

American University of Armenia

Department of English Programs

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree
of Masters of Arts in Teaching English as a Foreign Language (TEFL)

The Impact of Analytic Rubrics on the Teaching and Assessment of EFL Learners` Speaking Skills

Thesis committee:

Thesis Supervisor: Irshat Madyarov, Ph.D.

Thesis Reader: Rubina Gasparyan, MA

Statistics Consultant: Alexan Simonyan, Ph.D

by:

Luiza Grigoryan

Yerevan, Armenia

2013

American University of Armenia

We hereby approve that this thesis

by

Luiza Grigoryan

The Impact of Analytic Rubrics on the Teaching and Assessment of EFL Learners’ Speaking Skills

be accepted in partial fulfillment for the requirements of the degree

of

M.A. in TEFL

Thesis committee:

Thesis Supervisor: Irshat Madyarov, Ph.D

Thesis Reader: Rubina Gasparyan, MA

Associate Dean: Catherine Buon, Ph.D

Yerevan, Armenia

2013

Acknowledgements

I wish to express my gratitude to people who helped me to complete my thesis. I want to acknowledge my deep gratitude to Dr. Irshat Madyarov, my thesis supervisor, for his detailed feedback and thoughtful guidance to overcome difficulties and to complete my thesis.

I would like to express my gratitude to Mrs. Rubina Gasparyan, my thesis reader, for her suggestions and insightful comments that helped me to improve my work.

I am also very grateful to Dr. Alexan Simonyan, statistics consultant, who assisted me a lot with statistical analyses of my study.

My gratitude is also great to Dr. Catherine Buon, Associated Dean of the Department of English Programs as well as all instructors of CTEFL and MATEFL programs for theoretical and practical knowledge they gave us.

Table of Contents

List of Tables	vii
Abstract.....	viii
Chapter 1: Introduction	1
1.1 Background of the Study	1
1.2 Statement of the Problem.....	3
1.3 Purpose of the Study	4
1.4 Significance of the Study	4
Chapter 2: Review of Literature.....	6
2.1 Nature of Speaking Skill.....	6
2.2 Background to Teaching Speaking.....	8
2.3 Issues in Teaching Speaking in an EFL context	9
2.4 What is Assessment?	11
2.5 Types of Assessment	12
2.6 Purpose of Assessment	13
2.7 Assessment in Speaking.....	14
2.8 Background and Definitions of Rubrics.....	17
2.9 Types and Main Features of Rubrics	18
2.10 Viewpoints about Integrating Scoring Rubrics in Education	18
2.11 Summary of Related Literature	26
Chapter 3: Methodology.....	28
3.1 Research Design.....	28
3.2 The Setting of the Study	28
3.3 The Participants of the Study.....	28

3.4 Materials.....	29
3.5 Instrumentation and Procedure of the Study.....	29
3.5.1 Speaking proficiency pre and post tests.....	30
3.5.1.1 Pilot test.....	30
3.5.1.2 Pre and post tests reliability.....	31
3.5.1.3 Pre and post tests procedure.....	32
3.5.1.4 Assessment of pre and post tests.....	33
3.5.2 Speaking assignments during the course.....	34
3.5.2.1 Assessment of speaking assignments	35
3.5.3 Questionnaire.....	36
Chapter 4: Results of the Study.....	37.
4.1 Analysis of the Data for the First Research question	37
4.2 Analysis of the Data for the Second Research Question.....	38
4.2.1 Mann-Whitney U tests for between group comparisons.....	38
4.2.2 Wilcoxon Signed-Ranks tests within group comparisons	39
4.2.2.1 The effect size for Wilcoxon Signed- Ranks test.....	40
4. 3 Analysis of the Data for the Third Research Question.....	40
4.4 Analysis of the Data for the Fourth Research Question.....	41
4.4.1 Analysis of the quantitative data	41
4.4.2 Analysis of the qualitative data.....	44
Chapter 5: Discussion and Conclusion.....	46
5.1 Discussions on the Findings Related the Research Questions	47
5.2 Delimitations of the Study	49
5. 3 Limitations of the Study	50

5.4 Suggestions for Further Research.....	51
5.5 Pedagogical implications.....	52
References	53
Appendix A.....	60
Appendix B	62
Appendix C	64
Appendix D.....	66
Appendix E	67
Appendix F	68

List of Tables

Table1. Reliability Statistics for Pre and Post-Tests and four Assignments.....	37
Table 2 Mann-Whitney U Tests of Pre-Test and Post -Test Scores.....	38
Table 3. Wilcoxon Test for Experimental and Control Groups.....	39
Table 4 Mann-Whitney U Tests for between Group Comparisons.....	41
Table 5 Students' exposure to rubrics	42
Table 6: Students` attitude towards rubrics	43

Abstract

The aim of this research study is to investigate the impact of analytic scoring rubrics on Armenian EFL learners' speaking skills. It also investigates whether the use of rubrics promotes inter-rater reliability in assessment of speaking. Besides, it explores Armenian EFL learners' attitude towards the use of rubrics for teaching and assessment processes.

The study was conducted in the Experimental English Classes (EEC) of the Department of English Programs (DEP) at the American University of Armenia (AUA). The participants of the study were 19 EEC Communication Level 3 students. The study employed a quasi-experimental design and the participants were grouped into two groups: experimental and control. The experimental group received treatment; particularly the students used rubrics for preparing oral assignments and received feedback based on rubrics. The control group did not use rubrics for preparing their assignments and received the traditional method of assessment.

The results of the study revealed that there is no impact of rubrics on the improvement of Armenian EFL learners' speaking skills. Besides, the results showed that the students from the experimental group did not perform better on their speaking assignments than the students from the control group. However, the study displayed high inter-rater reliability of the scores of two raters guided by analytic rubrics to assess students' speaking skills. It also showed the learners' positive attitude towards the use of rubrics during the course.

Chapter 1: Introduction

1.1 Background of the Study

Nowadays, along with the strengthening position of English as a dominant language for communication, the teaching of speaking skill has become significant in the English as a foreign and second language (EFL/ESL) context. The great number of the world's EFL and ESL students learns English to develop their proficiency in speaking. Moreover, they are considered successful if they can communicate effectively in the target language (Riggenback & Lazaraton, 1991 as cited in Widiati & Cahyono, 2006, p.269). According to Ur (2006), the majority of second and foreign language learners, when learning the target language, are concerned about learning to speak, to communicate in that language. Alderson and Bachman (2004) state that the ability to speak and understand the foreign language is the goal of many learners. The mastery of speaking skills in English is a priority for many learners. Learners often evaluate their success in learning as well as the effectiveness of the course based on how much they have improved in their spoken language proficiency (Richards, 2008). Thus, speaking has been considered as an important language skill for second and foreign language learners.

The development of speaking skill in a foreign language context is rather challenging for learners as they do not have an opportunity to practice, use the target language for interaction outside of classrooms. Thus, development of oral English proficiency is the concern of many researchers and educators. According to Widiati and Cahyono (2006, p. 287), speaking is

a central element of communication and it needs special attention in an EFL context. Variety of classroom activities, tasks and teaching materials appear to have been used to deal with the complex nature of speaking.

Based on the review of literature, Widiati and Cahyono (2006, pp. 277-288) discuss recommendations on how to achieve a high quality of teaching speaking in EFL context. It is highlighted that in addition to the types of activities and materials in EFL speaking classrooms, it is also important to be aware of aspects related to assessment procedure of students' performance. Munoz and Alvarez (2009) state that the assessment system is of growing importance in an EFL context. Based on the results of the empirical study, Munoz and Alvarez (2009) conclude that if the students understand the assessment system and scoring scales they perform better and increase their language store.

