balloon (in the sub-group of four students) or two (in the sub-group of six students).

Prior to this task- in the lecture preceding it- the students were asked to choose themselves occupations, on the condition that not to have the same occupation within the same sub-group. They were asked to choose the occupations, which they think, are important to the society and to know how to justify their importance if asked. They were not informed of the task itself, so they thought it was a common discussion of occupations. Once started the task, the students were very motivated and acted in much interest- for the sake of keeping their lives. The challenge was how to convince others of the significance of one's occupation and keep one's life. At the end of the discussions, the leaders told who were their survivors and the whole group- with the researcher- had an argument upon each sub-group's decision.

6. *Journey Plan:* Based on discussion in some online forums, this task was created by the researcher to practice planning and sequencing. The students were told that they were going to have a six-day vacation. Along that period, they would be travelling around Iraq. They had the task of preparing a plan concerning the locations they like to visit, how long they will stay at each place, how to spend their time there and what kind of transportation to take. They had to write that plan on a sheet of paper to be shared later with the other sub-groups.

The students started their discussions with some interest. Having exchanged some opinions, they started to show more interest and acted in a more interactive way. Some of them suggested nice ideas like staying at their relatives' for saving money and going to more places. At the end, their plans showed many similarities, but the sequencing and timing differed significantly. They talked very eagerly about their journeys and

they ended the lecture by saying that they imagined each step in the journey and acted as if it were real.

4.5.5.6. 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. Meanwhile, the basic goal was to help initiate an interactive student-centered classroom, that is why the new rubric included criteria like:

1. *Interaction*: to direct the students within their sub-group discussions and encourage them to participate in various ways.
2. *Contents*: to encourage the students giving valuable information in regards to the subject matter, and not to take a passive role.
3. *Humor*: to make the discussions as formal as possible and not to deviate away from the classroom atmosphere. Since the students have the control over their give-and-take, this criterion was important to keep the discussions within the academic principles.
4. ***Assessment Criteria of the Discussion Leader (managing discussion within the time limit, and engaging group members):*** to guide and direct the leader through the discussions and to keep the allotted time prescribed earlier.

In addition to that, two sub-criteria of 'pronunciation' were excluded from the original rubric because they proved unclear to understand by the students, and to achieve and assess in turn. They were 'the use of intonation and rhythm' and 'loudness of voice'. This modification was made because more important criteria emerged during the pilot study as introduced before. See (Appendix VI) for the final rubric.

As far as the scoring of the rubric is concerned, each criterion was given two scores except the no use of Arabic and interaction, which were given four

scores. 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. Concerning the discussion leaders, the criteria that are found in their table gained two scores for each and substituted the interaction criterion found in the main table. The total score of the rubric for any student is 20 marks. Table (4.21) below illustrates the scoring of the rubric.

Table (4.21) Scoring of the 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*	Managing the discussion within the time limit	2	1.5	1	0.5
	Engaging group members in the discussion	2	1.5	1	0.5
Total score of the rubric		20	* the criteria of the discussion leader are to replace interaction in the leaders' assessment		

4.5.5.7. The Peer Assessment

When asking the students participating in the pilot study about the peering, they preferred if they can have new peers in each lecture. Hence, since the peering has not proved by earlier studies to be necessarily static, the researcher gave the freedom to the students to choose their peers along the study. Some of them kept their peer 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 student's performance, thus such an assessment was a bare necessity in

assisting the teacher to control and guide the large class. Additionally, to guarantee the objectivity of the rubric, the researcher explained to the students that the words 'some' and 'hardly' would 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. In the case of odd sub-groups, the leaders were excluded from the peer assessment, but were still assessed by all the members in the table specified for them.

4.5.5.8. 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 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 would take the same procedure of the pre-test, i.e. a ten-minute discussion within the sub-groups, one after another, and their assessment would 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 exam, the researcher asked for 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 could assess the students performance for the second term results while she assessesed them for the study

purposes. Getting the approval, she prepared the topics of the discussion (Appendix VII) 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.

At 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 out of 20 which was the total mark of the second term exam (and which was the same total score of the rubric).

4.6.   The Students' Questionnaire

Based on the questionnaire of White (2009), a students' questionnaire (Appendix VIII) was set to find out the students' perspectives of the practices of the rubric, the peer assessment, and the sub-group technique. Since the rubric used in White's study had no descriptors, the items of his questionnaire did not go along with the current study, so the idea was adopted from his work with certain modifications to meet the objective of the questionnaire of this study. Consequently, its items were set by the researcher involving the advantages and disadvantages of the three practices as far as students are concerned with the intention of checking their validity.

The questionnaire included 20 items with three options for each; agree, neutral, and disagree. Finishing the post-test, the students were asked to fill the questionnaire and thanked for their cooperation. They were told that their opinions were very important because they were the first in the department to experience such a study and their frankness is much valuable to the research. To feel free in their answers and comments, the students were directed by the

researcher to answer the questionnaire sheets and hand them to the secretariat of the Department of English, who delivered them in turn to the researcher.

4.7.   The Interrater Reliability Test

The following sections tackle the interrater reliability test which was conducted for the sake of checking the reliability of the rubric designed by the researcher for the study.

4.7.1. The Raters

To check the interrater reliability of the rubric, the researcher asked the assistance of 10 teachers at the Department of English, College of Education for Humanities to do the interrater reliability test. All the teachers have an experience in teaching conversation classes and five of them were teaching conversation at the same year of the study. Appendix (IX) shows their academic information.

4.7.2. Context of the Test

The test took place in the laboratory of the Department of English, College of Education for Humanities. Using a smart board, one video from the pre-test was selected randomly to be shown to the teachers. Since the application of the rubric by the teachers was the objective of the test, then no importance was given to the students' levels in the video. The recording was for a group consisting of four females and two males, one male was the leader of the discussion.

The test lasted for about an hour in which the researcher explained all the criteria levels in the rubric with their descriptors. After agreeing upon the meanings of the words used in the descriptors- as happened with the students in the peer assessment- the video was played and the teachers were asked to assess the students' performance in the video regardless of their prior knowledge of the students' capabilities. In addition to that, because it was the teachers' first use of

the rubric, the researcher had to stop the video once each student finishes her/his role in the discussion so as to make their ticks on the rubric. The teachers were provided with pencils to make the assessments, so that they could erase and re-tick the rubric levels when necessary.

4.7.3. Application of the Test

Starting the video, the teachers used to assess the students concentrating upon the linguistic criteria, while leaving the interactional criteria to the end of the discussion, which lasted 13.15 seconds (i.e. they exceeded the time-of 10 minutes- allotted for the discussion). Assessing the leader's interaction criteria was also postponed to the end of the video so that the teachers could see how he managed leading his group.

The teachers were very cooperative and they supplied the researcher with various comments, which were much helpful to the study. On the top of that comments, Mr. Saad Chasib Daghir suggested making a comparison between the traditional assessment and the rubric assessment in a way to find out which method was more reliable. As a result to the discussion with the raters concerning this point, the researcher was convinced to add this point to the methodology. In consequence, the teachers were asked to grade the students following the traditional assessment after finishing the rubric assessment. The teachers had their own criteria according to which they evaluated the speaking skill of the students. To identify those criteria, the teachers were asked to list them on the back page of the rubric. They are listed in Table (4.22) given below.

Table (4.22) Criteria of the Raters' Traditional Assessment

Rater	Criteria				
1	Fluency	Grammar			
2	Fluency	Communication			
3	Grammar	Pronunciation	Interaction	Fluency	Content Meaning (Semantics)
4	The Same Criteria Of The Rubric				
5	Fluency	Grammar	General Impression		
6	Fluency	Grammar	Speed Without Pauses		
7	Fluency	Grammar	Interaction		
8	Speaking	Listening	Fluency	Structure	General Impression
9	Participation	Fluency	Ability To Persuade Others	Grammar And Pronunciation	Self-Confidence
10	Fluency	Grammar	Interaction		

With the end of the interrater reliability test, the researcher gained two types of assessments, the alternative assessment with the hypothesized objective rubric and the traditional one with the subjective ratings.

CHAPTER FIVE: RESULTS AND DISCUSSION

5.1. Introduction

The data gathered out of the methodologies presented in Chapter Four were collected and computed for statistical analyses using the Social Sciences Statistical Package (SPSS v.18). Different analyses were applied for this purpose. The descriptive frequencies were used to analyze the students' responses regarding the use of rubric and peer assessment practices and the group discussion technique. A Cronpach's alpa was conducted to measure the interrater reliability of the rubric when used by the teachers, and the Coefficient of Variation was applied to make a comparison between the two types of assessment, namely, the rubric and the traditional assessments. Finally, a paired t-test was used to measure the differences in the performance of the students before and after applying the study (the Pre- Post tests). The following sections tackle the results of each analysis with the related discussion.

5.2. Students' Questionnaire

To test the results of the students' questionnaire, descriptive frequencies and a percentage analysis were applied. To guarantee a reliable analysis, the questionnaire was divided into three categories, each category focuses upon a particular dimension of the questionnaire: the construction and use of the rubric (items 1, 2, 3, 4, 5, and 18), practicing peer assessment (items 6, 7, 8, 9, 10, 11, 12, 13, 14, and 19), and the application of the group discussion technique (items 15, 16, 17, and 20). Items 18, 19, and 20 were intended at the end of the questionnaire to give a final judgement of experiencing the three practices of alternative assessment implemented in the study. The following sections show the results of each category.

5.2.1. Students' Perspectives of the Construction and Use of the Rubric

**Item 1**: From Table (5.1) given below, it can be seen that the descriptors of the rubric were clear enough to be understood and used in the classroom. The overall understanding of the descriptors is very important in judging the level of the peers and having an idea of the teachers' expectations of the different levels of a given criterion. The table shows that 48 students agree upon understanding the descriptors of the rubric, showing a high percentage of agreement (72.7%) to this point. Only one student (1.5%) showed a disagreement concerning this point, while 16 students did not find the descriptors of the rubric as clear enough, gaining a percentage of (24.2%) of the total number of the students. However, one student did not answer this item. Figure (5.1) illustrates the results of Item 1.

Table (5.1) Frequencies of the Students' Questionnaire: Item 1

Item 1: Criteria descriptors on the rubric were easy to understand			
		Frequency	Percent
Valid	Agree	48	72.7
	Neutral	16	24.2
	Disagree	1	1.5
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
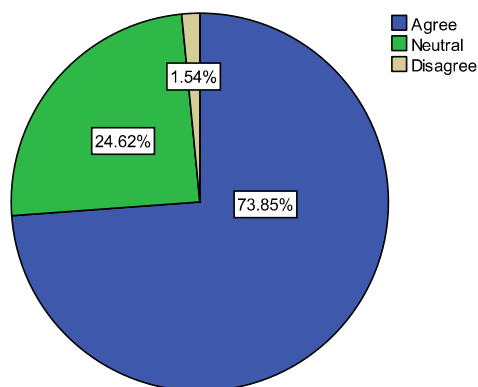
Figure (5.1) Frequencies of the Students' Questionnaire: Item 1

Item 2*:* It was not a surprise finding two-thirds of the students with no definite decisions concerning the levels of all the criteria of their peers (Table 5.2). This can be justified due to the novelty practice of the peer assessment with the use of the detailed rubric with no previous and/or long training (taking into consideration the group who took part in the pilot study). Figure (5.2) illustrates the results of Item 2.

Table (5.2) Frequencies of the Students' Questionnaire: Item 2

Item 2**: It was difficult to decide the level of my peer for all the criteria**			
		Frequency	Percent
Valid	Agree	12	18.2
	Neutral	44	66.7
	Disagree	10	15.2
	Total	66	100.0
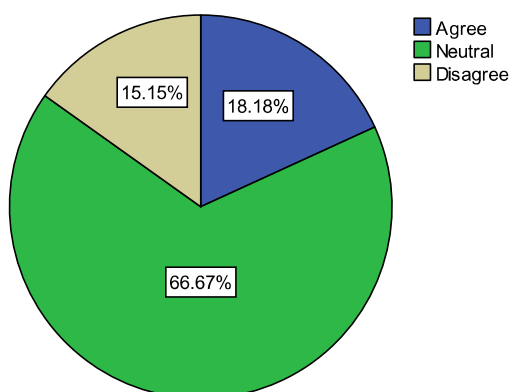
Figure (5.2) Frequencies of the Students' Questionnaire: Item 2

Item 3*:* One of the advantages of using the rubric in the classroom is to raise the level of the students' awareness of their weaknesses and strengths in the skill being assessed. As Table (5.3) shows, there is a high percentage of agreement (83.3%) among the students which gives evidence that using the rubric inside the classroom helped them be aware of their strengths and weaknesses. Figure (5.3) illustrates the results of Item 3.

Table (5.3) Frequencies of the Students' Questionnaire: Item 3

Item 3: I knew my strengths and weaknesses in speaking through using the rubric		
	Frequency	Percent
Valid Agree	55	83.3
Neutral	6	9.1
Disagree	5	7.6
Total	66	100.0
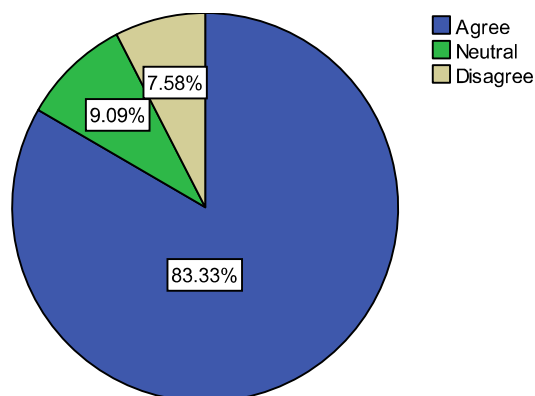
Figure (5.3) Frequencies of the Students' Questionnaire: Item 3

Item 4: The students' activity was clear enough to be realized along the period of the study. As shown in Table (5.4), 69.7% of the students (46 students) showed agreement to become more active after the use of the rubric, while only 1.5% (one student) disagreed on this point. However, some of the students were not strict to decide since some of the students were already active in the classroom discussions. This confirms Black and Wiliams' conclusion presented in Chapter Two that formative assessment is powerful in enhancing the students' achievement particularly "low achievers" (Black and Wiliams, 1998: 4). Figure (5.4) illustrates the results of Item 4.

Table (5.4) Frequencies of the Students' Questionnaire: Item 4

Item 4: I became more active in conversation classes after using the rubric in our daily assessment		Frequency	Percent
Valid	Agree	46	69.7
	Neutral	18	27.3
	Disagree	1	1.5
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0

Figure (5.4) Frequencies of the Students' Questionnaire: Item 4

___Item 5___*:* Another advantage of the rubric is tested in Item 5. The analytic rubric with the descriptors of the levels enabled the students to know what is expected of them to get a certain level for a given criterion. Thus, as shown in Table (5.5), 78.8% of the students (52 students) agreed that the rubric helped them know the characteristics of an active- as well as a passive- speaker. Figure (5.5) illustrates the results of Item 5.

Table (5.5) Frequencies of the Students' Questionnaire: Item 5

Item 5: The rubric told me what's expected of me to achieve the best performance in the speaking skill		
	Frequency	Percent
Valid Agree	52	78.8
Neutral	14	21.2
Total	66	100.0

Figure (5.5) Frequencies of the Students' Questionnaire: Item 5

Item 18*:* Being the only sample participating in the study, it was important to have an idea concerning the students' perspectives of using the rubric in future classes. As shown in Table (5.6), more than half the students (35 students 53%) recommended the use of rubrics, while only eight of them disagreed (12.3%). The neutral choices reflected the third percentage that showed no confirmed view towards the use of the instructional rubric. This can be supported by some of the comments added by the students when responding to the questionnaire (to be listed after the discussion of Item 20).