Various researchers highlight the importance of an analytic approach in speaking classes and the assessment based on the analysis of features of oral production. The analytic approach includes rubrics with evaluation of more specific skills and concepts rated by level of performance (Brown, 2008). The empirical research by Tuan (2012) shows an interconnection between assessment based on an analytic approach and students' improvement in speaking. This study also shows students' positive attitude towards an analytic scoring approach in teaching and assessing speaking skills. According to Mukminatien (2000), the analytic approach is suitable for EFL classroom purposes as a part of teaching and assessment process.

Due to various benefits of scoring rubrics many educators and researchers accept analytic scoring and use of rubrics for assessment. One of the benefits highlighted by Jonsson and Svingby (2007) is that detailed

evaluations of work and the use of analytic rubrics promote the reliable scoring in assessment. However, there are specialists in education that highlight the disadvantages of analytic approach and oppose the use of rubrics for assessment. Kohn (2006) states that detailed and frequent evaluation of students' accomplishments can be counterproductive. He criticizes the use of rubrics for assessment and feedback and opposes the distribution of rubrics to students. Besides, as Wolf and Stevens (2007) state, though the benefits of rubrics are highlighted in many studies, there are also limitations of rubrics. They also find that poorly designed rubrics can diminish the learning process.

1.2 Statement of the Problem

Research has revealed a number of benefits that scoring rubrics bring into learning and teaching processes (Jonsson & Svingby, 2007). However, there are viewpoints that the use of rubrics in education should be reconsidered. Besides, although rubrics have received theoretical support as an important measure for assessing and improving students' learning, there is a dearth of empirical work establishing the same. Reddy (2007) states that "studies on rubrics, a direct measure of students' learning, are far outnumbered by those using indirect measures such as surveys and questionnaires". Besides, the use of scoring rubrics in an EFL context, particularly for developing students' speaking skill, has not been investigated thoroughly yet. Moreover, there are no studies done particularly in an EFL Armenian setting to determine to what extent the use of analytic scoring rubrics may be effective for developing speaking as well as for promoting reliability in the assessment of speaking performance. Besides, almost

nothing is known about Armenian learners' attitude towards the use of rubrics for language learning. Therefore, it seemed reasonable to conduct a research and try to fill these gaps.

1.3 Purpose of the Study

The current study aims at revealing the impact of rubrics on Armenian EFL learners' speaking proficiency. Particularly, it aims to reveal the impact of analytic rubrics on preparing speaking assignments and measure the influence of rubrics on the improvement of EFL learners' speaking skills. Besides, the current study discloses the potential of rubrics to promote reliable assessment of speaking performances. Finally, it aims to show Armenian EFL learners' attitude towards the use of rubrics for teaching and assessment processes.

1.4 Significance of the Study

The significance of the present study stems from the following factors: It attempts to explore the influence of rubrics on the improvement of students' speaking skill viewed as one of the difficult skills to develop and master in EFL settings. In this respect, the results can help to evaluate the effectiveness and impact of rubrics on Armenian EFL students' speaking proficiency. The findings can promote interest among course and curriculum designers and those teachers who want to implement scoring rubrics in EFL classrooms.

Moreover, the results of the study may provide an insight concerning rubrics usage to assure reliability in assessment which is one of the major issues in education.

Finally, the result of the study may show students' perceptions about the use of rubrics in learning process. This information can also be taken into account by course designers and teachers.

Chapter 2: Review of Literature

In this chapter relevant literature on theoretical framework and historical background on speaking skill as well as the usage of scoring rubrics in education are explored. It presents viewpoints of various authors about the nature of speaking and discussions on the use of scoring rubrics in teaching and learning processes. To this effect, the discussions lead to the formulation of research questions presented at the end of the chapter.

2.1 Nature of Speaking Skill

“Spoken language, as has often been pointed out, happens in time, and must therefore be produced and processed ‘on line’. There is no going back and changing or restructuring our word as there is in writing; there is often no time to pause and think, and while we are talking or listening, we can not stand back and view the discourse in spatial or diagrammatic terms...”

(Cook, 1989, p115)

Speaking is a productive oral skill that consists of productive systematic verbal utterances to convey meaning (Nunan, 2003). Celce-Murcia and Olshtain (2000) consider speaking skill as one of the most difficult skills to develop as it requires command of both listening comprehension and speech production sub skills such as vocabulary, choice of grammatical

patterns, and so forth. Moreover, to communicate and to produce the messages on the spot requires various other skills and knowledge, such as selecting vocabulary that is understandable and appropriate for the audience, the topic being discussed, or applying strategies to enhance comprehensibility, such as emphasizing key words, rephrasing, or checking for listeners' comprehension (Brown, 1994 as cited in Florez, 1999).

Widiati and Cahyono (2006, 271) state that in the context of language learning it is commonly believed that to communicate in a written form (writing) is more difficult than orally (speaking). Though the complexity of spoken and written languages differs, it doesn't mean that one is easier than other (Artini, 1998 as cited in Widiati & Cahyono, 2006, p. 271).

Nunan (2003) finds that the difficulty of speaking is embodied in the following reasons: it take place in *real time*, in unpredictable and unplanned situations. Thus, while speaking there are limited opportunities to edit and revise output. Celce-Murcia and Olshtain (2000) state that in spoken and not preplanned interaction decisions and choices need to be made under the pressure of trying to communicate messages.

2.2 Background to Teaching Speaking

The history of language teaching started with the adoption of the approach dominated European and foreign language teaching from the 1840s to the 1940s. Under that approach, known as Grammar Translation Method, a foreign language teaching was focused on detailed analysis of grammar rules, translating sentences and texts into and out of the target language (Richards & Rodgers, 2002, pp. 5-6). That approach required reading and writing to be in the focus of language teaching, and the developing skills to speak a foreign language were regarded as irrelevant (Prator, 1991 as cited in Widiati & Cahyono, 2006, p.275). However, the opposition to the Grammar Translation Method gradually developed and laid the foundation for rethinking of foreign language teaching methodology. In the mid of 1950s the new language teaching method came to be known as Audiolingualism

In Audio-Lingual Method (ALM) speaking was taught by having students repeat expressions and memorize dialogs (Nunan, 2003). According to Brown (2001), this method of teaching was developed based on habit formation model that involve drills and language pattern practices. The ALM was criticized for not providing language learners with spontaneous use of the language. Memorizations and drills have questionable value for developing skills to communicate and share information. Various studies led to conclusion that people learn a language by interacting with each other. As a result, in 1970s many researchers focused on “communicative potential of the target language” (Widdowson, 1972; Brumfit & Johnson, 1979 as cited in Wenjie, 2009. p.19). In the center of language teaching was the development of communicative skill (Richard & Rodgers, 2001). Thus, the method called communicative language teaching arose and current teaching methodology

moved towards more and more communicative activities and tasks. The emergence of communicative language teaching led to changed views of syllabuses and methodology (Richards, 2008). This method requires from foreign language teachers to provide students with an opportunity to talk by using group and pair works, plan speaking tasks that involve negotiation for meaning, and use classrooms techniques and tasks, such as information gap, Jigsaw activates, role-plays, simulations, discussions (Nunan 2003, pp. 55-57).

2.3 Issues in Teaching Speaking in an EFL context

According to Celce-Murcia and Olshtain (2000), the teaching and learning of spoken language in the language classrooms is a very difficult task both for the teachers and the students. In order to communicate in another language, it is usually not enough to reach a good level of linguistic knowledge. In order to be an effective oral communicator in another language, there are prerequisites that should be taken into account as a part of discourse knowledge. Some of such prerequisites are:

- a) knowing the vocabulary relevant to the situation
- b) ability to use discourse connectors such as well; I see; okay
- c) ability to use “opening phrases” and “closing phrases” such as *Excuse me* or *Thank you for your help*, etc.
- d) ability to comprehend and use reduced forms
- e) knowing how to use the interlocutor’s reaction and input
- f) awareness of the various conversational rules that facilitate the flow of talk, etc

The mentioned areas should be taken into account in organizing speaking in EFL classrooms.

Besides, when planning speaking classes for non native language learners it is important to define the goals of the program. Besides, the focus should be on the most important features and principles of teaching speaking (Celce-Murcia & Olshtain, 2000)

According to Nunan (2003), there are core principles that should be addressed for teaching speaking. These principles are described as following:

1. Be aware of the differences between SL and FL learning contexts
2. Give learners practice with both fluency and accuracy
3. Provide opportunity for students to talk by using group work and limiting teacher talk
4. Plan speaking tasks and involve negotiation for meaning
5. Design classroom activities that involve guidance and practice of both traditional and interactional speaking (p.59)

Besides of complicated nature of speaking, the learners' different personalities can also bring difficulties for developing speaking skills. For example, there are students who are shy and their lack of self competence make them keep silence in the class (Tutyandari, 2005; Celce-Murcia, 2001). According to Munjayanah (2004), cited in Sari (2008), learners often shy to use the target language in classroom as they worry about mistakes.

Another issue that makes the development of the target language difficult is the use of mother tongue. It is easier for students to use their mother tongue in their classes (Sari, 2008, p.17). Thus, one of the challenges many second language teachers face is motivating their students to speak in the target language.

Due to its complicated nature as well as the different personalities of non native language learners that bring additional challenges in the learning process, various educators consider speaking as the most challenging skill. The review of literature on teaching speaking propounds a wide techniques and activities to improve learners' speaking production. In addition to the types of activities and materials in EFL speaking classrooms, it is also important to be aware of aspects related to the assessment procedure of students' performances. Munoz and Alvarez (2009) state that the assessment has a significant role in education and assessment system is of growing importance in EFL context as well.

2.4 What is Assessment?

“No assessment device is good or bad in itself but only in context. Only when we know what we are seeking to discover can we claim that a particular kind of assessment is appropriate or not.”

(White, 1995, p. 35)

To many teachers as well as students the word “assessment” means giving students grades. McAlpine (2002, p.5) describes grades as a concise summary of the students' ability. However, assessment is much more than just providing grades. Brissenden and Slater (n.d) define assessment as a mechanism for providing instructors with data for improving their teaching methods and for guiding and motivating students to be actively involved in their own learning.

2.5 Types of Assessment

There are various types of assessment, such as formative or summative, formal or informal, final or continuous, process or product. Each type of assessment has its purpose and should be used appropriately. McAlpine (2002) describes different types of assessment as follows:

Formative vs. Summative Assessment

Formative assessment is designed to assist the learning process; particularly it provides feedback to the learners which can help to improve their performance. It can improve students' overall performance by giving them more detailed information on their abilities and concentrating students' efforts on the more appropriate areas. Thus, the results of the formative assessment are used internally, i.e., by students, teachers, etc. Besides, formative assessment is conducted continuously during the course or program.

Summative assessment is for progression and external purposes, given at the end of the course and designed to judge the students overall performance. Thus, it gives a concise summary of students' ability, such as a *pass* or *fail* or a grade (*A-E*, etc.) Summative assessment is most appropriate to show students' ability to external interested parties.

Formal vs. Informal Assessment

In formal assessments students are aware that their performance is for assessment purposes, for example, tests, exams, etc. Most of formal assessments are summative in nature and are associated with increased stress for the learners. However, the formal assessment can motivate students and may be useful to encourage achievement. While informal

assessment can reduce stress as it is often used to provide feedback to students.

Final vs. Continuous Assessment

Final assessment takes place at the end of the course. It is used for summative purposes and it is appropriate when learning is assessed as a complete whole rather than constituent parts. Continuous assessment occurs throughout the course. Due to continuous assessment both students and teachers can obtain continuous feedback across the course which helps to have a picture of students' abilities. This information can be used to improve both teaching and learning processes.

Process vs. Product Assessment

Product assessment aims to evaluate the results or outcome of a process and it is used for summative purposes. Product-based assessment is easier to implement as it has tangible criteria to assess. Process assessment focuses on the steps of undertaken process. It is used for formative purposes, provides with much more detailed information, therefore, the process assessment is useful for learners to acquire new skills and/or improve their performance (pp. 6-10).

2.6 Purpose of Assessment

Different types of assessment can be implemented in class, however, according to McAlpine (2002), they should be designed based on the reason of assessment and should fit the purpose. The assessment works better when its purpose is clear and identifying the aim of any assessment is very important for its effective use. Ko (2004) mentions that assessment should grow out of the learning objectives established for students. Earl and Katz

(2006) identify the following purposes of classroom assessment: *assessment for learning*, *assessment as learning*, and *assessment of learning*. Earl and Katz (2006) describe them in the following way:

Assessment for learning is a process of giving teachers streamline and target instruction and resources. Teacher can use the information to see not only what students know, but also to determine whether students apply what they know and provide feedback to students to help them advance their learning . It is formative in nature and provides information on what an individual student needs to practice and learn next.

Assessment as learning is designed to develop metacognition for students. It helps students to develop capacity to evaluate and adapt their own learning.

Assessment of learning is summative in nature and is used to see whether students have achieved the curriculum outcomes and to determine what the students learn. Thus, it checks learning for reporting.

Earl and Katz (2006) state that there are the varieties of assessment tools that can be used depending on assessment purposes.

2.7 Assessment in Speaking

Most studies examine language and classroom assessment in general. However, according to Mukminatien (2000), assessment of speaking production has got little attention from language teachers and experts. Usually teachers use their own techniques in assessing students` speaking by making notes on the basis of their direct observation on learners` performance in the classroom activities. By doing so, the teacher determines the score on the basis of his/her impression without making analyses of

students' strengths and weaknesses. Such an approach will not have a positive influence on students' learning. That's why various specialists state that assessment shouldn't be ignored in teaching and learning processes and the appropriate assessment techniques should be used in the speaking classes.

Luoma (2004) states that speaking skill is an important part of the curriculum in language teaching and it should make speaking an important object of assessment. From an assessment perspective, speaking is special because of its interactive nature and has various challenges. Besides, speaking assessment is time-consuming and requires a fair amount of work. It includes task design and general action plan. There is a need to develop materials that examinees use while performing the task. Besides, there is a need to develop instruments to assess the examinees' performance (p.51).

In addition, Luoma (2004) states that assessing speaking is also challenging because there are so many factors that influence the raters' impression of how well someone can speak a language. Moreover, there can be variability in the rating process because it involves human raters. This means that special procedures are needed to ensure the reliability of the scores (p.170).

According to Brown (2008), language learning involves experiments. Besides, to develop competence in a foreign language speaking takes a long time. Thus, teachers should use various assessment approaches and tools to help students in their learning process.

In EFL settings assessment of language learners should focus on improving their learning (Pitoniak et.al, 2009). For improving students' learning, *assessment for learning* and *formative assessment* is considered to

be one of the most important and beneficial mechanisms. That's why, the assessment in education is going through a shift from traditional testing of knowledge to *assessment for learning* (Dochy & Segers, 2006 as cited in Jonsson & . Svingby, 2007). Tierney and Simon (2004) state that scoring rubrics are useful in assessment for learning because they contain qualitative descriptions of performance criteria that work well within the process of formative evaluation. One of the feathers of assessment for leaning is an effective feedback. Luoma (2004) states that assessment of speaking skills should be based on more organized strategies for reporting useful feedback to learners. Useful feedback is described as concrete and descriptive, and it relates descriptions of a good performance. Evaluative comments like 'good job' can not help students to develop their speaking, because they do not tell learners about what they did well, or what they might do better. Reddy (2007) states that rubrics are assessment tools which facilitate the process of effective feedback.