Table (5.6) Frequencies of the Students' Questionnaire: Item 18

Item 18: I recommend using the instructional rubric in conversation classes		Frequency	Percent
Valid	Agree	35	53.0
	Neutral	22	33.3
	Disagree	8	12.1
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
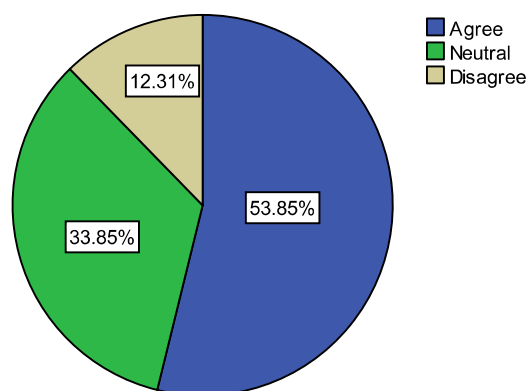
Figure (5.6) Frequencies of the Students' Questionnaire: Item 18

5.2.2. Students' Perspectives of Practicing Peer Assessment

Item 6: As shown in Table (5.7), a high percentage of 78.8% of the students have agreed upon the advantage of peer assessment in giving and receiving feedback. However, a percentage of 21.2% of them were neutral in giving their views. This may be attributed to some uninterested students in the use of peer assessment. Figure (5.7) illustrates the results of Item 6.

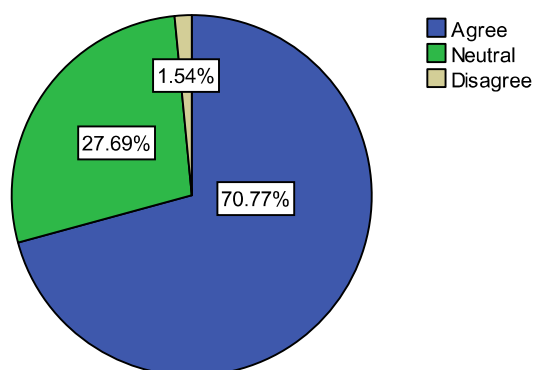**Table (5.7)** Frequencies of the Students' Questionnaire: Item 6

Item 6: Peer assessment allowed me to gain experience in giving and receiving feedback		
	Frequency	Percent
Valid Agree	52	78.8
Neutral	14	21.2
Total	66	100.0
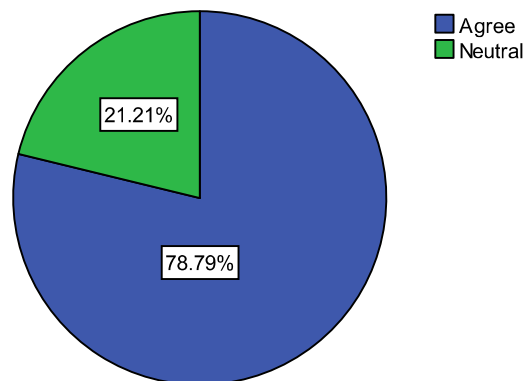
Figure (5.7) Frequencies of the Students' Questionnaire: Item 6

**Item 7**: One of the main objectives of peer assessment is giving a timely feedback. Regardless of the validity of the peers' feedbacks, their significance stems from motivating the students and raising their awareness to participate in the classroom activities. As shown in Table (5.8), two-thirds of the total number of the students (41 students, 62.1%) supported this objective, while the left third held a neutral view. The reason behind this may be attributed to the idea that some students were uncomfortable to have certain peerings, especially at the days of their friends being absent. Besides, some active students, when assessed by passive peers used to complain from not receiving a valid evaluation. In addition to that, only five students showed no comfort in acting as judges of their peers. Figure (5.8) illustrates the results of item 7.

Table (5.8) Frequencies of the Students' Questionnaire: Item 7

Item 7: Peer assessment supports me with detailed and timely feedback to my speaking skill			
		Frequency	Percent
Valid	Agree	41	62.1
	Neutral	23	34.8
	Disagree	1	1.5
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
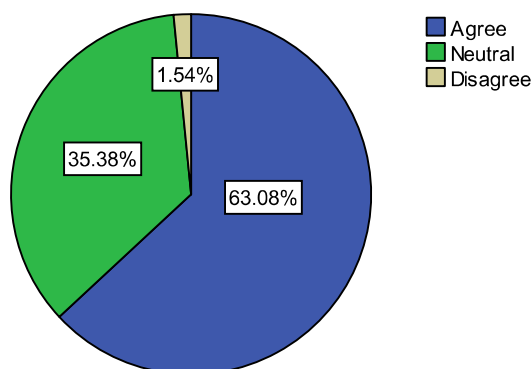
Figure (5.8) Frequencies of the Students' Questionnaire: Item 7

*__Item 8__: Due to the participants' personal relations with each other, the validity of peer assessment may be negatively affected. However, as shown in Table (5.9), about half the students (32 out of 66) disagreed on this point, showing a confidence concerning their own judgements of the performance of their peers. On the other hand, 14 of 66 students (21.2%) agreed on such an impact. The neutral third (20 out of 66, 30.3%) had no exact opinion, reflecting a probability of the validity of this item. Figure (5.9) illustrates the results of Item 8.

Table (5.9) Frequencies of the Students' Questionnaire: Item 8

Item 8: Relationships with my peer (friendships, etc.) may have influenced the overall assessment I give to him/her		
	Frequency	Percent
Valid Agree	14	21.2
Neutral	20	30.3
Disagree	32	48.5
Total	66	100.0
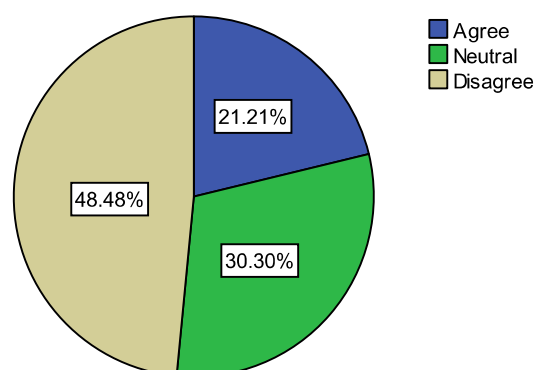
Figure (5.9) Frequencies of the Students' Questionnaire: Item 8

Item 9: Experiencing a new practice, 68.2% of the students (45 students) showed a comfortable status in taking part in their peers' assessment, as shown in Table (5.10). However, 24.2% of the students (16 out of 66) took a neutral stand to this point. This may be attributed to the unstable and/or unequal peering of the students. Figure (5.10) given below illustrates the results of item 9.

Table (5.10) Frequencies of the Students' Questionnaire: Item 9

Item 9: I was comfortable being a judge and assessing my peer's speaking skill		
	Frequency	Percent
Valid Agree	45	68.2
Neutral	16	24.2
Disagree	5	7.6
Total	66	100.0
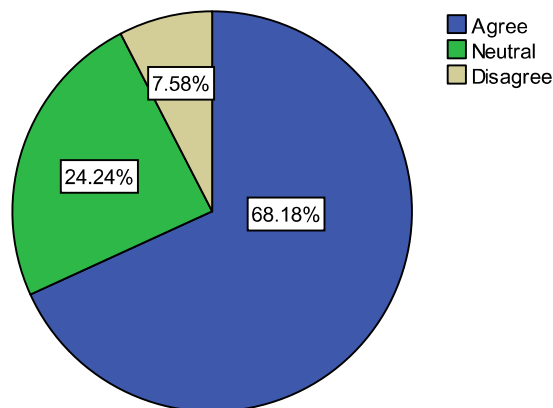
Figure (5.10) Frequencies of the Students' Questionnaire: Item 9

Item 10*:* Not to be mixed with the previous point, the students' views about having their speaking skill being assessed by their peers showed a significant percentage of 54.5% (36 students) of agreement as shown in Table (5.11). Still, one-third (34.8%) of the total number of the students held a neutral stand concerning this point, which may be attributed to the same reason for being uninterested in practicing peer assessment.

Table (5.11) Frequencies of the Students' Questionnaire: Item 10

Item 10: I was comfortable having my speaking skill judged and assessed by my peers			
		Frequency	Percent
Valid	Agree	36	54.5
	Neutral	23	34.8
	Disagree	7	10.6
	Total	66	100.0
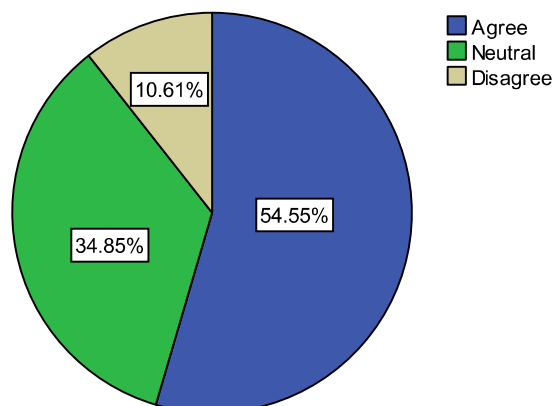
Figure (5.11) Frequencies of the Students' Questionnaire: Item 10

Item 11: This point reveals some association to item 10. As shown in Table (5.12), more than half the students took a neutral stand showing no final view regarding the overall judgement they got from their peers. This is normal, since the peerings of some students were not stable along the study, and some of them spoke about this point directly to the researcher. One third of the total number of the students (36.92%, 24 students) were satisfied with the assessments of their peers, while only 10.77%.of the students (7 students) showed a disagreement on this point.

Table (5.12) Frequencies of the Students' Questionnaire: Item 11

Item 11: The overall assessment my peers gave me were fair and reasonable		Frequency	Percent
Valid	Agree	24	36.92
	Neutral	34	52.31
	Disagree	7	10.77
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
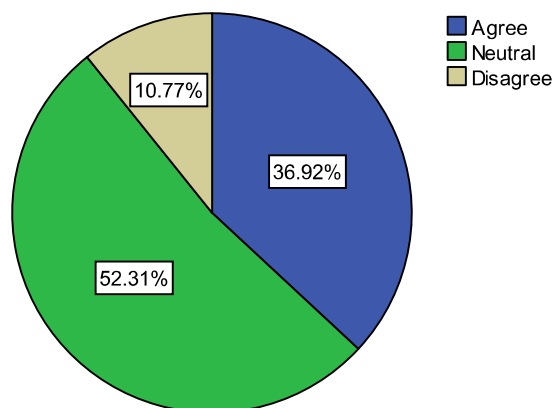
Figure (5.12) Frequencies of the Students' Questionnaire: Item 11

Item 12: From Table (5.13), it is clear that a high percentage of agreement (73.85%) reflects the impact of the peer assessment in raising the awareness among the students of their own skills. Paying attention to the peers' performance enabled the assessors to compare their own performance to their

peers in a way of improvement. However, a neutral percentage of 23.08% was held by 15 students reflecting an unclear opinion towards the peer assessment practice. Again, one student missed answering this question, which may be due to her/his little interest in doing so.

Table (5.13) Frequencies of the Students' Questionnaire: Item 12

Item 12: Assessing other students' speaking skill helped me plan to improve my own			
		Frequency	Percent
Valid	Agree	48	73.85
	Neutral	15	23.08
	Disagree	2	3.08
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
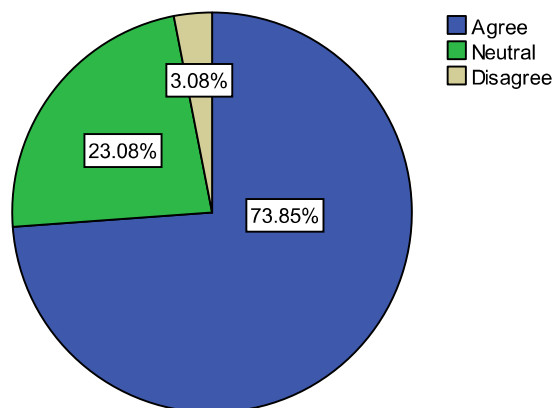
Figure (5.13) Frequencies of the Students' Questionnaire: Item 12

**Item 13**: Having Item 3 asked the students about knowing their strengths and weaknesses throughout using the rubric, Item 13 asked if the assessments of their peers raised their awareness of their weaknesses and strengths. As shown in Table (5.14), while more than half the students (34 students, 52.3%) showed agreement on this case, only five of them disagreed. Still, more than one third of the students (26 students, 39.4%) had a neutral view. This may be attributed to the same reason for the unequal peering, especially when taking into consideration the relatively same percentage gained for the neutral choice.

Table (5.14) Frequencies of the Students' Questionnaire: Item 13

Item 13: PA comments from the first use made me aware of my weaknesses and strengths		Frequency	Percent
Valid	Agree	34	52.31
	Neutral	26	40.00
	Disagree	5	7.69
	Total	65	98.50
Missing	9.00	1	1.50
Total		66	100.00
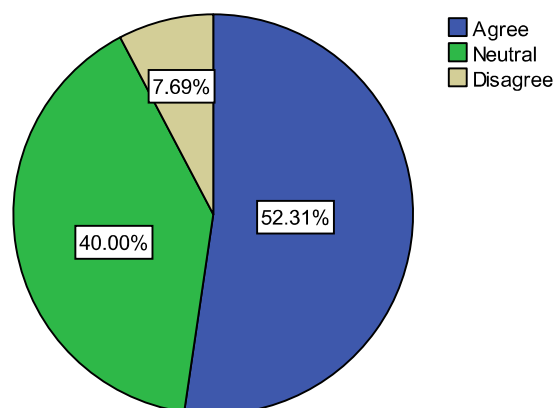
Figure (5.14) Frequencies of the Students' Questionnaire: Item 13

<u>***Item 14***</u>*:* Having applied the group discussion technique, the presence of the teacher is restricted to one group at a given time. Establishing such an interactive class may limit the teacher's role as an observer of all the students in the classroom simultaneously. In this case, applying the peer assessment is very significant. The students play their roles as observers in the groups and that guarantee the class to act within the classroom restrictions.

As shown in Table (5.15) given below, a significant percentage of 62.1% (42 students) agreed on having performed in their best even in the absence of the teacher. This reflects an important objective of peer assessment and encourages its use as a primary tool within the group technique discussion. However, less than one-third (28.8%) of the students, had a neutral view

concerning this point. Only six students (9.1%) disagreed showing the need for the presence of the teacher to observe their performance.

Table (5.15) Frequencies of the Students' Questionnaire: Item 14

Item 14: Peer assessment guarantees we speak in our best performance even with the absence of the teacher's observation		
	Frequency	Percent
Valid Agree	41	62.1
Neutral	19	28.8
Disagree	6	9.1
Total	66	100.0
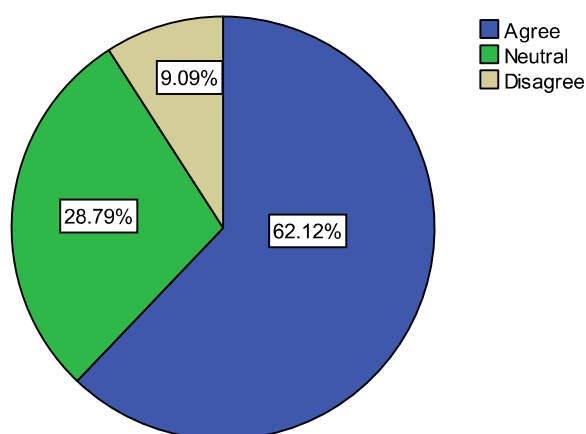
Figure (5.15) Frequencies of the Students' Questionnaire: Item 14

_**Item 19**_*:* The purpose behind this item is to get the students' final judgement concluding their experience in practicing peer assessment. As shown in Table (5.16), a disagreement of 40.9% on this item among the students showed agreement on participate in the practice of peer assessment. More than one-third of the students had a neutral view of this point, which gives support to the previous neutral percentages of the use of peer assessment. As a recommendation of the participants, the practice of peer assessment is proved successful with the use of the group discussion technique. However, one student missed answering this item which might be due to little interest in giving a final judgement of the experience.

Table (5.16) Frequencies of the Students' Questionnaire: Item 19

Item 19: Students should not be involved in assessing peers; assessment should be solely the teachers' job		Frequency	Percent
Valid	Agree	16	24.2
	Neutral	22	33.3
	Disagree	27	40.9
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
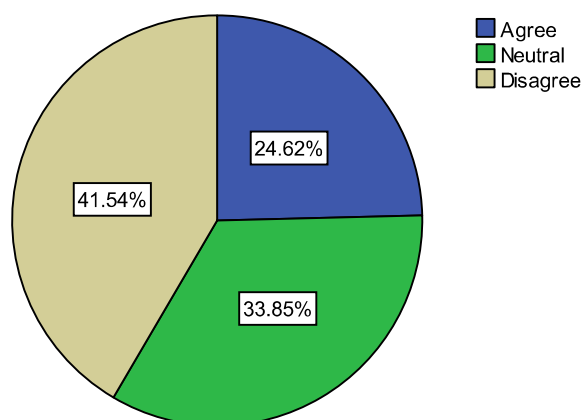
Figure (5.16) Frequencies of the Students' Questionnaire: Item 19

5.2.3. Students' Perspectives of the Group Discussion Technique

Item 15: Leading the group discussion throughout the lessons aimed to build up confidence within the students to perform as teachers (since they are in the preparation-process to become future teachers). As shown in Table (5.17), a significant percentage of the participants of 74.24% (49 students) agreed on the positive impact of this technique, which reflected a high acceptance among the students of its use. This was particularly important to train the shy students for leading their sub-groups, since many of them (especially females) had some difficulty concerning this point. Thus, leading a group of 4-5 students was encouraging and motivating, primarily when the group consists of a number of acquainted friends. Still, the neutral percentage is repeated for this item representing no exact view concerning the group discussion technique.

Table (5.17) Frequencies of the Students' Questionnaire: Item 15

Item 15: Leading my group discussion gave me more self-confidence to act as a teacher		Frequency	Percent
Valid	Agree	49	74.24
	Neutral	15	22.73
	Disagree	2	3.03
	Total	66	100.0
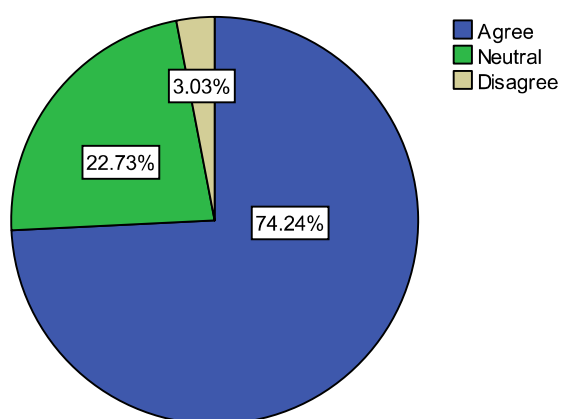
Figure (5.17) Frequencies of the Students' Questionnaire: Item 15

***Item 16*:** Another point supporting the use of group discussion is presented in Item 16 for establishing an interactive classroom. As shown in Table (5.18), the high percentage of agreement of 50 students (76.9%) showed a positive perspective of the use of the group discussion technique. However, 11 students having a neutral view, reflected no exact preference of the use of this technique. Four students (6.1%) disagreed on the use of this technique in the classroom discussion.

Table (5.18) Frequencies of the Students' Questionnaire: Item 16

Item 16: I felt more comfortable in discussing the topics with my classmates through group discussions than merely with the teacher			
		Frequency	Percent
Valid	Agree	50	76.92
	Neutral	11	16.92
	Disagree	4	6.15
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
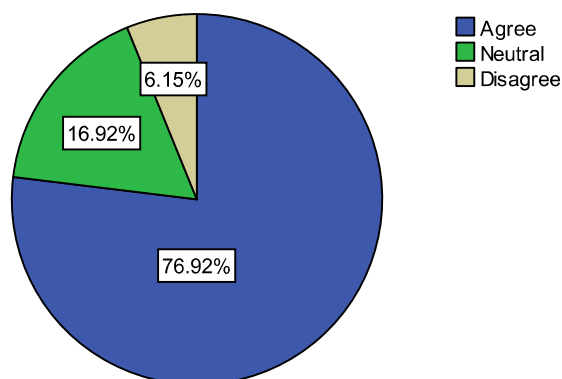
Figure (5.18) Frequencies of the Students' Questionnaire: Item 16

Item 17: The interaction of the students inside the classroom was so obvious to reflect the high percentage of agreement (80.4% of 53 students) of the use of the group discussion technique, as shown in Table (5.19). Their willingness to participate in each class assured such a percentage, unlike their passive views of the interview technique they used to practice earlier. Still, 13.6% of the students (9 students) had a neutral view, showing no absolute preference of the discussion technique. Some of the students were uncomfortable- in some lessons- to be within certain sub-groups. While some students stuck to their own sub-groups, some others had several changes due to the absentees in their teams, which had some impact on the students' views. In addition to that, two students did not answer this question, which might be due to uninterest in answering this question.

Table (5.19) Frequencies of the Students' Questionnaire: Item 17

Item 17: Group discussion encouraged me to participate in each lecture's discussion/ activity			
		Frequency	Percent
Valid	Agree	53	80.4
	Neutral	9	13.6
	Disagree	2	3.0
	Total	64	97.0
Missing	9.00	2	3.0
Total		66	100.0
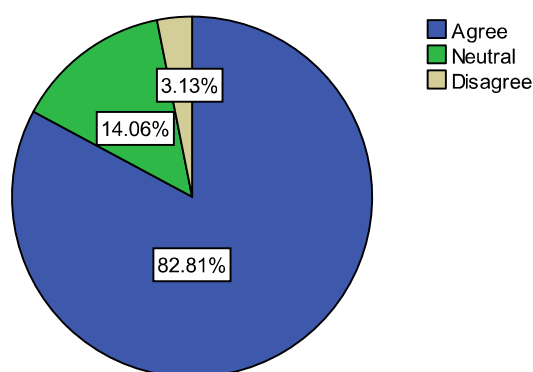
Figure (5.19) Frequencies of the Students' Questionnaire: Item 17

***Item 20*:** Although the majority of the students (72.7%) showed a comfort in participating in their group discussions (as presented earlier in Table (5.18)), still there is a percentage of 18.2% of the students (12 students) who had a neutral view about recommending the use of the group discussion technique in conversation classes (see Table (5.20)). However, one student did not answer this question, which may be due to her/his disinterest in giving a final judgement or answering this point. Nevertheless, the high percentage of agreement of 72.7% (48 students) among the participants showed an acceptance of using the group discussion technique in conversation classes.

Table (5.20) Frequencies of the Students' Questionnaire: Item 20

Item 20: I recommend using the group discussion technique in conversation classes		Frequency	Percent
Valid	Agree	48	72.7
	Neutral	12	18.2
	Disagree	5	7.6
	Total	65	98.5
Missing	9.00	1	1.5
Total		66	100.0
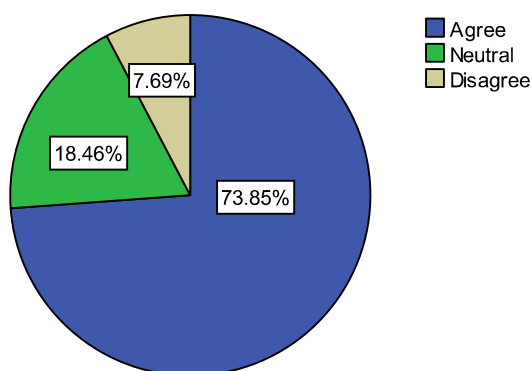
Figure (5.20) Frequencies of the Students' Questionnaire: Item 20

5.2.4. Students' Comments Added to the Questionnaire

In the last point of the students' questionnaire, the students were asked to give their comments- if they have any- concerning their experience in practicing the rubric, the peer assessment, and the group discussion technique. Only 18 students added their comments, which were classified into positive, negative, and suggestions. Table (5.21) shows the classification of the students' comments.

Table (5.21) Classification Frequencies of the students' comments added to the questionnaire

Comments	N	Percent of Cases
+ comment	13	100.0%
suggestions	3	23.1%
- comment	2	15.4%
Total	18	138.5%

Some of the positive comments repeated a number of the items of the questionnaire describing the advantages of the three practices mentioned earlier. Furthermore, whereas some other positive comments expressed likes and thanks for taking part in the experience, others added statements that are more informative. Some of the suggestions repeated some items of the questionnaire and others gave new ideas. On the contrary, the negative comments were direct in showing disinterest in the experience. The following lines list some of the positive comments, the suggestions, and all the negative comments of the participants (**verbatim**)[3].

a. *Positive Comments*

1. *"Using the rubric is a very useful way in conversation lecture. This way encourages all students to participate in discussion."*

2. *"This test was a great benefit to us. The rubric though us to give our judge faithfully."*

3. *"This is the best way I have known my speaking skills though it. The group discussion has the main job of developing my speaking skills. Peer assessment was a good but not so much."*

4. *"The rubric make us develop in speaking, makes us avoid use Arabic. The rubric force us to speak even if we dislike to speak."*

5. *"I think that it is useful for us to assess our skills. Also I like the group discussion, it makes the connection between us more easly. And it is fix our mistakes."*

6. *"It helps me to improve my influency. It makes me how to know my level. It helps me to know more vocabulary."*

b. *Suggestions*

1. *"rubric is something useful, but it needs long time to understand it."*

[3] The comments of the students are listed here verbatim (i.e. listing exactly the same words of the students' comments found in the questionnaire)

2. *"We should have more time in discussion."*

3. *The peers should be changed every week. The topic should be known from the preceding lecture."*

c. *Negative Comments*
1. *"Frankly speaking, I don't like the rubric. I think it is the job of the teacher to assess the students."*

2. *"My point of view that discussion with teacher will be more useful than with students because the teacher will correct my mistakes in speaking. Thank you."*

5.3.   Interrater Reliability Test

It is evident that the reliability of a certain scale will lead to its validity (Peterson, 1994: 381). Thus, checking the reliability of the rubric designed for the study is of more than one advantage. Hence, the results gathered from the interrater reliability test, which was conducted by ten teachers in rating six students, were analyzed statistically by two measures: Cronbach's alpha and the Coefficient of Variation. Cronbach's alpha was used to measure the interrater reliability of the analytic rubric designed by the researcher for the study and the Coefficient of Variation was used in comparing the two types of assessments given by the ten raters in the interrater reliability test, namely, the rubric and the traditional assessment.

The following sections discuss the results of the two types of measures.

5.3.1. Interrater Reliability of the Rubric

Cronbach's alpha was developed by Cronbach in 1951 as a basic measure for internal consistency for rating scales with multi-criteria and became the most widely used measure for estimating the internal consistency of rating scales (ibid: 382). What characterizes Cronbach's alpha among the other reliability measures is that it is used to measure the interrater (and the intrarater) reliability among multiple raters, while the other measures (like Pearson's correlation and

Spearman rank order correlation) are used to measure the interrater reliability between two variables (raters) only.

For a rating scale to achieve an acceptable interrater reliability with the Cronbach's alpha measure, the alpha should be "within the range of 0.50- 0.92, with most values above 0.70". (Jonsson and Svingby, 2007: 135) However, if the Cronbach's alpha is low, then it means that "the majority of the variance in the total composite score is really due to error variance and not to true score variance (Crocker and Algina, 1986; cited in Stemler, 2004, n.p.). The following sections tackle the Cronbach's alpha measuring the rubric as a whole and then, criterion by criterion.

5.3.1.1. Interrater Reliability of the Overall Rubric

As introduced in Chapter Four, the levels of each criterion of the rubric were given specified scores and the total score of the rubric is 20. To make possible their statistic analysis, the assessments of the six students were converted into the final scores according to the values of each criterion. Having the summation of each student, the final scores were analyzed using the Cronbach's alpha measure for the test of interrater reliability. This section deals with the analyses of the overall interrater reliability of the rubric using the total scores of all the students given by the ten raters who are labeled numerically from 1 to 10.

a. *Overall alpha:* As the results in Table (5.22) show, the overall alpha is .934, which is very high and it indicates a strong internal consistency among the ten raters. Accordingly, the analytic rubric designed for the study has proved to be reliable and in conclusion valid as well.

Table (5.22) The overall statistics of the rubric

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.934	.951	10

b. Corrected Item-Total Correlation: In Table (5.23) given below, the column "Corrected Item-Total Correlation" displays the correlation between the score of a given item (rater) and the sum score of the other nine items (raters). The results shown in the column reflect a strong, positive correlation between the scores of one rater and the combined scores of the other nine raters. For example, the correlation between rater 6 and the sum of the other nine raters is 0.981, which means a strong positive correlation. However, raters 1, 3, 4, and 7 show the lowest correlations with the other raters in the corrected item-total correlation.

Table (5.23) Item-Total Statistics of the ten raters

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	117.250	206.275	.630	.936
2	118.333	234.067	.788	.931
3	119.667	208.867	.637	.934
4	117.000	227.400	.675	.931
5	118.167	223.767	.906	.925
6	119.083	203.442	.981	.915
7	116.333	216.867	.650	.931
8	119.000	179.300	.966	.914
9	120.000	196.300	.807	.924
10	119.417	226.342	.856	.927

c. Cronbach's Alpha if Item Deleted: This column displays the Cronbach's alpha measure that would result if a given item (rater) was deleted. Like the item-total correlation, this column of information is valuable for determining which items (raters) from among a set of items (the 10 raters) contributes to the total alpha. Hence, the value presented in this column represents the alpha value if the given rater was not included. This approach can identify the raters who detracted from the overall interrater reliability and show the poorer raters whose ratings, if deleted, will cause the alpha to increase.

According to the above discussion, it is evident from Table (5.23) that the alpha, if rater 8 is deleted, would drop from the overall .934 to .914. Since the alpha would drop with the removal of rater 8, then this rater appeard to be useful and contributed to the overall reliability of the ratings. On the contrary, rater 1 seems less useful. The alpha would increase from .934 to .936 if rater 1 was deleted. Also, the alpha would remain the same if rater 3 was deleted, which means that the ratings of rater 1 and 3 do not contribute much to the reliability of the rubric assessment. Nevertheless, since the differences between the overall alpha and the 'alpha if item deleted' are not significantly high, then there is no real reason to drop the two raters from the total interrater reliability test.

5.3.1.2. Interrater Reliability of the Detailed Criteria of the Rubric

In an analytic rubric with different criteria, it is significant to find out which criterion has contributed to the overall reliability of the rubric. Thus, Cronbach's alpha was applied again to each of the eight criteria of the rubric, namely: fluency, grammar, pronunciation, no use of Arabic, use of English, interaction, content, and humor.

It is found that the reliability varies from a criterion to another according to significance differences in alpha. However, prior to the interrater reliability test, some teachers showed disagreement on the group of the students selected for this issue. To clarify, the researcher explained that she is not concerned with the levels of the students, but with the adequate use of the rubric. In this regard, raters 3, 4, and 8 showed their personal impression concerning the students prior to watching the video, which means their rubric assessment might be affected by their general impression.

The following sections deal with the Cronbach's alpha of each criterion.

1. Fluency: Tables (5.24) and (5.25) illustrate the alpha results of the fluency assessment. From Table (5.24), it appears that the alpha of fluency is .364, which means a weak alpha among the ten raters and reflects a lack of interrater reliability of fluency.

Tables (5.24) Reliability Statistics of Fluency

Cronbach's Alpha	N of Items
.364	10

As shown in Table (5.25) given below, it can be seen that rater 8 is the most reliable rater. While the corrected item-total correlation of rater 8 is .926, which means a high correlation, the Cronbach's alpha if rater 8 was deleted would be decreased from .364 to -.201. On the contrary, whereas the corrected item-total correlation of rater 4 is -.280 (which means a lack of correlation), the Cronbach's alpha if rater 4 is deleted would be increased from .364 to .468.

Tables (5.25) Item-Total Statistics of Fluency

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total correlation	Cronbach's Alpha if Item Deleted
1	9.000	.800	.000	.445
2	9.167	.967	.000	.369
3	9.167	.767	.000	.465
4	9.083	1.042	-.280	.468
5	9.167	.967	.000	.369
6	9.167	.967	.000	.369
7	9.167	.767	.181	.318
8	9.167	.467	.926	-.201[a]
9	9.250	.675	.745	.111
10	9.167	.967	.000	.369

a. The value is negative due to a negative average covariance among items.

2. Grammar: Tables (5.26) and (5.27) illustrate the alpha results of the grammar assessment. From Table (5.26), it appears that the alpha of grammar is .779, which means a good alpha among the ten raters and reflects a good interrater reliability of grammar in turn.

Tables (5.26) Reliability Statistics of Grammar

Cronbach's Alpha	N of Items
.779	10

As shown in Table (5.25), raters 1 and 4 are the most significant in the interrater reliability of grammar. While the Cronbach's alpha if rater 1 is deleted increases from .779 to .804, it would be decreased if rater 4 was deleted. In opposition to fluency assessment, rater 4 is seen more reliable in grammar assessment.