Munoz and Alvarez (2009) consider rubrics as one of the components of the Oral Assessment System (OAS). OAS was developed in 2001 with the aim to improve the teaching and learning of oral language. The principles of the system suggest that assessment should be done through clear delineated assessment criteria that are understandable to teachers and student, assessment of different aspects of oral language and detailed and specific feedback. According to Munoz and Alvarez (2009), rubrics are designed to measure separate components of language along grade level descriptors. Mukminatien (2000) finds that assessing learners' speaking ability by detailed descriptions of the components of speaking can be beneficial for students

learning. However, there is little information in the literature about empirical studies on the effectiveness of rubrics used for oral production.

2.8 Background and Definitions of Rubrics

In the past thirty years, along with traditional forms of language assessment, a wide range of alternative assessment forms has become popular. Different scholars support assessment with rubrics as a useful tool that has many advantages and can be beneficial for students.

Schneider (2006) mentions that historically *rubric* was a theological term used to describe an outline, which may be a plan of action. During the late 1980s, rubrics originated in the study of religion. From 1990 rubrics became also a grading tool in education.

There are different definitions of rubrics. Mertler (2001) defines scoring rubric as an effective assessment tool, which provides both a grade (summative assessment) and detailed feedback to improve future performance (formative assessment). According to Schafer (2001), rubrics are vehicles that provide a useful mechanism to translate achievements into assessment terms.

Rubrics are aligned with outcomes based approach to education (Andrady & Du, 2005, Moscal, 2000; Isaacs, 2001; Holmes & Smith, 2003 as cited in Reddy, 2007). In the outcome based approach the focus is on Student Learning (SL) defined as “the change in knowledge, understanding, skills and attitudes brought about by experience and reflection upon that experience” (Reddy, 2007).

Walser (2011) states that rubrics are popular because they can be created and adapted to a variety of subjects and situations. They can be used

across a broad range of subjects (Moscal 2000; Tierney & Simon, 2004 as cited in Reddy, 2007). Thus, rubrics can be used with benefits in EFL classes as well.

2.9 Types and Main Features of Rubrics

There are two types of rubrics: *holistic and analytic*. Nitko (2001), cited in Mertler (2001), summarized them as follows:

- *Holistic rubrics* require the teacher to score the overall process or product as a whole, without judging the component parts separately
- Analytic rubrics require the teacher to identify and assess performance components separately.

According to Popham (1997), cited in Reddy (2007), rubrics have the following essential features: *Evaluation Criteria, Quality Definitions and Scoring Strategy*. Reddy (2007) described these features as the following:

Evaluation criteria are the factors that a teacher considers to evaluate the quality of a student's work. *Quality definitions* give a detailed description and explanation of what a student must do for a particular level of performance. And finally, *scoring strategy* of rubrics could be for scoring the performance entirely (holistic). Scoring strategy could also be for scoring each criterion separately and then form an overall score (analytic).

2.10 Viewpoints about Integrating Scoring Rubrics in Education

Many researchers and assessment specialists recommend using rubrics because of various advantages that help students in their learning. One widely cited effect of rubric use is the promotion of students' learning. Jonsson and Svingby (2007) review empirical research on performance

assessment in combination with rubrics and propose that using scoring rubrics in performance assessment has its benefits and gives positive influence on learning. Andrade (2000) argues that rubrics are easy to use and easy to explain, and when rubrics are used correctly, they serve the purposes of evaluation as well as learning. There are studies showing that learning and academic performance is positively enhanced when rubrics are used to evaluate students' performance (Jonsson & Svingby, 2007).

Andrade (2000) believes that students who use rubrics learn more than students who do not. There are studies reporting and showing the same. Howell (2011) tests the following hypothesis: *Grading rubric use exerts a positive impact on students' academic performance.* The study was conducted with the students of undergraduate elective course, Juvenile Delinquency, offered during spring and fall 2009 at a research university in a Southeastern U.S. town. Students from both semesters received the same instructional materials, were taught by the same instructor using the same course content, were taught in the same manner (that is, Power Point lectures, class discussion, videos, etc.), and received the same number of assignments. The only difference was that the students from the spring semester (control group) completed assignments without rubric; meanwhile the fall semester students (treatment group) used rubrics during their course. Data derived from 80 students participated in the study shows that the students of the treatment group performed better on assignments than the students from the control group.

However, some researchers state that the quality of students' work does not always depend on using rubrics by the students. Green and Bowser (2006) show no differences in the quality of the work done by students with

and without rubrics. They tested rubrics in Best Practices University (BPU) during the fall term in 2004. To test the efficacy of the instrument, it was decided to compare the samples of literature reviews written by the BPU students with rubrics with the samples written without rubrics. The researchers invited independent raters to read and rate samples of literature reviews using rubrics. Due to that experiment no significant differences are found between the scores of samples written without rubrics and the scores of samples written with rubrics. However, Green and Bowser (2006) mention that simply handing out a rubric cannot be expected to have an impact on students' work. Students must be taught to use a rubric.

In opposition to Green and Bowser (2006), Reitmeier, Svendsen, and Vrchota (2004) suggest that rubrics has potential to help students in their learning. They suggest that involving students in the development and sharing rubrics prior to the submission of an assignment is associated with improvements in academic performance. One of the main advantages of scoring rubrics is that rubrics help students to improve their performance by clearly showing what is expected from their work and how it will be evaluated (Schafer, 2001). Rubrics make the learning target clear. If students know what the learning target is, they are able to hit it better (Stiggins, 2001 as cited in Wolf & Stevens, 2007).

According to Nunan (2003), teachers often hear from their students the following words: "I just don't understand what you want". The use of rubrics can make the expectations clear and it will give students an opportunity to improve their achievements. There are studies in the literature review supporting Schafer's, (2001) as well as Nunan's (2003) ideas. A study was conducted in the course called Scientific Study of Food. Throughout the

course students had communicative activities and oral presentations about food. A rubric was developed and distributed among the students beforehand to help them to work on their tasks. As a result, in the interactive classroom students improved their oral presentation skills and the average grade of students was 94%, compared to 86% of the previous semester. The instructors conclude that the rubric described clear expectations and requirements of the assignment which help the students to perform at a higher level.

Andrade (2000) states that rubrics help students think more deeply and understand qualities of good performance. Rubric presents criteria to students and gives knowledge of the qualities of effective work. Explicitness of criteria is important and it can promote students' learning (Arter & McTighe, 2001 as cited in Jonsson & Svingby, 2007).

However, according to Andrade (2001), handing out rubrics can increase students' knowledge of the criteria and knowledge of the qualities of effective performance, but it does not necessarily mean that it will promote students' learning. Andrade (2001) conducted an experiment to examine the impact of rubrics showing the criteria and gradations of quality for essays on students' writing. The participants of the experiments were eighth grade students. Students in the treatment group were provided instructional rubrics for writing three essays. Students in the control group were assigned to write the same essays but did not receive the rubric. At the end of the experiment, students in the treatment group received, on average, higher scores only on one of the three essays. The researcher also administered questionnaire in the experimental and control groups. The analysis of students' responses to the questionnaire revealed significant differences between the treatment and

control groups. It revealed that students in the treatment group manage to identify more criteria by which their writing was evaluated. Thus, the researcher concluded that rubrics can increase students' knowledge of the criteria for writing but 'translating that knowledge into actual writing is more demanding'.

Besides, some researchers find that students may not consider all the criteria in the rubrics, and that's why it is useful to involve students in the creation of rubrics so that to facilitate their comprehension and applications (Huba & Freed, 2000; Stix, 1997; Taggart, Phifer, Nixon & Wood, 2001 cited in Garsia-Ros, 2011).