Tables (5.27) Item-Total Statistics of Grammar

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	10.250	3.475	.000	.804
2	10.167	3.267	.202	.786
3	10.417	2.742	.405	.773
4	10.167	1.967	.884	.677
5	10.250	2.975	.561	.750
6	10.250	2.475	.701	.720
7	10.333	3.367	.178	.785
8	10.417	2.742	.668	.732
9	10.083	2.842	.728	.732
10	10.417	3.542	.000	.789

3. Pronunciation: Tables (5.28) and (5.29) show the alpha results of the pronunciation assessment. From Table (5.28), it appears that the alpha of pronunciation is .389, which means a weak alpha among the ten raters and reflects a lack of interrater reliability.

Table (5. 28) Reliability Statistics of Pronunciation

Cronbach's Alpha	N of Items
.389	10

As shown in Table (5.29) given below, individual raters 1, 2 and 3 correlate negatively with the other nine raters. However, raters 6 and 9 are the most contributing raters to the alpha since the deletion of either of them would cause the alpha to decrease into negative values.

Table (5.29) Item-Total Statistics of Pronunciation

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	10.667	1.267	-.363	.503
2	11.000	1.200	-.224	.469
3	11.083	1.342	-.386	.671
4	10.667	.767	.933	.098
5	10.583	1.142	.000	.394
6	10.833	.567	.635	$-2.873\text{E-}14^{a}$
7	10.833	.867	.392	.260
8	10.917	.942	.266	.319
9	11.083	.642	.790	$-2.523\text{E-}14^{a}$
10	11.083	1.142	.000	.394

a. The value is negative due to a negative average covariance among items.

4. No Use of Arabic: Tables (5.30) and (5.31) illustrate the alpha results of the no use of Arabic assessment. From Table (5.30), it appears that the alpha of this criterion is .705, which means a good alpha reflecting a good interrater reliability of the no use of Arabic among the ten raters.

Table (5.30) Reliability Statistics of the No Use of Arabic

Cronbach's Alpha	N of Items
.705	10

From Table (5.31), it is clear that raters 4 and 8 contribute highly to the overall alpha of the no use of Arabic since the alpha decreases from .705 to .554 if they were deleted. On the contrary, if any of the raters 1, 2, 5, 6, 7, and 9 were deleted, the alpha would increase to .714.

Table (5.31) Item-Total Statistics of the No Use of Arabic

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	33.67	3.467	.000	.714
2	33.67	3.467	.000	.714
3	34.00	2.400	.500	.656
4	33.83	2.167	.943	.554
5	33.67	3.467	.000	.714
6	34.67	3.467	.000	.714
7	33.67	3.467	.000	.714
8	33.83	2.167	.943	.554
9	34.33	1.467	.674	.665
10	33.67	3.467	.000	.714

5. Use of English: Tables (5.32) and (5.33) show the alpha results of the use of English assessment. From Table (5.32), it appears that the alpha is .822, which means a very good alpha reflecting a very good interrater reliability of the use of English in turn.

Table (5.32) Reliability Statistics of the Use of English

Cronbach's Alpha	N of Items
.822	10

From Table (5.33), it is evident that raters 6, 7, 8, and 9 contribute to the overall alpha of English since the alpha decreases if any of them was deleted. Conversely, the alpha would increase if any of raters 1, 2, 3, 4, 5, and 10 was deleted.

Table (5.33) Item-Total Statistics of the Use of English

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	9.167	3.467	.190	.844
2	9.250	3.875	.000	.832
3	9.500	3.700	.095	.841
4	9.167	3.167	.423	.817
5	9.167	3.467	.482	.811
6	9.250	2.875	.839	.770
7	9.167	2.567	.967	.745
8	9.083	2.542	.896	.752
9	9.250	2.475	.853	.758
10	9.250	3.875	.000	.832

6. Interaction: Tables (5.34) and (5.35) show the alpha results of the interaction assessment. From Table (5.34), it appears that the alpha is .946, which means a very high alpha reflecting a very high interrater reliability of interaction as well.

Table (5.34) Reliability Statistics of Interaction

Cronbach's Alpha	N of Items
.946	10

From Table (5.35) given below, it can be seen that there is no real significance in the alpha if any of the raters was deleted. In this case, raters 6 and 10 are the ones contributing mostly to the alpha, whereas raters 7 and 9 affect the alpha negatively since the deletion of any of them causes the alpha to increase.

Table (5.35) Item-Total Statistics of Interaction

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	20.583	47.042	.827	.939
2	20.250	52.175	.797	.940
3	21.167	46.567	.894	.935
4	19.667	53.867	.890	.940
5	21.083	47.642	.820	.939
6	20.500	49.100	.957	.933
7	19.583	54.242	.456	.955
8	21.000	49.600	.913	.934
9	21.167	58.967	.483	.952
10	21.250	47.975	.923	.933

7. Content: Tables (5.36) and (5.37) show the alpha results of the content assessment. From Table (5.36), it appears that the alpha is .761, which means a good alpha reflecting a good interrater reliability of content in turn.

Table (5. 36) Reliability Statistics of Content

Cronbach's Alpha	N of Items
.761	10

As shown in Table (5.37) given below, the rater 8 is the most contributing rater to the alpha, since the alpha would increase from .761 to .619 if it was deleted. However, though not significantly high, the alpha would increase from .761 to .787 if the rater 2 was deleted.

Table (5.37) Item-Total Statistics of Content

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	11.250	3.675	.467	.735
2	11.333	4.767	-.150	.787
3	11.500	3.700	.497	.730
4	11.333	4.267	.435	.747
5	11.750	4.675	.000	.770
6	11.667	3.267	.657	.700
7	10.917	4.342	.248	.760
8	11.583	2.242	.961	.619
9	11.667	3.267	.657	.700
10	11.750	4.675	.000	.770

8. Humor: Tables (5.38) and (5.39) show the alpha results of the humor assessment. From Table (5.38), it is evident that the alpha is -.132, which means a negative alpha reflecting a lack of interrater reliability of humor among the ten raters.

Table **(5.38)** Reliability Statistics of Humor

Cronbach's Alpha[a]	N of Items
-.132	10

a. The value is negative due to a negative average covariance among items.

From table (5.39) given below, while the alpha increases if any of the raters 1, 2, 4, and 8 was deleted, it decreases if any of the raters 3, 6, 7, 9, and 10 was deleted, and decreases significantly if the rater 5 was deleted.

Table **(5.39)** Item-Total Statistics of Humor

Rater	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	13.833	.867	-.196	3.947E-14
2	14.500	.900	-.258	3.360E-14
3	14.000	.700	.293	-.321[a]
4	14.000	.900	-.258	3.360E-14
5	13.667	.867	-.175	-.043[a]
6	13.917	.542	.614	-.692[a]
7	13.833	.567	.485	-.596[a]
8	14.167	1.067	-.469	.492
9	14.333	.567	.159	-.397[a]
10	14.000	.700	.293	-.321[a]

a. The value is negative due to a negative average covariance among items.

5.3.2. Comparison between the Two Types of Assessment: the Rubric and the Traditional Assessment

In order to make a comparison between the two types of assessment, i.e. the rubric and the traditional (summative) assessment, a Coefficient of Variation was performed between the total rubric scores and the traditional scores given by the ten raters to the six students at the end of the interrater reliability test. Figure (5.21) illustrates the means of the two types of assessment; the rubric scores and the traditional scores.
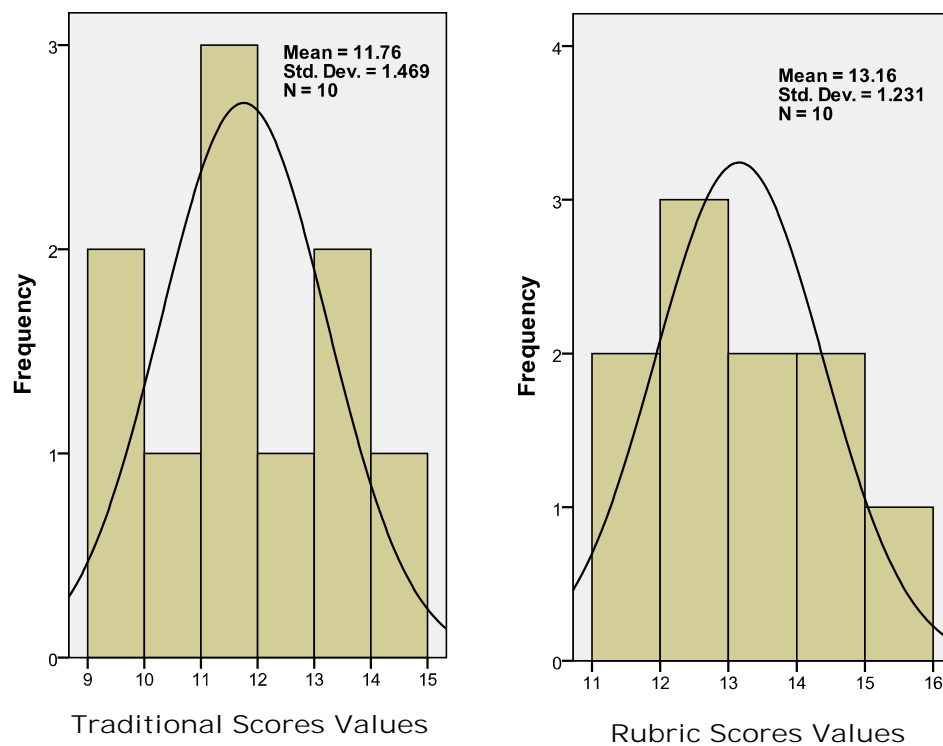
Figure (5.21) The Means of the Students' Traditional and Rubric Scores

In Figure (5.21), the horizontal dimension represents the scores values given to the students and their range of the total score (i.e. 20), whereas the vertical dimension shows the frequency of a certain score given by the teachers. Hence, from the norm curve of the two histograms, it can be seen that the scores of the rubric assessment are less dispersed from the scores of the traditional

assessment and much concentrated on the centre of the norm curve. This gives the hypothesis that the rubric assessment is much consistent than the traditional assessment. To check the validity of this hypothesis, the coefficient of variation measure should be calculated. The coefficient of variation measure ($\hat{C}v$) is defined as the ratio of the standard deviation ( ) to the mean ($\bar{X}$) multiplied by 100%. Thus, the $\hat{C}v$ equation will be:

$$\hat{C}v = \frac{S}{X} * 100\%$$

Since the coefficient of variation measure is one of the dispersion measures, then the lower its value, the better its meaning. Accordingly, the $\hat{C}v$ of the rubric assessment and the traditional assessment will be calculated as follows.

$$Rubric\ \hat{C}v = \frac{1.231}{13.16} * 100\% = 9.3541$$

$$Traditional\ \hat{C}v = \frac{1.469}{11.76} * 100\% = 12.4914$$

Noting the results of the two calculations, it is concluded that the $\hat{C}v$ of the rubric assessment is lower than that of the traditional assessment, which gives evidence that the overall rubric assessment was better than the traditional assessment.

In the following sections, the same procedure is followed to compare the consistency of the two types of assessment to each of the six students in the reliability test. The students are named according to the initials of their first names as B, Kh, D, A, E, and J.

1. *Coefficient of Variation of Student B*

Table (5.40) given below shows the descriptive statistics of student B.

Table (5.40) Descriptive Statistics of Student B

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
B_ Rubric	10	11.5	18.0	14.800	1.7981	12.1493
B_ Traditional	10	9	17	12.90	2.923	22.6589
Valid N (listwise)	10					

From Table (5.40), it is found that the rubric assessment is more consistent for student B than the traditional assessment because the value of $\hat{C}v$ of the rubric assessment is less than that of the traditional assessment.

Figure (5.22) given below shows the scores distribution of the rubric assessment and the traditional assessment of student B.
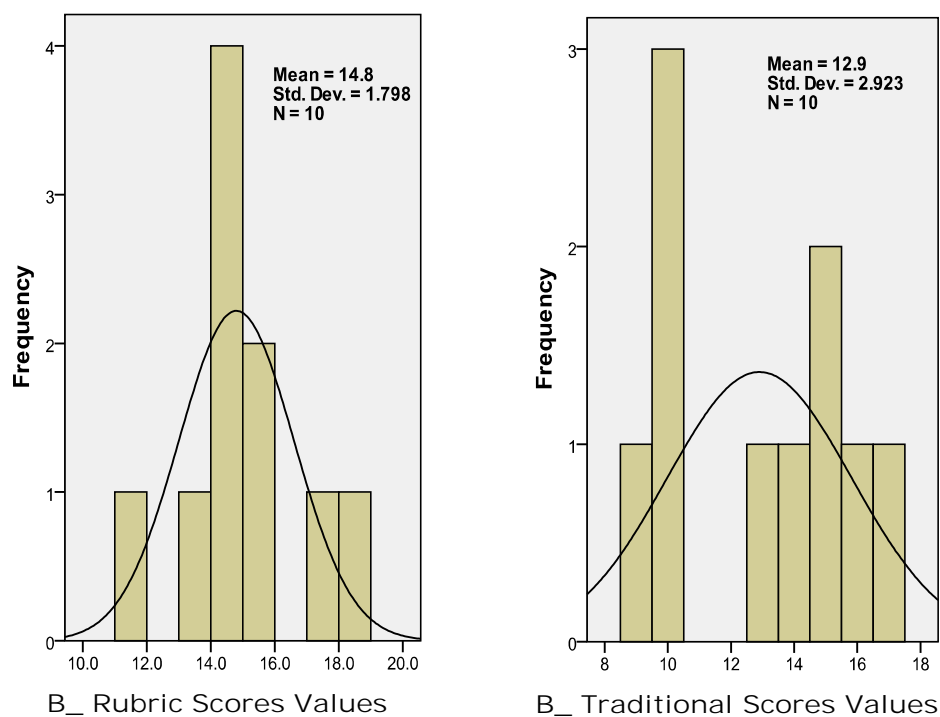

Figure (5.22) Traditional and Rubric scores of Student B

From Figure (5.22), it can be seen that the scores of student B in the rubric assessment histogram are less dispersed and more concentrated on the centre than those in the traditional assessment histogram.

2. *Coefficient of Variation of Student Kh*

Table (5.41) given below shows the descriptive statistics of student Kh.

Table (5.41) Descriptive Statistics of student Kh

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
Kh_ Rubric	10	10.0	17.0	12.800	2.3118	18.06
Kh_ Traditional	10	9	15	12.00	2.000	16.66
Valid N (listwise)	10					

From Table (5.41), it is found that the traditional assessment is more consistent for student Kh than the rubric assessment because the value of $\hat{C}v$ of the traditional assessment is less than that of the rubric assessment.

Figure (5.23) given below represents the scores distribution of the rubric assessment and the traditional assessment of student Kh.
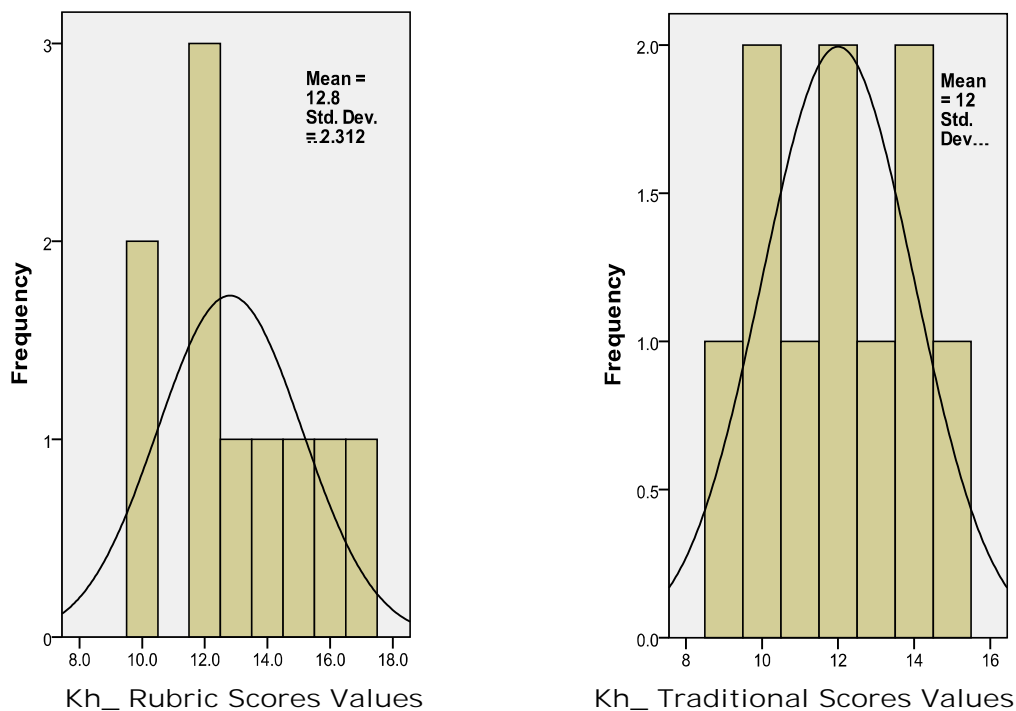

Figure (5.23) Traditional and Rubric Scores of Student Kh

From Figure (5.23), it is evident that the scores of student Kh in the traditional assessment histogram are less dispersed and more concentrated on the centre than those in the rubric assessment histogram.