One of the benefits of using scoring rubrics mentioned in the literature is that rubrics can promote reliability in assessment, which is an important property in educational measurement (Colton, 1997, p. 3 as cited in Oakleaf, 2006). The study conducted by Schneider (2006) provides evidence that the use of rubrics promotes reliability and increases objectivity in grading. In the frame of the study rubrics were introduced to fifty-five students in Kingsborough Community College of the City University of New York. The researcher developed two rubrics for two assignments. The first rubric was used to assess the students' assignments and the students received the rubric after completion of the project. The second rubric was distributed to the students before they completed their second assignment. All students strongly preferred having the rubric in advance as the rubric helped clarify the assignments. Besides, three instructors were trained and used the rubric to grade students' papers. The analysis of the scores given by different raters shows high level of inter-rater reliability. Thus, the researcher concluded that

rubrics give an opportunity to the instructors reliably assess students' performances and can increase objectivity in grading.

Moreover, the study conducted by Simon and Giroux (2001) indicates that the use of rubrics can lead to the relatively common interpretation of students' work by instructors and students. The study was conducted in a Canadian University with students of four graduate and undergraduate courses that used rubrics for self-assessment. The instructor also used the same rubrics to assess the students' portfolios. Analysis of the grades provided both by the students and the professor shows that the average percentage agreement between professor's assessment and students' self-assessment was 75 %. Therefore, it was concluded that rubrics help to have common understanding and interpretation of students' performance, which means that rubrics may promote the consistency of teachers' and students' score.

However, Orsmond and Marry (1996) state that there may be disagreements between the instructors' and students' interpretations of the work. Students might not find the qualities in their works as they have a less developed sense of how to interpret the criteria.

Besides, Jonsson and Svingby (2007) review various articles and state that the use of rubrics can not fully guarantee the consistency in assessment. The variation in raters' grading can occur because of several factors. One of the main reasons of disagreement can be the differences in raters' experience to use rubrics for assessment purposes.

There is evidence of disagreements between raters' scores that used rubrics for assessing students' work. Oakleaf (2006) investigates the viability of a rubric approach in assessment and analyze inter rater reliability of

rubrics. The participants of the study were 25 raters that received training to use rubrics for the assessment purposes. The data for the study was gathered from students' responses to open-ended questions in the Library Online Basic Orientation Comparing (LOBO) online tutorial. Comparing the scores of the raters with the scores of the researcher, the study reports wide variations across raters in terms of consensus.

Another idea that is often discussed in the literature is the potential of rubrics to provide detailed ongoing feedback (Jonsson & Svingby, 2007). As Wiggins (1998) states, providing ongoing feedback is one of the essential elements of assessment that can promote students' learning. Smith and Davies (n.d) state that 'rubrics are essentially qualitative appraisals in that a performance is judged by selecting the pattern of descriptors in the cells of the rubric that best matches the qualities of a performance. Thus, by returning the completed rubric to students they can get diagnostically useful feedback on their work'.

According to Hattie and Timperley (2007), effective feedback has powerful impact on learning and achievement. Effective feedback must answer three major questions asked by a teacher and/or by a student: *Where am I going?* (What are the goals?), *How am I going?* (What progress is being made toward the goal?), and *Where to next?* (What activities need to be undertaken to make better progress. From this perspective the rubrics is an effective assessment tool.

Cothran (n.d.) states that in order to design more effective tools it is important to take into consideration students' points of view regarding the use of rubrics. According to Zeidner (1987, p. 353), cited in Cothran (n.d.), "those who experience an assessment are one of the best sources of information".

To find out learners' attitude towards the use of rubrics, Cothran (n.d.) undertook a study that involved 39 college students enrolled in *an Introduction to Teaching* course at a university in the United States. The students of that course were provided with scoring rubrics to complete their writing assignments. The instructor of the course used the same rubrics to assess students' work. At the end of the course the students filled in a questionnaire and 12 students participated also in the face-to face interview. Results showed students' positive attitude towards the use of rubrics during the course.

A similar study was conducted by Andrade and Du (2005) in the University at Albany. 14 undergraduate students in the focus group were interviewed in the frame of the study. Analyzing focus group data, the researches highlight that students' comments regarding the rubric use were positive. The participants in this study believed that there were positive outcomes associated with rubric use, including better and fairer grades, improvements in the quality of their work, and less anxiety about assignments.

Another study showing the students' attitude towards the use of rubrics was conducted by Ros (2011). The participants of the study were 64 students in Educational Physiology degree program in the 2008-2009 academic years that were provided with rubrics for oral presentations. The results of the responses to the questionnaire about the usefulness of rubrics showed that students especially highlighted the usefulness of rubrics for "having knowledge about grading criteria".

However, Kohn (2006) states that not all students like to use rubrics and that too much attention to the quality of one's performance can be

associated with more superficial thinking and less interest in whatever one is doing. Besides, detailed assessment of students' accomplishment can be counterproductive. A 6th grade student said that "The whole time I am writing, I am not thinking about what I am saying or how I am saying it. I am worried about what grade the teacher will give me, even if she's handed out a rubric. I am more focused on being correct, than on being honest in my writing" (Perchemlides & Coutant , 2004).

2.11 Summary of Related Literature

Reflecting upon the results of different studies, the researchers mainly highlight positive influence of rubrics in education. Mertler (2001) states that rubrics make possible to create a "profile" of students' strengths and weaknesses and provide students with an opportunity to improve their learning. Besides, assessment with rubrics can provide students with a practical "road map" guiding them to evaluate and improve their achievements.

However, there are also controversial ideas about the effectiveness and usefulness of rubrics. There are studies suggesting that there is no difference in students' performance with or without rubrics. Researchers also discussed controversial ideas about the potential of rubrics to promote reliability, which is considered to be an important aspect in assessment.

Therefore, taking into consideration all the controversial discussions mentioned above, it becomes difficult to draw definite conclusions about the effectiveness of rubrics.

Last but not least, as it was mentioned above, the better understanding of students' perspectives on educational innovations like rubrics can provide

key insights into understanding and designing more effective assessment tools (Cothran, n.d.). Thus, learning students' attitude is also needed to have a clear picture about the usefulness of this tool in education.

In addition, previous research focuses on the use of rubrics in education in general. However, there is a lack of studies on rubric and its effect in EFL settings. Moreover, there are no studies investigating the use of rubric in Armenian EFL classes for development speaking skills. Therefore, it is reasonable to conduct an empirical research to understand whether rubrics have positive influence on Armenian EFL learners' speaking. The current study addresses the speaking skill to be surveyed since it is viewed as the most challenging one to be mastered. Besides, it is also reasonable to learn about Armenian students' perception about the use of rubrics in order to use them with benefits.

The research questions of the current study are as follows:

1. To what extent do two raters agree on the scores on speaking tasks when guided by rubrics?
2. What is the impact of analytic rubrics on the improvement of Armenian EFL learners' speaking skills?
3. Do Armenian EFL students who are guided by analytic rubrics perform better on speaking assignments than those who are not?
4. What is Armenian EFL learners' attitude towards the use of rubrics for teaching and assessment processes?

Chapter 3: Methodology

This chapter presents detailed description of the methodology applied in the current study. It introduces the design, setting as well as the participants of the study. It also presents the instrumentation and the employed procedure of data collection.

3.1 Research Design

Due to non-randomized sampling the study has a quasi-experimental design. To answer the research questions of the study both quantitative and qualitative approaches were implemented.

3.2 The Setting of the Study

The study was conducted in EFL context, specifically in the Experimental English Classes (EEC) offered by the Department of English Programs (DEP) in the American University of Armenia (AUA). The course duration was 7 weeks. The classes met twice a week with duration of half an hour.