3. *Coefficient of Variation of Student D*

Table (5.42) given below shows the descriptive statistics of student D.

Table (5.42) Descriptive Statistics of student D

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
D_ Rubric	10	13.5	17.0	15.350	1.1316	7.371
D_ Traditional	10	12	18	14.40	1.776	12.347
Valid N (listwise)	10					

From Table (5.42), it is found that the rubric assessment is more consistent for student D than the traditional assessment because the value of $\hat{C}v$ of the rubric assessment is less than that of the traditional assessment.

Figure (5.24) given below shows the scores distribution of the rubric assessment and the traditional assessment of student D.

Figure (5.24) Traditional and Rubric scores of Student D

From Figure (5.24), it is obvious that the scores of student D in the rubric assessment histogram are less dispersed and more concentrated on the centre than those in the traditional assessment histogram. This judgement might not be so clear from examining the histograms for the first time, but taking into consideration the range of the scores of each type of assessment will give a clear understanding to this point. From the two histograms, it can be noted that in the rubric scoring, the scores values extend between the values 13.5- 17 (i.e., the scores range is only 3.5 scores), whereas the scores values of the traditional assessment extends between the values 12- 18 (i.e., the scores range is 6 scores). Accordingly, it is evident that the rubric assessment is more consistent to student D than that of the traditional assessment.

4. *Coefficient of Variation of Student A*

Table (5.43) given below shows the descriptive statistics of student A.

Table (5.43) Descriptive Statistics of student A

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
A_ Rubric	10	7.50	15.00	11.0500	2.2663	20.5095
A_ Traditional	10	7.00	12.00	9.4000	1.5776	16.7829
Valid N (listwise)	10					

From Table (5.43), it is found that the traditional assessment is more consistent for student A than the rubric assessment because the value of $\hat{C}v$ of the traditional assessment is less than that of the rubric assessment.

Figure (5.25) given below illustrates the scores distribution of the rubric assessment and the traditional assessment of student A.
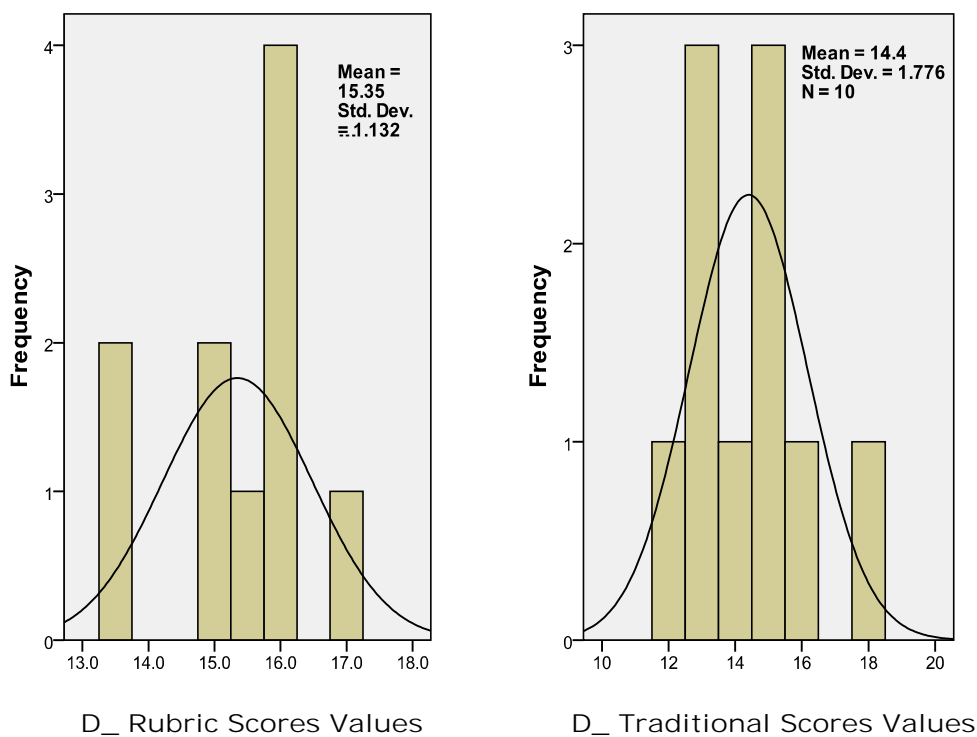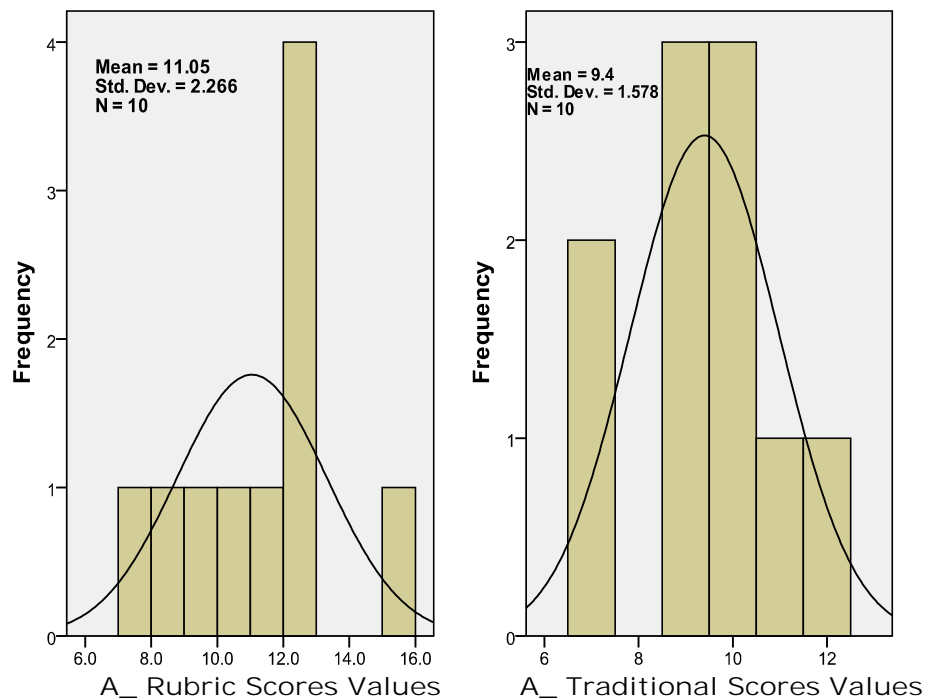

Figure (5.25) Traditional and Rubric Scores of Student A

From Figure (5.25), it is evident that the scores of student A in the traditional assessment histogram are less dispersed and more concentrated on the centre than those in the rubric assessment histogram.

5. *Coefficient of Variation of Student E*

Table (5.44) given below shows the descriptive statistics of student E.

Table (5.44) Descriptive Statistics of student E

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
E_ Rubric	10	11.00	16.00	12.5000	1.4337	11.4696
E_ Traditional	10	9.00	12.00	10.4000	1.0750	10.3365
Valid N (listwise)	10					

From Table (5.44), it is found that the traditional assessment is more consistent for student E than the rubric assessment because the value of $\hat{C}v$ of the traditional assessment is less than that of the rubric assessment.

Figure (5.26) given below shows the scores distribution of the rubric assessment and the traditional assessment of student E.
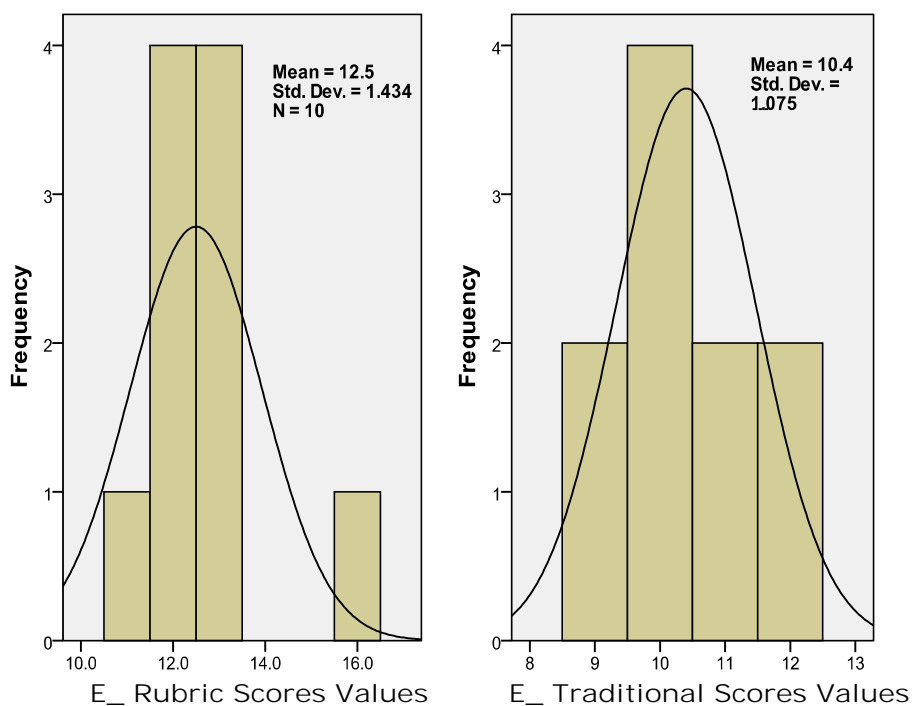

Figure (5.26) Traditional and Rubric scores of Student E

From Figure (5.26), it is apparent that the scores of student E in the traditional assessment histogram are less dispersed and more concentrated on the centre than those in the rubric assessment histogram.

6. Coefficient of Variation of Student J

Table (5.45) given below shows the descriptive statistics of student J.

Table (5.45) Descriptive Statistics of student J

	N	Minimum	Maximum	Mean	Std. Deviation	Coefficient of Variation
J_ Rubric	10	12.00	14.00	12.6111	.6509	5.1613
J_ Traditional	10	9.00	15.00	11.7778	1.8559	15.7576
Valid N (listwise)	10					

From Table (5.45), it is found that the rubric assessment is more consistent for student J than the traditional assessment because the value of $\hat{C}v$ of the rubric assessment is much less than that of the traditional assessment.

Figure (5.27) given below shows the scores distribution of the rubric assessment and the traditional assessment of student J.
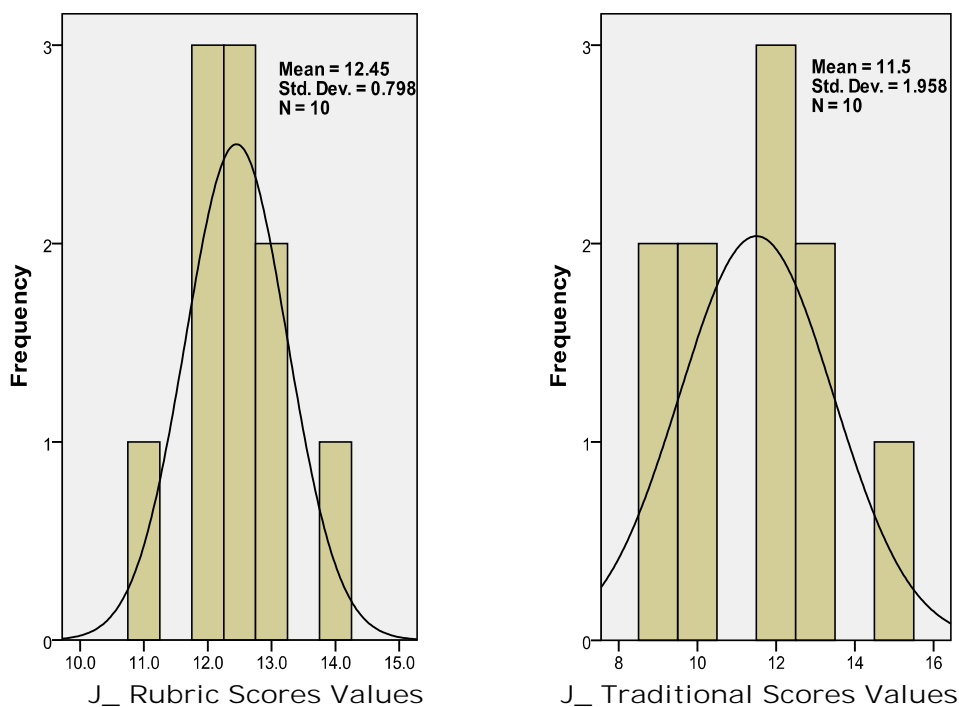

Figure (5.27) Traditional and Rubric Scores of Student J

From figure (5.27), it is clear that the scores of student J in the rubric assessment histogram are less dispersed and more concentrated on the centre than those in the traditional assessment histogram.

5.4. The Pre-Post Tests

The Paired-Samples T Test procedure was performed to measure the differences between the participant' performance before and after applying the study methodology. T Test is the most common test used in comparing the means of two data and it is available 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). Elliott and Woodward (2007:70) state that Paired T Test "computes the differences between values of the two variables for each case and tests whether the average differs from 0."

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 (5.28) shows the plots for the pre and post tests data in a way of simple comparison.
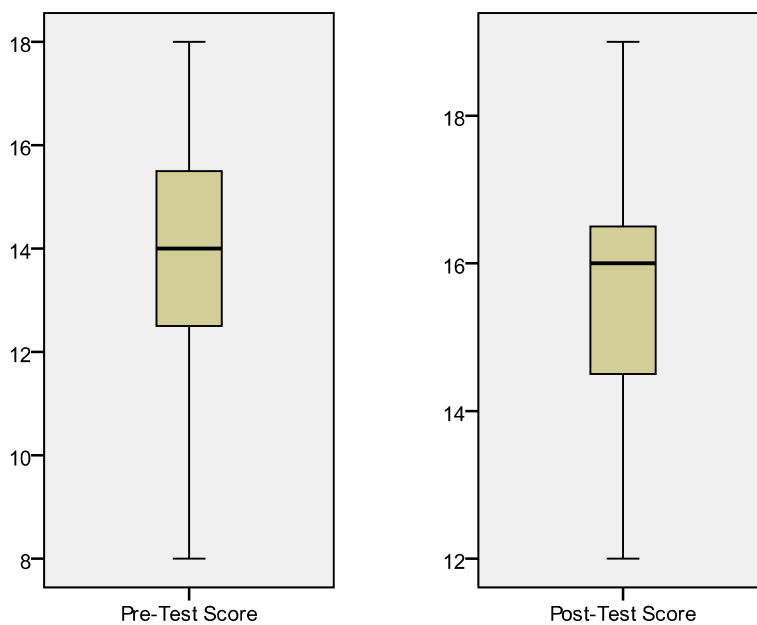

Figure (5.28) Two Boxplots Showing the Distribution Differences of the Pre and Post Tests Scores

From Figure (5.28), it is obvious that the scores of the participants in the pre test are distributed within the range of 8+ to 18, with a concentration of sores at score 14. Yet, 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 participants' performance is 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 in Chapter Four. Although the total number of the participants was 74, the scores of only 66 were computed in the paired *t*-test. This is due to the lack of the scores (of either of the tests) of eight participants (due to their absence in the days of either of the two tests).

To get adequate results, a calculation of *Post minus Pre* tests scores must 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 procedure and the results obtained are shown in Table (5.46) given below.

Table (5.46) Paired Samples Test of the Post -Pre test

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

As illustrated in Table (5.46), the Post - Pre test mean is 1.6716 with a standard deviation of the differences equals 1.8961. The calculated t-statistics with the differences of the 66 scores (66 df) equals 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 was effective in promoting the participants' speaking skill.

CHAPTER SIX: CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS

6.1. Introduction

In this concluding chapter, the overall results of the study investigations are presented for the aim of implying that the aims of the study introduced in Chapter One are accomplished. The following sections summarize the findings of the study with the concluding remarks, propose some recommendations based on the empirical work, and give suggestions for future implementations of similar empirical studies.

As introduced in Chapter One, the main aim of the study was to investigate the impact of using the scoring rubric and peer assessment on promoting Iraqi EFL learners' speaking skill at the university level. The empirical study was limited to a convenience sample of the target population from the Department of English, College of Education for Humanities, University of Basrah. The following sections summarize the conclusions of the study with reference to the problem, hypotheses, and aims of the study introduced in Chapter One.

6.2. Conclusions

6.2.1. General Conclusion with Reference to the Problem

Taking the problem of the study into consideration, the researcher found that the use of the alternative assessment practices, namely the rubric, the peer assessment and the group work technique, generally has a positive effect on each of the teaching, learning, and assessment processes. As far as the learning process is concerned, the students started to be aware of their strengths and weaknesses in speaking. This kind of awareness particularly evolved throughout the use of the rubric in peer assessment. The descriptors of the different levels

of each criterion were influential in informing (and continuously reminding) the students of the objectives meant behind the course. Taking part of their own assessments, the students shared the responsibility for their learning, which is highly encouraged for polishing their professional skills for future employment. As for teaching, the instructions found in the rubric descriptors made it easy for the researcher not to repeat what is demanded of the students in each class. Thus, the rubric was a good instructional device in monitoring the students' interaction as well as behaviours. As far as the assessment process is concerned, the formative assessment of the peers, though not adequate always, was helpful in giving the students formative assessment for their daily performance, which in turn made them eager to participate in class and aware of the quality of their performance. These conclusions reflect the fact introduced in Chapter One that alternative assessment is an integration point between learning, teaching, and assessment.

6.2.2. Conclusions with Reference to the Hypotheses and Aims of the Study

1. In answering the first hypothesis of whether scoring rubrics can promote the students' speaking skill, the researcher implemented a pre- posttest procedure through which she investigated the differences in the students' performance before and after applying the practices of the rubric and the peer assessment. According to the results of the statistical analysis tackling this point in Chapter Five, i.e. the paired T-test, it is found that the rubric and the peer assessment are effective tools in improving the students' speaking skill since they enhance the means of the students' score. In this case, the findings go along with those concluded earlier by Jonsson and Svingby (2007), White (2009), Reddy and Andrade (2010), and Kutlu et al (2010).

2. As far as the second hypothesis is concerned, it is evident from the lectures of conversation that the rubric, when used in peer assessment and group work technique, was successful in creating a highly motivated student- centered classroom. She found that the students, even the very passive ones, lack motivation and practice more than they lack linguistic skills. Applying the alternative assessment, particularly with the group technique, the role of the researcher was no longer the central in the class, but an observer of the activities and performance of the students. However, when used alone, the rubric was not as effective as it was with the application of peer assessment. The students did not realize the importance of the rubric, nor was the researcher able to use it as a fair assessment tool due to the large class. Accordingly, when used with the application of peer assessment, the rubric proved to be a valuable tool of assessment as well as teaching and learning.

3. The third hypothesis deals with the issue of subjectivity in the assessment of speaking. According to the statistical analyses of the Cronbach's alpha, it is found that the overall rubric has achieved a high reliability, which is, in turn, evidence that the rubric has achieved validity (as proposed by Peterson (1994: 381), in Chapter Five, that the reliability of a rating scale leads to its validity). Moreover, the interrater reliability test has shown that the criteria descriptors were clear enough to achieve a good reliability for 'No use of Arabic', 'Grammar', and 'Content', very good reliability for 'Use of English', and excellent reliability for 'Interaction'. However, the unacceptable reliability of 'Fluency' and 'Pronunciation', and the negative reliability of 'Humor' may indicate unclear definitions of the levels of criteria.

Yet, a point of importance is to be mentioned. Not all the teachers participating in the interrater reliability proved to rate the students according to the rubric criteria. Some of them reflected their personal judgements, which they admitted later to the researcher. Others, on the other hand, explained that the rubric was much detailed and that they needed more time of training, since the teachers had only an hour of explanation of the criteria of the rubric and how to use it consistently.

4. As far as the comparison of the rubric and traditional assessment is concerned, i.e. the fourth hypothesis, it is found that the rubric assessment was more reliable than the traditional assessment with three of the assessed students, namely, B, D, and J. On the other hand, the traditional assessment was more reliable than the rubric assessment for the students Kh, A, and E. This result gives evidence to the raters' personal judgements that have affected their rubric assessments. The teachers made several remarks on the last three students in being passive ones, though some of them described student B as such. This is a clear indicator that some teachers have involved their personal judgements in rating some students.

Ultimately, it is evident that a reliable rubric can achieve an objective assessment of the students' speaking skill, if the raters were to receive more training on its use, and more importantly, if they make use of the criteria of the rubric as a guide in the assessment of the students' performance. Accordingly, alternative assessment can be more reliable than traditional assessment.

6.2.3. Summary of Findings

In conclusion to the study, the researcher sums up the findings in the following points (including both positive and negative findings):

1. Scoring rubrics, if shared with students in peer assessment, are very effective tools of assessment and teaching as well as learning.
2. Scoring rubrics, if developed with clear criteria descriptors and teachers have training on applying them, are effective in reducing teachers' subjectivity in the assessment of the speaking skill.
3. Peer assessment is an effective tool in raising the students' motivation, cooperation, and awareness of the quality of their performance, as well as in shaping their profession and confidence.
4. Implementing the group technique with peer assessment is very successful in managing large classes.
5. A high reliability of scoring rubrics is achieved if the criteria descriptors are clear enough and understood by all raters.
6. Setting clear criteria descriptors is a challenging matter, however it is solved by the time of application. The criteria should be developed in a way simulating actual performance of students.
7. Students take time to understand and master peer assessment.
8. Peer assessment needs teachers' continuous observation.

6.3. Recommendations

In the light of the findings of the empirical study, the researcher recommends the following:

1. Teaching alternative assessment practices to EFL students in the College of Education for Humanities, so that they can make use of their application in their future employment.
2. Training the university teachers at the practices of the alternative assessments.
3. Implementing the use of rubrics, peer assessment, and group work technique in conversation classes to help teachers manage large classes and create student-centered classrooms.

4. Developing rubrics for all courses to be handed to students (at least, if not to be implemented in peer assessment) in the first classes. This is primarily for the sake of sharing the goals of the course and the teachers' expectations with the students in advance, and secondly in order to guarantee a transparent assessment of the students' performance.

5. In examining the students' perspectives of the use of rubrics, peer assessment, and group work in conversation classes in the students' questionnaire, it is found that the students have a general acceptance of the use of the three practices. However, more training is recommended to enroll students in the development of rubrics and the application of the peer assessment.

6. Analytic rubrics are highly recommended than holistic rubrics, since they detail the skill being assessed and make clear the reason behind a specific assessment.

6.4. Suggestions for Future Studies

The following points make suggestions for any future similar study:

1. Taking into consideration the few number of males participating in the study, a similar study is suggested to be undertaken with an equal number of participants from both genders.

2. Since the current study covers several practices of the alternative assessment, the scoring of the peer assessment was not considered in the study. The main purpose of the peer assessment was to raise the students' awareness of the quality of their performance and draw their attention to the class activities. So, no real interest was given to the adequacy of the students' peer assessments. Thus, a study devoted to the application of peer assessment is highly encouraged, particularly if compared with teachers' assessment.

3. The practices of rubrics and peer assessment are widely applied worldwide in the assessment of writing and they were found to be effective. Thus, a similar study investigating this point with Iraqi EFL learners is suggested.

4. Interrater reliability can be studied more specifically with a larger number of raters with an option of investigating the raters' bias in the assessment of students from both genders. (The genders of the raters themselves may be involved in such a study).

BIBLIOGRAPHY

5.

Anderson, Lorin W. (2003). Classroom Assessment: Enhancing the Quality of Teacher Decision Making. London: Lawrence Erlbaum Associates, Inc.

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, [Online], 10 (3), April. n.p. Available at: http://pareonline.net/pdf/v10n3.pdf

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.p. Available at: http://cie.ed.asu.edu/volume4/number4

_____ (2005). "Teaching with rubrics: The Good, the Bad, and the Ugly". College Teaching, 53(1), 27-32.

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. Retrieved January, 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 MultiDiscipliniary Evaluation, 6 (12), 52- 67.

Brown, H. Douglas. (2003). Language Assessment: Principles and Classroom Practices. California: Longman Publishing.

Brown, James Dean. (2004). "Formative Assessment: Existing Literature and Directions for Research". Second Language Studies, 22(2) Spring, 91-139.

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.

CITL (Center for Innovative Teaching and Learning). (2007). Rubric Design Guide 'RDG': Establishing Rubrics to Guide Your Grading. Bloomington: Indiana University.

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.

Genesee, Fred and John A. Upshur. (1996). Classroom-based Evaluation in Second Language Education. Cambridge: Cambridge University Press.

Guskey, Thomas R. (2003). "How Classroom Assessment Improve Learning". Educational Leadership, 60 (5), February, 6- 11.

Harmer, Jeremy. (2001). The Practice of English Language Teaching. 3rd ed. London: Longman Press.

Harris, David P. (1969). Testing English as a Second Language. New York: MicGrow-Hill University.

Jonsson, Anders and Gunilla Svingby. (2007). "The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences". Educational Research Review, (2), 130-144.

Kutlu, Omer, Ozen Yildirim, and Safiye Bilican. (2010). "The Comparison of the Views of Teachers with Positive and Negative Attitudes towards Rubrics". Procedia Social and Behavioral Sciences, 9, 1566–1573

Lazaraton, Anne. (2001). "Teaching Oral Skills". In, Marianne Celce-Mucia (Ed.). Teaching English as a Second or Foreign Language, 3rd ed. New York: Heinle & Heinle Publishers Inc.

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. Retrieved Dec. 2011, from http://creativecommons.org/licenses/by-nc-nd/3.0/

Luoma, Sari. (2004). Assessing Speaking. Cambridge: Cambridge University Press.

Moskal, Barbara M. (2000). "Scoring Rubrics: What, When and How?". Practical Assessment, Research and Evaluation, [Online], 7(3), n.p. Retrieved September, 2010 from http://PAREonline.net/getvn.asp?v=7&n=3.

Moskal, Barbara M. and Jon A. Leydens. (2000). "Scoring Rubric Development: Validity and Reliability. Practical Assessment, Research and Evaluation, [Online], 7 (10), n.p. Retrieved 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. Chicago: Addison-Wesley Longman, Inc.

Perlman, Carole L. (2003). "Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics". In: <u>Measuring Up: Assessment for Teachers, Counsellors, and Administrators</u>, 497-506. ERIC: Educational Resources Information Center.

Peterson, Robert A. (1994). "A Meta-Analysis of Cronbach's Coefficient Alpha". <u>Journal of Consumer Research</u>, 21(2), September, 381-391.

Popham, James W. (1997). "What's Wrong- and What's Right- with Rubrics". <u>Educational Leadership</u>, October,72- 75.

Raza, W. (2011). "A Rating Scale to Assess English Speaking Proficiency of University Pakistan". <u>Teaching and Learning Language Current Trends and Practices</u>, 242-256. (Gained via email from the author).

Reddy, Y. Malini. (2007). "Effect of Rubrics on Enhancement of Student Learning". <u>Educate</u>, 7(1), 3-17.

Reddy, Y. Malini and Heidi Andrade. (2010). "A Review of Rubric Use in Higher Education'. <u>Assessment and Evaluation in Higher Education</u>, 35(4), 435- 448.

Richards, Jack C. (2006). <u>Communicative Language Teaching Today</u>. Cambridge: Cambridge University Press.

Richards, Jack C. and Richard Schmidt. (2002). <u>Longman Dictionary of Language Teaching and Applied Linguistics</u>. London: Pearson Education Limited.

Savignon, Sandra J. (2001). "Communicative Language Teaching for the Twenty First Century". In, Marianne Celce-Mucia (Ed.). <u>Teaching English as a Second or Foreign Language</u>, 3rd ed. New York: Heinle & Heinle Publishers Inc.

Scott, J. (2006). "Authentic Assessment Tools". <u>California Adult Literacy Professional Development Project</u>, 33- 48. Retrieved August, 2011, from <u>http://www.calpro-online.org/eric/docs/custer/custer5.pdf</u>

Stemler, Steven E. (2004). "A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability". <u>Practical Assessment, Research and Evaluation</u>, [Online], 9(4), n.p. Retrieved December, 2012 from <u>http://PAREonline.net/getvn.asp?v=9&n=4</u>.

Stevens, Dannelle D. and Antonia Levi. (2005). <u>Introduction to Rubrics</u>. Virginia: Stylus Publishing.

Stiggins, Richard J. (1992). "High Quality Classroom Assessment: What Does it Really Mean?". <u>ITEMS: Instructional Topics in Educational Measurement</u>, 211- 215.

Stiggins, Rick and Jan Chappuis. (2005). "Using Student-Involved Classroom Assessment to Close Achievement Gaps". <u>Theory into Practice</u>, 44 (1), 11-18. Retrieved March, 2012 from <u>http://www.jstore.org.tiger.sempertool.dk/stable/3496986</u>

Topping, Keith J. (1998). "Peer Assessment Between Students in College and University". <u>Review of Educational Research</u>, 63 (3), 249-276.

_____ (2012). "Peers as a Source of Formative and Summative Assessment". In, J. H. McMillan (Ed.), <u>Sage Handbook of Research on Classroom Assessment</u>. London: Sage Publications, Inc. (Forthcoming Book. The Chapter is gained by direct contact with the author via email in Dec. 2011)

Ur, Penny. (1996). <u>A Course in Language Teaching: Practice and Theory</u>. Cambridge: Cambridge University Press.

Vu, Thu Thuy & Gloria Dall' Alba. (2008). "Exploring an Authentic Approach to Assessment for Enhancing Student Learning". [A

Refereed paper presented in a symposium entitled 'Re-imagining Higher Education Pedagogies' AARE conference, Brisbane], pp.1-11.

Wheater, C. Philip, A. Mark Langan, and Peter J. Dunleavy. (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.

APPENDICIES

Appendix I: Questionnaire of the Teachers of Conversation

Letter of questionnaire

Dear Jury member..

Kindly check the attached questionnaire for its validity and consistency.

The questionnaire is to be introduced to teachers of conversation in the Department of English (in the College of Education and the College of Arts) and the Department of Translation (in the College of Arts).

The questionnaire aims at sighting teachers' methods of assessment they use in assessing students' speaking skill in conversation classes. What kind of criteria they focus upon in their assessments? What type of teaching method they follow? Their perspective about students' performance in the speaking skill, and several other issues related to the study intended.

Please check the box in front of each item, whether approved or disapproved. Then kindly fill in the following form.

I have checked the attached questionnaire and approved using it in the study of the MA candidate Zainab Jaafar Auda, entitled "**The impact of using scoring rubrics in peer assessment on promoting Iraqi EFL learners' speaking skill at the university level**".

<u>Jury member information</u>

Name:

Academic degree:

Teaching experience: (   ) years

Signature:

Date:

Cont.

Teachers Questionnaire

Dear Teacher:

Please, kindly fill in this questionnaire by ticking the space you select for each item. Your cooperation in this regard will be highly appreciated.

Thank you in advance.

Basic information about thesis topic :

In conversation classes, teachers differ in the ways of assessing students' speaking skill. Sometimes they differ even in assessing the same student's performance. This questionnaire is made particularly to gather teachers' opinions about the assessment of EFL learners' speaking skill at the University of Basrah, and as a discovery to the different ways of assessment they use in conversation classes.

It is part of an MA study for assessing the impact of alternative assessment on promoting students' speaking skill.