3.3 The Participants of the Study

The research study was conducted with the EEC students of Communication Level 3 (Com. 3) The level of students' English proficiency was low-intermediate. The total number of the students of two groups was 29, specifically 15 students were in the experimental group and 14 students were included in the control group. However, the data from 19 students (10 in the experimental and 9 in the control group) was analyzed since only 19

students took their pre and post tests as well as completed all the assignments in the frame of the treatment. The participants of the study were 12 males and 17 females. Their age ranged from 12 to 16. The participants were included into the groups based on the results of the EEC placement test they had taken beforehand. Thus, the students of both groups had the comparable level of language proficiency.

3.4 Materials

As the experiment was conducted with the participants from the same level of EEC, the materials used during the course were the same for both groups. The course textbook used for Com.3 is “English in mind 1” by Brian Hart, Herbert Puchta and Jeff Stranks, 2010. It includes 14 units. Based on the syllabus designed for Com 3, students covered five units during the course (Units 1-5). The book contains Student Book, Work Book and two CDs. Besides, the teacher provided the students of both groups with supplementary materials based on needs.

The teacher of both groups was the researcher herself.

3.5 Instrumentation and Procedure of the Study

The instruments for the study were the following:

- speaking proficiency pre and post tests (Appendices A and B)
- four speaking assignments during the course
- a questionnaire with close and open ended items (Appendix C)

3.5.1 Speaking proficiency pre and post tests

Speaking proficiency pre and post tests were conducted with the participants of both experimental and control groups.

The purpose of the pre-test was to find out the initial level of the participants' speaking proficiency. The aim of the post-test was to see whether there was an improvement in speaking proficiency of both groups at the end of the course. Due to the pre-test and post-test results it became possible to compare the improvement in the experimental and control groups and to assess the effectiveness of the treatment in the experimental group.

The pre-test and post-test were designed based on the samples of Cambridge Key English Test for School (KET). KET is a qualification that shows whether a student can deal with everyday written and spoken English at a basic level. The tests were adapted to the students' level.

3.5.1.1 Pilot test

In general, a speaking test consists of two parts: the speaking prompt (SP) or an instrument to elicit students' utterance and a scoring sheet (SS) which contains scoring criteria based on which raters assess students' speaking (Mukminatien, 2000). In the frame of this research the pre-test and post-test had the same format for both groups. Initially the tests included three tasks:

Task 1: Short exchanges with the interlocutor (3 minutes)

Task 2. Interactive task involving both students (2-3 minutes)

Task 3: Pictures description (3 minutes for each student)

In order to check whether the students can complete the above mentioned task types a pilot test was conducted. Two students from the

control group and two students from the experimental group took the pilot test. The pilot and the final tests had different contents so that not make students take the same test twice.

Based on the results of the piloting it was decided to remove one of the tasks. For that task prompt cards should be used to stimulate questions and answers of a non-personal kind. Teacher should give a question card to one candidate and an answer card to the other. Students should ask question and answer them based on the cards. An example of the cards is presented below.

Student A: answers

Student B: questions

<p>Skateboarding Competition</p> <ul style="list-style-type: none"> • for anyone 11 – 15 years old • at Green Park • 20 June • 1st prize A new skateboard 	<p>Skateboarding Competition</p> <ul style="list-style-type: none"> • where ? • for children ? • date ? • what / win ?
--	---

Because of the low level of English proficiency the students could not complete the task which required the use of appropriate tenses. Hence, the structure of the tests was changed and only two tasks remained: *short exchanges with the interlocutor* and *pictures description*.

3.5.1.2 Pre and post tests reliability

Stewart and Gibson (2010) suggest that the pre and post-tests must be equated for meaningful comparisons between their scores. To assess the effectiveness of the treatment in the experimental group at the end of the course, there was a need to assure equivalence between the pre and post tests. Thus, the tests were piloted with ten EEC students from Com.1. Each student took pre and post test. The students were recorded and graded by

two raters. To measure the equivalence between pre and post test scores Cronbach`s Alpha reliability scale was employed. Cronbach's alphas for pre and post results were. 984 (first rater) and .978 (second rater). According to Pallant (2007, p.98), “the values above .89 suggest very good internal reliability for the scale with this sample”. Thus, the results suggest high consistency of scores in the pre and post tests which means that pre and post tests are equivalent.

Besides, as two raters graded the students' speaking during pre and post tests piloting, inter-related reliability of the scores of two raters was also measured. Cronbach's alphas for inter-rater reliability were .987 (pretest) and. 996 (post test). The results show high inter-reliability of the scores of two raters during the pre and post test piloting.

3.5.1.3 Pre and post tests procedure

The pre-test was administered at the beginning of the course. The students took the test in pairs. Each pair came to the test 10 minutes after the previous one. In order to avoid the outflow of information regarding test content, the teachers accompanied the students to the door and invited the next pair to the classroom.

In the first part of the test each student interacted with the teacher, using a language normally associated with meeting people for the first time, giving personal information, such as name, favorite food, hobby, family, etc. Students were expected to be able to talk about daily life, interests, likes, and each student was asked 5-6 questions that required a simple everyday language. The students were expected to participate in conversation and to

provide reasonable answers to the questions. However, when students had difficulties in understanding the questions, the teacher repeated them or asked the same questions in a simpler way. Since two students were interviewing at a time, the questions to each student were different.

In the second part of the test each student was given a picture to think for 30 seconds and then to describe it. When the students had difficulties to start speaking and to describe the pictures, the teacher asked them “scaffolding “questions, such as *What can you see in this picture? Is it a beautiful room? How many children are there in the room? What are they doing? etc.*

The post test was administrated at the end of the course. The procedure of the post- test was the same as in the pre-test.

3.5.1. 4 Assessment of pre and post tests

Students were graded by the same two raters on a basis of an analytic rubric retrieved from The National Capital Language Resource Center (NCLRC) web-cite. The rubric was adopted to use in this study. It is focused on the assessment of pronunciation, fluency, grammar, vocabulary and communicative success (Appendix D). The students were recorded by the teacher and then graded. The recordings were passed to the co-rater to listen to and to grade students' speaking. It should be mentioned that before grading the students' speaking two raters discussed each criterion in the rubric. Besides, they assessed several students separately and then calibrated to see to what extent they agreed on the scores.

3.5.2 Speaking assignments during the course

Throughout the course the students of both groups were assigned four speaking assignments. The topics for the assignments were chosen based on the topics covered within the units of the book used during the course. The distribution of the assignments was the following:

Assignment 1: Presentation on the topic *Where I would like to spend my holidays* or *Where I spent my holidays last year*

Students were assigned to think about the topic and prepare a 3 minute long presentation. The teacher encouraged the students to bring and show pictures related to their presentations.

Assignment 2: Power Point Presentation on *Who`s your hero?*

While covering the unit about different heroes, the teacher encouraged the students to chose a hero and prepare at least 3 minute Power Point Presentation. All the students were experienced in preparing Power Point Presentations.

Assignment 3: Poster Presentation on *My favorite sport or /and sportsmen*

For the third assignment students were assigned to prepare at least 3 minute presentation about their favorite sport or /and sportsmen. They were also assigned to prepare a poster with handwritten or printed text, visuals, pictures, photos related to their presentations. Students, who wanted to speak about the same well-known sportman, were allowed to prepare a group presentation.

Assignment 4: Presentation on *My future profession*

Students were assigned to prepare a presentation about jobs they are interested in and to speak about their future profession. Again the time given for the presentation was at least 3 minutes.

Thus, the students of two groups were assigned to prepare the same presentations. However, at the beginning of the course the rubric was introduced to the students of the experimental group. The teacher explained each component and each criterion of the rubric. The experimental group carried out their speaking assignments by using an analytic rubric as a guideline. They learnt that they would be assessed based on the descriptions included in the rubric. The teacher graded all four tasks by making notes on the rubrics, adding comments, assigning a grade to the entire project (Appendix F). The students received the rubric with grades and comments from the teacher about their oral performances.