Personal Background:

1. Qualification: a) M.A ☐   b) PhD ☐

2. Teaching experience (in years):
   a) 1-5 ☐   b) 6-10 ☐   c) 11-15 ☐   d) 16-20 ☐   e) 21+ ☐

3. Experience in teaching conversation (in years):
   a) 1-5 ☐   b) 6-10 ☐   c) 11-15 ☐   d) 16-20 ☐   e) 21+ ☐

4. I teach conversation classes to:

a) First stage ☐   b) Second stage ☐   c) Third stage ☐   d) Two stages ☐   e) Three stages ☐

No.	Questionnaire Item
1.	***What are the objectives of the conversation course you teach?*** *(please tick the one/s suitable to you)*
	a) To teach students the principles of English language. ☐
	b) Help students understand the various relations among ideas. ☐
	c) Help students understand the synonyms of words, phrases and meanings. ☐

Cont.

	d) Help students understand the functions of words in sentences as far as meanings and conditions are concerned. ☐
	e) Help them understand some of the cultural, sociological, political and economical effects in language and literature. ☐
	f) Help them understand the harmony between ideas and their arrangements, logic and classifications. ☐
	g) Help them apply the new ideas they have acquired. ☐
	h) Help them explain and interpret what they read or hear and the like. ☐
	i) Help them choose the suitable meaning from a context. ☐
	j) All the above. ☐
	k) None of the above. ☐
	l) Others. Mention them please: a. b. c. *d.* *e.*
2.	*If you teach conversation classes to more than one stage, do you consider the stage-level variable in your assessment of students' performance?* *a) Yes ☐ b) No*
3.	*If you give conversation classes to more than one stage, do your objectives differ from one stage to another?* *a) Yes ☐ b) No ☐*
4.	*If you chose (yes) from the previous point , would you please list the item/s letter/s of the objective/s mentioned in no.1 that you decide for each stage?* *1st Stage:* _____ . *2nd stage:* _____ . *3rd stage:* _____ .
5.	*What are the criteria you consider in assessing the students' speaking skill?* a) Pronunciation ☐ b) Vocabulary ☐

Cont.

<table>
<tr><td></td><td>
c) Fluency ☐

d) Comprehension ☐

e) Grammar ☐

f) Interaction ☐

g) All the above ☐

h) General impression ☐

i) Others. Mention them please …………………………………………………………………………………………………
</td></tr>
<tr><td>6.</td><td>
Please, list the criteria you mentioned in item no.5 according to their importance in the following grid (from the most to the least)

<table>
<tr><td>1.</td><td>5.</td><td>9.</td></tr>
<tr><td>2.</td><td>6.</td><td>10.</td></tr>
<tr><td>3.</td><td>7.</td><td>11.</td></tr>
<tr><td>4.</td><td>8.</td><td>12.</td></tr>
</table>
</td></tr>
<tr><td>7.</td><td>
Is your assessment of the students' performance

a) daily ☐   b) weekly ☐   c) monthly ☐   d) semesterly
</td></tr>
<tr><td>8.</td><td>
What type of assessment do you follow in your class?

a) *Criterion-referenced assessment (i.e. assessing students against certain standards)*   ☐

b) *Norm-referenced assessment (i.e. assessing students by comparing them to their classmates)* ☐
</td></tr>
<tr><td>9.</td><td>
Do you make use of any text books?   ☐ Yes   ☐ No
</td></tr>
<tr><td>10.</td><td>
How do you choose the topics of the lectures? (you may choose more than one option)

a) I choose them myself.   ☐

b) I ask the students what do they like to talk about.   ☐

Cont. ⌐ the course syllabus.   ☐

d) I choose those which match the course objectives. ☐

e) I have another way. What is it please?   ☐`
</td></tr>
</table>

11.	*Is your class* ☐ student-centered, ☐ teacher-centered, or ☐ teacher- student - centered?
12.	*How do you describe the majority of students in conversation classes?* a) active learners ☐ b) passive learners ☐ c) neither active nor passive ☐
13.	*Which approach/method do you follow in the course?* a) *Communicative approach* ☐ b) *Grammar- translation method* ☐ c) *Audio- lingual method* ☐ d) *Direct approach* ☐ e) *Eclectic approach* ☐ f) *All the above* ☐ g) *Another, mention it please. …………………………………………………….*
14.	*What do you do concerning the students' mistakes?* a) *I correct them* ☐ b) *I over correct them* ☐ c) *I do not correct them* ☐
15.	*In your academic studies, have you studied assessment as one topic in the syllabus?* a) *Yes* ☐ b) No ☐
16.	*a) Are you interested in teaching conversation?* ☐ yes ☐ No ☐ Not much *b) If No, why please?*
17.	*a) Do you feel your students are interested in the conversation lessons?* ☐ yes ☐ No ☐ Not much *b) If NO, why do you think so?*

Appendix II: Academic Information of the Jury of the Teachers' Questionnaire

No	Name	Qualification	Position
1.	*Majeed Hameed Jasim*	*Professor*	*Faculty member in the Dept. of English, College of Education for Humanities, University of Basrah*
2.	*Alaa Hussein Oda*	*Professor*	*Faculty member in the Dept. of English, College of Education for Humanities, University of Basrah*
3.	*Faisal Abid Minshid*	*Professor*	*Head of the Continuous Learning Center, University of Basrah*
4.	*Jameel Qasim Hameed*	*Asst. Prof*	*Faculty member in the Dept. of English, College of Education for Humanities, University of Basrah*
5.	*Amal Abdul-Razzaq Al-Mansoori*	*Asst. Prof*	*Dean of the College of Education for Girls, University of Basrah*
6.	*Zainab Hayyawi Bedewi*	*Asst. Prof*	*Faculty member in the Psychology Dept., College of Education for Humanities, University of Basrah*
7.	*Intisar Adnan AbdulQadir*	*Lecturer*	*Faculty member in the dept. of English, College of Education for Humanities, University of Basrah*
8.	*Rif'at Abdullah Jassim*	*Lecturer*	*Head of the Psychology Dept., College of Education for Humanities, University of Basrah*

Appendix III: The Rubric of the Pilot Study

Speaking Rubric

My name is: stage: group: Date:

I am assessing: Assessment Grade:

Criteria	Superior	Advanced	Intermediate	Novice
Fluency	☐ speaks continuously with almost no pauses or hesitations usually filled with err, umm, you know, etc	☐ speaks fluently with hesitations and pauses to search for appropriate words, with some uses of fillers	☐ frequent use of pauses to organize thoughts and/or for lack of vocabulary in the target language	☐ frequent long pauses and silences with incomplete thoughts and/or loss of subject track
Grammar	☐ no or almost no grammatical errors	☐ some minor grammatical errors that do not obscure meaning	☐ many grammatical errors that do not obscure meaning	☐ major grammatical errors that obscure meaning
Pronunciation & voice	☐ no or almost no pronunciation errors ☐ intonation and rhythm used appropriately ☐ voice is loud enough to reach and attract the attention of the whole group	☐ clear pronunciation with some minor errors that do not affect communication ☐ use of intonation and rhythm but with some errors ☐ voice is heard and reaches the whole group but does not attract their attention	☐ an increase in pronunciation errors that affect communication ☐ use of intonation and rhythm inappropriately ☐ voice is heard but does not reach or attract the distant in the group	☐ major pronunciation errors that affect comprehension ☐ use of mother tongue intonation and rhythm ☐ voice is quite low and does not attract the attention of near colleagues
Vocabulary	☐ use of a rich vocabulary and idioms approximate that of a native speaker ☐ no use of native language	☐ use of a rich vocabulary with inadequate use of idioms and fixed expressions ☐ rare use of native language	☐ rephrase ideas with limited and repetitive vocabulary with no use of idioms & fixed expressions ☐ use of native language for about half his/her speech	☐ misuse of words because of lexical inadequacies ☐ frequent use of words and phrases from the native language
comprehension	☐ fully understands what is said and gives appropriate responses	☐ understands most of what is said and responds appropriately	☐ understands some of what is said and responds appropriately	☐ does not understand what is said. Needs paraphrasing or rewarding.

Appendix IV: Handouts Used in the Pilot Study

Students Handout

Topic: *Advertising*	Lesson:	Date:	Class:

JOBS /Men & Women

Fashion model: good looking, 18- 25 old. Ph. 652-454-5687

Seamstress: Experienced ladies dresses. Paid holidays, vacation, medical insurance.

Apply in person, 2-5pm.

Elegant Dress Shop, 1827, Main St.

Driver: delivery of valuable

Handy Vocabulary

To advertise	Advertising	Salary	Ph., phone, mobile ph.
To qualify, To have the qualifications	Advertisement ad	Product	P. O. Box, post office box
To be worthwhile	Want ad, classified ad	Radio announcer	St., Street
to pay	Sign	Script	Effective
To translate	Billboard	Television announcer	Full page
To buy	Highway	Television set	True/ false
To sell	Hotel, restaurant	Commercial (announcement)	Creative
To offer	Truck	Bus	Stunning
To invent	Newspaper	Slogan	Distracting
To publish	Magazine	Way	Model
To distribute	Position, job	Location	Career

Analysis of the drawing
1. Why is/isn't the highway a good place for billboard like the one in the picture?
2. What do you think the advertisement of the women face is about?
3. How many wants ads do you see? Describe them.
4. For which of the positions in the want ads would/ wouldn't you be qualified? Why?
5. What do you suppose the radio announcer is announcing for?

Points of interest
6. Name some ways of advertising a product.
7. What kind of products are advertised frequently on radio or television?
8. Describe a product you would like to advertise on radio or television and the sort of ad you would use.
9. Why does it seem to be more difficult to be a television announcer than a radio announcer?
10. Why are busses good places for advertisements?
11. What is a want ad?
12. One way to advertise something is to use a slogan. Translate one typical slogan used in your mother tongue.
13. Invent a want ad describing a position for which you would be qualified.
14. "It pays to advertise". Explain this opinion.
15. How it is known when or when not an ad is not effective?

Topic: *Gossip & behaviours*	Lesson:	Date:	Class:

Handy Vocabulary

To gossip	Mate	Decent	Reputation
To hinder	Rumour	Social bonding	Coercive
To participate	Private affairs	Misinformation	Idle talk
To tattletale	Flatter , pay tribute to	Elite	trivial
To affiliate	Newsmonger	Tattler/ chitchat/ dirt	Intentional
To ostracize	Lack of morality	Big mouth	Malicious
To inform	Dignity	Scandalmonger	Negative consequences
To Jeopardize	Aggression	Gossipmonger	Waste time
To criticize	An attack	Self-esteem	Unethical
To perpetuate	Moral	Social status, prestige	Erosion of trust
To sin	Remarks	Anxious	Validation

Analysis of the drawing
1. Is gossip a positive or negative behaviour?
2. With whom is the woman on the top right tattling?
3. What do you think the girl on the phone is talking about? For how long?
4. Can chitchat be hindered? How?
5. Whom do you consider yourself among the three girls in the top left picture?

Points of interest
6. "Gossip is idle talk or rumour, specially about personal and private affairs of others, meant to harm the subject's reputation in the community. It may break existing bonds by ostracizing individuals within a community." Is this true or false? Explain.
7. Gossip can be very serious depending upon the amount of power that the gossiper has over the recipient, which will in turn affect how the gossip is interpreted. Do you agree?
8. What are the main reasons behind gossip?
9. How do private information of celebrities or of elite (like scandals of cinema stars) spread? Are they considered as gossip?
10. What can one do to stop gossip?
11. Can chitchat play a role in building social bonds out of helping individuals to learn information about others in the community?
12. Why does Islam consider backbiting the equivalent of eating the flesh of one's dead brother? What do you think this analogy means?
13. 'Gossip is mainly a female behaviour". Do you agree? Why?

Students Handout

Topic: *Sports*	Lesson:	Date:	Class:

Handy Vocabulary

To play	Sportsmanship	Stadium	Basketball	Badminton
To win	Athlete	Line	Tape	Volleyball
To lose	Fan, spectator	Singles	Runner	Swimming
To score	Contest, game	Doubles	Baseball	Skating
To throw	Match, meet	Tennis	Football	Skiing
To kick	Team	Racket	Bat	Hockey
To hit	Referee	Net	Boxing	Champion
To tie	Rule	Soccer	Heavyweight	Amateur
To run, to jog	Point	Goal	Wrestling	Professional
To exercise	Field	goalkeeper	Golf	Exciting

Analysis of the drawing

1. Name the sports in the drawing.
2. Are the people in the Ping-Pong scene playing singles or double? How do you know?
3. How do the basketball players score points?
   Points of interest
4. What is an athlete?
5. How many players are necessary for a basketball game?
6. Name two sports that are played on a court and two that are played on a field.
7. What is a stadium?
8. Explain the difference between an amateur athlete and a professional athlete.
9. Tennis and badminton are similar. What is one difference between them?
10. In what sport does one kick the ball, and in what sports does one hit the ball with an object?
11. In your opinion what are the most popular sports in the world?
12. Describe an exciting game or match that you have watched.
13. What sports do you play? If you do not play sports, how do you exercise.
14. "Winning isn't everything." Do you agree if you are an athlete? Explain.
15. Which sports are good for children to learn? Why?
16. Which sport is the most dangerous? Why?

Students Handout

Topic: *Describing Peoples -Character*	Lesson:	Date:	Class:

1. *Intellectual ability:*
    - **Ability: intelligent bright clever smart shrewd able gifted talented**
    - **Lacking ability: stupid foolish half-witted simple silly**
    - **Clever, in a negative way, using brains to trick or deceive: cunning crafty sly**

2. *Attitudes towards life:*
    - **Optimistic or pessimistic : Looking on either the bright or the black side of things.**
    - **Extroverted or introverted: Outward-looking or inward-looking (i.e. to the world around one or to one's own inner world).**
    - **Relaxed or tense: Calm or not calm with regard to attitude to life.**
    - **Sensible / down-to-earth : Practical, not dreamy in approach to life.**
    - **Sensitive : Feeling things very intensively.**

3. *Attitude towards other people*
    - **Sociable or gregarious : Enjoying other's company.**
    - **Quarrelling / argumentative : Disagreeing with others.**
    - **Cruel / Sadistic: Taking pleasure in others' pain.**
    - **Easy-going / even-tempered: Relaxed in attitude to self and others.**
    - **Impolite rude ill-tempered: Not polite to others.**
    - **Honest / Trustworthy / Reliable / Sincere: Telling the truth to others.**
    - **Jealous / envious: Unhappy if others have what s/he doesn't have.**

4. *Magazines often publish questionnaires which are supposed to analyse your character for you. Look at the words below and then match them to the statements which aim to decide a person's character.*
    Example:
    If you arrange to meet at 7 p.m. and you arrive at 7 p.m., then you are: Reliable
                            (argumentative sensitive sociable   inquisitive)

    1. **You prefer to be in the company of other people?   --------------**
    2. **You frequently disagree with what other people say. --------------**
    3. **You lie awake at night if someone has said something unkind to you. --------------**
    4. **You always look out of the window if you hear a car draw up. --------------**

5. **Choose five characters that best describe your own and your best friend's character. Say how you demonstrate these characteristics.**
                    Example: I am sociable because I love being with other people.

6. **Choose five characters that best describe a person you do not like. (Without mentioning his/her name)**

Students Handout

Topic: *Shopping*	Stage:	Group:	Date:

Handy Vocabulary

To buy/ sale	shop assistant	product labels	Online shopping	recreational
To return	Cashier	Fit	discounted goods	leisure
To exchange	Window shopping	credit card	sold out	entertainment
To purchase	receipt	Shopping list	refund	vexation
To try on	refund	shopping carts	Clothes	inconvenience
To select	Goods	peddlers	Personal care	Shopping frenzy

Analysis of the drawings
1.   What's your favorite department?
2.   Is it difficult for you to choose what to buy? Why?
3.   Are you good at buying gifts?
4.   Why do you try on the things you buy?
5.   "Window shopping" is an activity that shoppers engage in by browsing shops with no intent to purchase, possibly just to pass the time between other activities, or to plan a later purchase.
6.   Do you read the labels on products? Why?
7.   Make a shopping list of the items you mostly go shopping for.
8.   Not all shop assistants accept giving refund for sold items.

Points of interest:

1. To some people, shopping is considered a recreational and leisure activity in which one visits a variety of stores with a premeditated intent to purchase a product. To others, shopping is a task of inconvenience and vexation.

2. Stores are divided into multiple categories of stores which sell a selected set of goods or services.

3. Regulations: Some nations regulate the operation of businesses for religious reasons and do not allow shopping on particular days or dates.

4. Shopping seasons: Shopping frenzies are periods of time where a burst of spending occurs—typically near holidays.

5. Pricing & negotiation: Often, prices are fixed and price discrimination can lead to a bargaining situation.

6. Home Shopping:
  * Home shopping: Home mail delivery systems and modern technology (such as television, telephones, and the Internet) allow consumers to shop from home.

  * Neighborhood shopping: Sometimes peddlers and ice cream trucks pass through neighborhoods offering services and goods

  * Party shopping: The party plan is a method of marketing products by hosting a social event to display and demonstrate the products to those gathered, and then to take orders for the products before the gathering ends.

Students Handout

Topic: *Technology*	Stage:	Group:	Date:
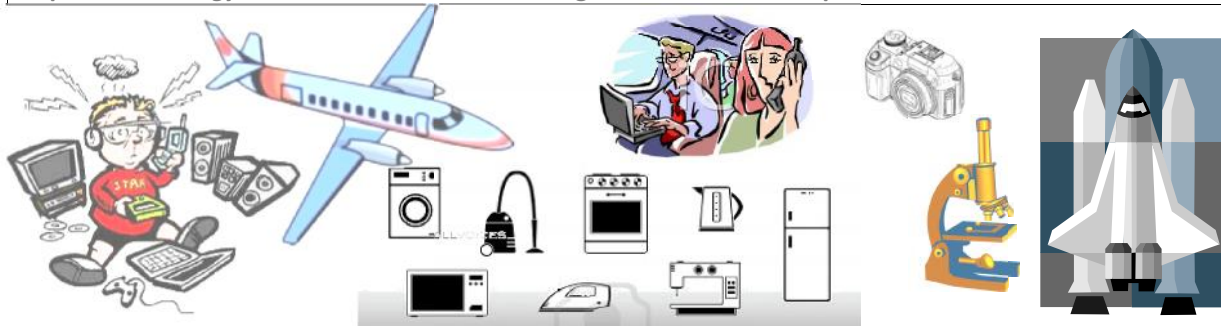
Handy Vocabulary

To browse	innovations	hardware	science fiction	Services
To download	IT	software	Chat	infrastructure
To deliver	achievements	Desktop	e-mail	Virus
To chat	advantages	Laptop	Account	Website
To access	disadvantages	CD	Face to face	social websites
To publish	cell phones	printer	Communication	Violent

Points of interest:

1. Technological innovations affect all aspects of our lives, personal and professional.
2. The children of today are the leaders of tomorrow
3. IT: Information Technology is the use of hardware, software, services, and supporting infrastructure to manage and deliver information.
4. What are some of the greatest technological achievements?
5. In your opinion, what are the advantages and disadvantages of technology?
6. In your opinion, what is the greatest technological invention? Why?
7. What do you think are the three most interesting technologies which have entered into Iraq market since 2003?
8. Do you think using cell phones too much is bad for our physical or mental health? Why?
9. What social changes have cell phones made?
10. How do you think face to face communication differs from communication using computers?
11. What are good and bad points of using computers?
12. What is your opinion about children playing violent video games or computer programs?
13. Have you used the Internet to learn English or read or talk in English?
14. Can you access the internet on your telephone?
    - If so, how often do you use it?
15. How often do you check email?
16. What science fiction movies have you seen?
    - Do you think that what you have seen in these movies is possible?
17. How often do you use the internet?
18. When did you first use the Internet?
19. Who uses the Internet the most in your family?
20. What are the sites you most commonly access?
21. What problems does the Internet create? What problems does it solve?
22. Do you have an e-mail address? What do you use it for?
23. Have you ever chatted on the Internet?
24. Is it dangerous to meet people on the Internet?
25. Do you think that the Internet safe for children? Why?
26. Can you believe all the information that is published (available) on the Internet?
27. Do you think that it is a good or bad habit for young people to play computer games?
28. How does the Internet help people from different countries to communicate with each other?
29. Do you have a Facebook account?
30. How many friends do you have on Facebook?
31. What do you do on Facebook?

Appendix V: The Rubric of the Pilot Study with Scoring

[Speaking Rubric for Conversation Classes]   stage: group: Date:

My name is:

I am assessing:

Criteria	Superior	Advanced	Intermediate	Novice	Topic
Fluency	☐ **continuous** speech with almost **no pauses** or hesitations	☐ **continuous** speech with **some pauses** to search for adequate words	☐ **frequent pauses** to organize thoughts and/or for lack of vocabulary	☐ **frequent long pauses** with incomplete thoughts	
Grammar	☐ **no** or almost no **errors**	☐ **some minor** errors that **do not** obscure meaning	☐ **many** errors that **do not** obscure meaning	☐ **major errors** that **obscure** meaning	
	☐ **no** or almost no **errors**	☐ **some minor errors** that **do not** affect communication	☐ **many errors** that **do not** affect communication	☐ **major errors** that **affect** comprehension	
Pronunciation & voice	☐ **appropriate** use of **intonation** and rhythm	☐ **some errors** in using **intonation** and rhythm	☐ **inappropriate use** of **intonation** and rhythm	☐ use of **mother tongue intonation** and rhythm	
	☐ **loud & attractive** voice addressed to whole group	☐ **loud** voice but **not attractive**	☐ **low & unattractive** to whole group	☐ **quite low & unattractive** voice **to near** colleagues	
Vocabulary	☐ **rich** use of **handout vocabulary**	☐ **some** use of **handout** vocabulary	☐ **rare** use of handout vocabulary	☐ **misuse** of words for lack of vocabulary	
	☐ **no** use of **native** language	☐ **rare** use of **native** language	☐ use of **native** language for about **half** the speech	☐ **frequent use** of words and phrases from the **native** language	
Comprehension	☐ **fully understands** what is said and gives **appropriate responses**	☐ **understands most** of what is said and **responds** appropriately	☐ **understands some** of what is said and **responds** appropriately	☐ **doesn't understand** what is said. **Needs paraphrasing** or rewarding.	

Scoring Information

Criteria	Superior	Advanced	Intermediate	Novice
Fluency	☐ 4	☐ 3	☐ 2	☐ 1
Grammar	☐ 2	☐ 1.5	☐ 1	☐ 0.5
Pronunciation & voice	☐ 2	☐ 1.5	☐ 1	☐ 0.5
	☐ 2	☐ 1.5	☐ 1	☐ 0.5
Vocabulary	☐ 2	☐ 1.5	☐ 1	☐ 0.5
	☐ 4	☐ 3	☐ 2	☐ 1
comprehension	☐ 2	☐ 1.5	☐ 1	☐ 0.5

Total:

Appendix VI: the Scoring Rubric of the Study

[Speaking Rubric for Conversation Classes]

My name is: _____   I am assessing: _____   Discussion leader: _____

group: _____   Date: _____   Topic _____

Criteria	Superior	Advanced	Intermediate	Novice
Fluency	**continuous, clear** and **well organized** speech	**continuous** speech with **some pauses** to search for suitable words and/or organize thoughts	**frequent pauses** to organize thoughts and/or for lack of vocabulary	**frequent long pauses** with incomplete utterances
Grammar	**no errors**	**some minor** errors that **do not** obscure meaning	**many** errors that **do not** obscure meaning	**major errors that obscure** meaning
Pronunciation	**no errors**	**some minor errors** that **do not** affect communication	**many errors** that **do not** affect communication	**major errors** that **affect** comprehension
	no use of Arabic	**hardly** uses **Arabic** words	almost **half** the speech is in Arabic	**most** of the speech is in Arabic
Vocabulary	**Rich** use of **vocabulary** (including fixed expressions and phrases)	**often adequate** and accurate use of **vocabulary**	**repetition** of words because of **limited** adequate and accurate vocabulary	**Inadequate** and **inaccurate** use of vocabulary
Interaction	Raises questions, responds, comments, helps to clarify others' ideas, and **encourages others to participate in discussion**	only raises questions and responds to others	Only responds to questions if asked by name	**no attempt to participate** in discussion
Content	**Interesting details** related to the topic including facts, evidences, personal stories,… / uses **interesting humor**	questions and responses related to the topic with **no details** / **restricted** use of **humor**	**Circumlocutes** around the topic with **no valuable content** / focuses on **humor** in each statement	**very short answers** (yes/no or 2-5 words responses) / **humor dominates** speech

Criteria	Superior	Advanced	Intermediate	Novice
Discussion leader	Starts, develops, maintains and closes discussion affectively within time limit / Engages **all members in** discussion (by asking questions, requesting for clarifications, commenting, clarifying misunderstandings, correcting others' utterances, etc)	Starts, develops, and maintains discussion with **no closing** within time limit / Engages all members in discussion by asking questions only	Starts discussion but **loses control** over discussion & time / Engages specific members by asking questions **and/or** dominates discussion him/herself	**unable** to start or control discussion / **unable** to engage others in discussion

Appendix VII: 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.

Appendix VIII: Students' Questionnaire Concerning their Perspectives of the Application of the Alternative Assessment Practices, Namely: the Rubric, the Peer Assessment, and the Group Work

Students' Questionnaire

Dear students,

The aim behind this questionnaire is to discover your perspectives of using the instructional rubric, the peer assessment, and group discussions that have been applied on your conversation classes in the second semester. The data will be used for research purposes, so your actual responses will be highly appreciated. Thank you in advance.

How to answer

For each statement, you may choose any of the three options, agree, disagree, and neutral. Agree means 'yes', disagree means 'no' and neutral means 'in between'.

Note: PA refers to peer assessment.

No.	Item	Agree	Neutral	Disagree
1.	Criteria descriptors on the rubric were easy to understand			
2.	It was difficult to decide the level of my peer for all the criteria			
3.	I knew my strengths and weaknesses in speaking through using the rubric			
4.	I became more active in conversation classes after using the rubric in our daily assessment			
5.	The rubric told me what's expected of me to achieve the best performance in speaking skill			
6.	PA allowed me to gain experience in giving and receiving feedback			
7.	PA supports me with detailed and timely feedback to my speaking skill			
8.	Relationships with my peer (friendships, etc.) may have influenced overall assessment I give to him/her.			
9.	I was comfortable being a judge and assessing my peer's speaking skill.			
10.	I was comfortable having my speaking skill judged and assessed by my peers.			
11.	The overall assessments my peers gave me were fair and reasonable.			
12.	Assessing other students' speaking skill helped me plan to improve my own			
13.	PA comments from the first use made me aware of my weaknesses and strengths			

14.	PA guarantees we speak in our best performance even with the absence of the teacher's observation (like participating in each discussion, not turning the discussion into laughter, avoiding speaking in Arabic, etc)			
15.	Leading my group discussion gave me more self-confidence to act as a teacher			
16.	I felt more comfortable in discussing the topics with my classmates through group discussions than merely with the teacher			
17.	Group discussion encouraged me to participate in each lecture's discussion/ activity			
18.	I recommend using the instructional rubric in conversation classes.			
19.	Students should **not** be involved in assessing peers; assessment should be solely the teachers' job.			
20.	I recommend using the group discussion technique in conversation classes			

If there are any comments you'd like to add concerning your experience in the use of the rubric, the peer assessment, and the group discussions, please write them below.

Your Comments

1.

2.

3.

4.

5.

Appendix IX: Academic Qualification of the Jury of the Reliability Test

No	Name	Qualification	Workplace
1.	*Saad Chasib Daghir*	*Lecturer*	*Department of English Language, College of Education for Humanities, University of Basrah*
2.	*Abdul-Razzaq Darweesh Abdul-Razzaq*	*Lecturer*	
3.	*Amin Ukaal Ghailan*	*Lecturer*	
4.	*Mahdi Mohsin Mohammed*	*Lecturer*	
5.	*Neda Salih AbdulRidha*	*Lecturer*	
6.	*Ibtisam Hussein Al-Abudi*	*Lecturer*	
7.	*Na'il Sadiq Wakir*	*Assist Lecturer*	
8.	*Wafaa Shakir Ibrahiem*	*Assist Lecturer*	
9.	*Rasha Ali Sehu*	*Assist Lecturer*	
10.	*Nawress Sabah Al-Mudhafar*	*Assist Lecturer*	

من أجل التحقق من فرضيات الدراسة، أتبَعت الباحثة أولاً طريقة الاختبار ما قبل الدراسة و ما بعدها حيث أُستخدمت كلاً من التطبيقات الثلاثة للتقييم البديل (أي الجدول التقييمي "Rubric" و تطبيق تقييم النظير "Peer Assessment" و تقنية العمل الجماعي "Group Work") فيما بين الاختبارين على عينة مناسبة من الطلبة (أربع و سبعون طالباً و طالبة) من متعلمي اللغة الانجليزية في قسم اللغة الانجليزية في كلية التربية للعلوم الإنسانية، جامعة البصرة. ومن أجل التحقق من النتائج أُستُخْدِمَ الاختبار "ت" حيث أظهر تأثيرات إيجابية لتطبيقي الجدول التقييمي "Rubric" و تقييم النظير "Peer Assessment" على إنجاز عينة الطلبة. ثانياً، ساعدت التطبيقات البديلة على خلق صف تفاعلي ذي فعالية عالية محوره الطلبة. ثالثاً، بعد إختبار الجدول التقييمي "Rubric" المُعَد للدراسة بمساعدة عشرة مقيّمين في القسم المذكور سابقاً لتقييم مهارة التكلم لدى عينة من الطلبة (ستة طلبة) و بإستخدام تحليل كرونباك للمصداقية، تبين أنه حقق مصداقية عالية تبلغ 0,934. و الخلاصة؛ تبعاً لنتائج الدراسة المذكورة فإن طريقة الدراسة كانت ناجحة في تحقيق أهدافها .

الخلاصـــة

شهد مجال القياس و التقويم منتصف التسعينيات تحولاً من التقييم التقليدي إلى التقييم البديل. و نتج عن هذا التحول العديد من التطبيقات الجديدة التي تهدف إلى ربط مجال التقييم بمجالي التعلُّم و التعليم من أجل الحصول على أفضل فوائد التقييم. و من ضمن هذه التطبيقات، وُجِدَ بأنَ تطبيق الجدول التقييمي "Rubric" و تطبيق تقييم النظير "Peer Assessment" لهما تأثير إيجابي على كل من المتعلمين و المدرسين و المناهج. فضلاً عن ذلك، وُجِدَ بأنَّ الجدول التقييمي "Rubric" أكثر فعالية في تقييم المهارات التي عادة ما تُقيَّم بصورة شخصية مثل مهارة التكلم. لذا فإنَّ الهدف الرئيس من وراء هذه الدراسة هو التحقق من تأثير إستخدام تلك الجداول التقييمية في تطوير مهارة التكلم لدى متعلمي اللغة الانجليزية في دروس المحادثة.

بعد تطبيق الجدول التقييمي "Rubric" في الدراسة كوسيلة تقييم مجردة، لم يحصل تأثير حقيقي في أداء الطلبة. لِذا أُستُخْدِمَ الجدول التقييمي "Rubric" في تقييم النظير "Peer Assessment" داخل الصفوف. و بناءً على ذلك، تفترض الدراسة عدداً من النقاط. أولاً، إن لإستخدام الجداول التقييمية "Rubric" في تطبيق تقييم النظير "Peer Assessment" في دروس المحادثة تأثيراً إيجابياً في تطوير مهارة التكلم لدى الطلبة داخل الصف. ثانياً، عند تنفيذ التطبيقين مع تقنية العمل الجماعي "Group Work" سيكونان فعالين في خلق صف تفاعلي محوره الطلبة. ثالثاً، يمكن تحقيق مصداقية عالية بين مقيّمين مختلفين من خلال إسنخدام جدول تقييمي "Rubric" ذي صياغة حسنة.

جامعة البصرة
كلية التربية للعلوم الإنسانية
قسم اللغة الانجليزية

تأثير إستحدام الجداول التقييمية "Rubrics"

في تقييم النظير "Peer Assessment"

لتطوير مهارة الكلام لدى طلبة الجامعة العراقيين

من متعلمي اللغة الانجليزية لغة أجنبية

رسالة مقدمة

إلى مجلس كلية التربية للعلوم الإنسانية ـ جامعة ألبصرة

و هي جزء من متطلبات نيل شهادة الماجستير آداب في اللغة الانجليزية

نقدمت بها ألباحثة

زينب جعفر عودة

بإشراف

أ. د. بلقيس عيسى كاطع الراشد

1434 هـ   2013 م