Meanwhile, the control group didn't have any exposure to rubric to prepare their assignments. For grading the control group students the teacher used the same rubric as in experimental group, however, the students were not provided with detailed feedback via rubrics. The feedback was provided by the teacher orally in form of grades without detailed discussions and comments about their presentations.

3.5.2.1 Assessment of speaking assignments

Like in the pre-test and post-test, the students' oral production was assessed with two raters. The raters used an analytic rubric developed based on a rubric retrieved from The National Capital Language Resource Center (NCLRC) web-site. However, as the same rubric was distributed among the

students, it was simplified to make criteria clearer for the students. The same rubric was used for all four assignments and was focused on the assessment of pronunciation, fluency, grammar and vocabulary. It also included *Question and Answer (Q & A)* and *Task* sections describing the specific requirements of the assignments. The rubric ranges from *Excellent* to *Need Improvement* for all the components (Appendix E).

3.5.3 Questionnaire

The next procedure of the study was the administration of the questionnaire at the end of the treatment. The questionnaire was administered in the experimental group only, since the participants of the control group did not have any exposure to rubrics.

Total number of the students filled the questionnaire was 10. The questionnaire of 12 items, ten of which were close-ended items and two last questions were open-ended. Quantitative data was collected with first 10 open-ended items. Two items (Q1, Q2) were based on Likert-type scale and participants chose one of the given options (always, sometimes, rarely, never) that better described their answers. The remained close-ended items (Q.8-Q.10) were based on Likert scale ranging from *Strongly Agree* to *Strongly Disagree* in addition with a *Not Sure* option. The qualitative data was collected with two open-ended questions (Q. 11, 12) and was analyzed descriptively.

The questionnaire was translated into Armenian in order to enhance students' comprehension of the items (Appendix C). Besides, the students were informed that the questionnaire was anonymous to encourage them to give honest answers.

Chapter 4: Results of the Study

This chapter presents quantitative and qualitative data analyses to provide answers to the proposed research questions of the study.

4.1 Analysis of the Data for the First Research question

Trying to reveal to what extent two raters agree on the scores on speaking tasks guided by rubrics, statistical analysis was conducted using the reliability test. Cronbach's Alpha reliability scale was employed to measure the inter-rater reliability of the scores of pre test, post test and four speaking assignments. The results in Table 1 show high internal consistency of two raters' scores since, as it was already mentioned "the values above .89 suggest very good internal reliability for the scale with this sample" (Pallant, 2007, p.98). Hence, the results of the data reveal high inter-rater reliability of two raters' scores.

Table 1. Reliability Statistics for Pre and Post-Tests and four Assignments

	Cronbach's Alpha	N of Items
Pre-Test	.950	2
Post-Test	.932	2
Assignments 1	.974	2
Assignment 2	.944	2
Assignment 3	.946	2
Assignment 4	.946	2

4.2 Analysis of the Data for the Second Research Question

The quantitative data to answer the second research question and find out the impact of analytic rubrics on the improvement of Armenian EFL learners' speaking skills, was gathered through pre and post tests. The data was analyzed with Mann-Whitney U test (between group comparisons) and Wilcoxon Signed- Ranks test (within group comparisons).

4.2.1 Mann-Whitney U tests for between group Comparisons

With the aim to determine the results of the pre and post tests between the experimental and control groups, non parametric Mann-Whitney U tests were employed.

Table 2 shows that there is no significant difference in the pre-tests results of the experimental and control groups since the probability value is larger than 0.05 ($Z = -1.721$, $p = .095$). Thus, the students of two groups were at the similar level of speaking skills at the beginning of the course. Besides, the results show that there is no significant difference in the post test results of two groups as well ($Z = -0.657$, $p = .549$). Thus, the students in the experimental and control groups had the similar level of speaking skills at the end of the course.

Table 2 Mann-Whitney U Tests of Pre-Test and Post -Test Scores

	Pre test	Post test
Mann-Whitney U	24.000	37.000
Wilcoxon W	79.000	92.000
Z	-1.721	-.657
Asymp. Sig. (2-tailed)	.085	.511
Exact Sig. [2*(1-tailed Sig.)]	.095^a	.549^a

4.2.2 Wilcoxon Signed-Ranks tests within group comparisons

To analyze and explore to what extent the students of both groups improved their speaking skills Wilcoxon Signed-Ranks test were employed. For this purpose the average ranks of scores of the pre and post tests of each group were compared.

Table 3 displays that there is a significant difference between the pre-test and post- test scores of the experimental group in favor of the post-test as the probability value is less than .05 ($p=.005$). Thus, it can be concluded, that the students of the experimental group improved their speaking skills as they performed better in the post-test implemented at the end of the course. Besides, the results show that there is a significant difference between pre and post tests results in the control group in favor of the post test ($p=.018$). Therefore, the students of the control group also benefited from the course and performed better during their post- test.

Table 3 Wilcoxon Test for Experimental and Control Groups

	posttest – pretest Experimental group	posttest – pretest Control group
Z	-2,812 ^a	-2.375 ^a
Asymp. Sig. (2-tailed)	.005	.018

To sum up, the results revealed the progress in both groups. We may conclude that the instructions of both groups were effective and helped students to improve their speaking skills. However, in order to clarify what

effect the instruction of both groups had on the students' speaking skills, the effect size for Wilcoxon Signed-Ranks tests was calculated.

4.2.2.1 The effect size for Wilcoxon Signed- Ranks test

The effect size is calculated with the aim to quantify the difference between or within two groups and emphasize the size of differences between two samples. According to Coolidge (2000), effect size refers to the effect of the influence of the independent variable upon the dependent variable. The formula for calculating effect size is:

$r = Z/\text{square root of } N$ where N =total number of cases. Thus, the effect size of Wilcoxon Signed- Ranks test for the experimental group was calculated as the following: $r=2.812/\text{square root of } 20 =0.63$. The effect size of Wilcoxon test for the control group is $r=2.375/\text{square root of } 18 =0.56$

According to Cohen`s (1998), criteria, 0.1 stands for small effect, 0.3 stands for medium effect and 0.5 stands for large effect. As we can see, the effect sizes of Wilcoxon Signed-Ranks test for the experimental group (0.63) and control group (0.56) are larger than 0.5 which means large effect in both groups. Therefore, we may conclude that there is no impact of analytic rubrics on the improvement of Armenian EFL learners' speaking skills.

4. 3 Analysis of the Data for the Third Research Question

To provide an answer to the third research question and to find out whether the EFL students who are guided by analytic rubrics perform better on their speaking assignments than those who are not,

results of four speaking assignments implemented during a 7 week course were analyzed with Mann-Whitney U test. The results of four assignments compared across the experimental and control groups show that there is no significant difference in students` performance since the significance level of p for all four assignments are larger than 0.05 (See Table 4). Thus, we may conclude that Armenian EFL students who are guided by analytic rubrics do not perform better on speaking assignments than those who are not.

Table 4 Mann-Whitney U Tests for between Group Comparisons

	<i>Assignment 1</i>	<i>Assignment 2</i>	<i>Assignment 3</i>	<i>Assignment 4</i>
Mann-Whitney U	39.500	43.500	36.000	37.000
Wilcoxon W	84.500	98.500	91.000	82.000
Z	-.451	-.123	-.740	-.656
Asymp. Sig. (2-tailed)	.652	.902	.459	.512
Exact Sig. [2*(1-tailed Sig.)]	.661^a	.905^a	.497^a	.549^a

4.4 Analysis of the Data for the Fourth Research Question

To answer the third research question and disclose the learners' perceptions towards the use of rubrics for teaching and assessment processes both quantitative and qualitative data was collected with an attitudinal anonymous questionnaire administered at the end of the course.

4.4.1 Analysis of the quantitative data

The tables below demonstrate the results of the questionnaire in terms of percentages

Table 5 *Students' exposure to rubrics (N=10)*

Questionnaire items	Always	Sometimes	Rarely	Never
1. Did you use scoring rubrics before this course?	0%	0%	0%	100%
2. You had four speaking tasks during the course. Did you use the rubrics while preparing for these tasks?	70%	30%	0%	0%

Table 5 includes questions regarding students' exposure to rubrics. The first item in the questionnaire aims at finding out whether the students had an experience of using rubrics before this course. The result shows that all the students (100 %) never used rubrics before. Thus, I may assume that may be the concept of assessing through scoring rubrics theoretically familiar to Armenian teachers, lecturers and students, the assessment through rubrics hasn't been practically implemented in Armenian educational institutions. From the next item (item 2) we can see that 70% of the respondents stated that they always used the scoring rubrics in the frame of the speaking tasks of the course and the rest 30% of the respondents used rubrics sometimes. Therefore, we may suppose that all the participants of the experimental group have an exposure to rubrics.

Table 6: *Students` attitude towards rubrics (N=10)*

Questionnaire items	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Sure
3. Rubrics help me to prepare better for speaking tasks	80%	20%	0%	0%	0%
4. Rubrics show me what I need to do for the tasks	90%	10%	0%	0%	0%
5. Rubrics show me how my work will be graded	70%	30%	0%	0%	0%
6. Use of rubrics is useless for me	0%	0%	40%	60%	0%
7. Rubrics show me my achievements	60%	30%	0%	0%	10%
8. Rubrics give good feedback of my work/performance	70%	30%	0%	0%	0%
9. Use of rubrics is waste of time for me	0%	0%	40%	60%	0%
10. Rubrics show me my strengths and weaknesses	70%	30%	0%	0%	0%

Table 6 includes eight close-ended items and answers of participants regarding their perception about using rubrics during the course. As Table 6 demonstrates, all the participants are positively disposed towards the statement that scoring rubrics help them to prepare better for speaking tasks (item 3: 80% strongly agreed and 20% agreed). All the participants also agreed with the next statement (item 4) that scoring rubrics show them what they need to do for the task (90% strongly agreed, and 10% agreed). Besides, 100 % of participants either strongly agreed (70%) or agreed (30 %) that rubrics show the students how their work will be graded (item 5)

To the negative statements that the use of rubrics is useless for students (items 6) as well as that the use of rubrics is just waste of time for

them (item 9), the majority of the learners (60%) strongly disagreed and the rest of them (40%) simply disagreed. These are “signal” items to see whether respondents read attentively questions to answer or just automatically choose one of the given options. Thus, we may suppose that the students provided truthful answers to the questions.

The majority of the students, i.e. 60%, strongly agreed, 30 % agreed, and only 10% was not sure that scoring rubrics show them their achievements.

The next item (item 8) refers to feedback that students have from the rubrics. All participants agreed to the posed idea that rubrics provide students with good feedback of their work. Table 6 also shows that 100 % of participants either strongly agreed (70%) or agreed (30 %) that scoring rubrics show the students their strengths and weaknesses (item 10).

4.4.2 Analysis of the qualitative data

The last two items of the questionnaire (items 11, 12) are open-ended to give the participants an opportunity to express their attitude in a less structured way. Item 11 is formulated as *What do you like best about the rubrics used for speaking tasks?* One of the common answers was that scoring rubrics clearly gave them good feedback on their mistakes.

Here are some quotations:

“ I like to use rubrics as rubrics show me my mistakes. When I see my mistakes I try not to repeat them again. Besides, I can see how my performance was graded”

“I like rubric because it shows my mistakes, for example, what words were mispronounced or what kind of grammar errors I had”

“ I like to use rubric, because it makes clear for me at what points I should pay attention and how should I work. I better remember my mistakes when I see them mentioned in the rubrics”.

Besides, students mentioned that they like to use rubrics to know the criteria of the assignments. Below are some quotations:

«I like to have rubrics and to see how my work will be graded'

“With rubrics I know what I need to do for assignments”

“Rubric gives me a good guidance on how to prepare the task”.

Item 12 was formulated as *What would you like to change in rubrics for speaking tasks?* Nine students out of ten mentioned that they would change nothing about the rubric. Only one student answered that he would like to use a shorter rubric.

Summing up the results of the questionnaire we may conclude that Armenian EFL learners have positive attitude towards the use of rubrics for teaching and assessment processes.

Chapter 5: Discussion and Conclusion

This chapter presents the discussion of the results of both qualitative and quantitative data and presents concluding remarks on the findings of the study. It also includes the delimitations and limitations of the research, pedagogical implementations as well as provides some recommendations for further research studies.

The current study investigates to what extent raters guided by rubrics agree on the scores on speaking tasks. It also investigates the effect of rubrics on the improvement of EFL learners' speaking skills. Besides, this study examines whether the students who use scoring rubrics perform better on their oral assignments than those who do not. It also aims to investigate the students' attitude towards use of rubrics during the course for oral assignments. Therefore, the study aims to answer the following research questions:

1. To what extent do two raters agree on the scores on speaking tasks when guided by rubrics?
2. What is the impact of analytic rubrics on the improvement of Armenian EFL learners' speaking skills?
3. Do Armenian EFL students who are guided by analytic rubrics perform better on speaking assignments than those who are not?
4. What is Armenian EFL learners' attitude towards the use of rubrics for teaching and assessment processes?

5.1 Discussions on the Findings Related the Research Questions

To answer the first research question the results of Cronbach`s Alpha reliability tests employed in the frame of this study were taken into account. From the results of analyses we determined high inter-rater reliability of two raters` scores guided by rubrics to assess students` speaking. However, it is worth mentioning that to assure inter-reliability, two raters discussed the criteria of rubrics (a rubric used for pre and post tests, and a rubric used for speaking assignments) and collaborated to secure internal reliability in assessment. Such preparation for using the rubrics to assess students` work may influence positively on high inter-reliable assessment between raters.

In order to provide an answer to the second research question, the results of the pre and post tests were analyzed using Mann-Whitney U tests to compare the results between two groups as well as Wilcoxon Signed Ranks tests for comparisons within groups.

According to the findings, both the experimental and the control groups had significant improvement in speaking, as both groups performed much better during their post test. Besides, the effect size for Wilcoxon Signed-ranks tests showed that both experimental and non-experimental methods had large effect on the improvement of students` speaking skills. So, we may conclude that there is no impact of analytic rubrics on the improvement of Armenian EFL learners speaking skills.

Due to the results from Mann-Whitney U tests it becomes obvious that the students of two groups were at the similar level of speaking skills at the beginning of the course. There is also no significant difference in the post test results of two groups

There may be various reasons for the above mentioned results. One of the reasons can be the short duration of the study. Seven weeks is not enough time for the new pedagogical tool to take effect. May be a longitudinal study could affect the results of the experiment and provide with better prospective about the use of rubrics for developing EFL students' speaking.

To answer the third research question the data from four speaking assignments was used. The results of Mann-Whitney U tests revealed that there is no significant difference in scores across the four speaking tasks of the experimental and control groups.

Having no significant difference in two groups for the speaking assignments also could be caused by a lack of time to explore rubrics and get used to it. Besides, even though the teacher introduced and discussed with the students the criteria of rubrics, as Orsmond and Marry (1996) state, students might have a less developed sense of how to interpret the criteria in the rubrics. That`s why it is useful to involve students in the creation of rubrics so that to simplify their comprehension and applications (Huba & Freed, 2000; Stix, 1997; Taggart, Phifer, Nixon & Wood, 2001 as cited in Ros, 2011).

In addition, no significant difference in favor of the experimental group also could be caused by overloaded course. Due to EEC syllabus for Communication level 3, the students should cover 5 Units of the book used in the course. They should meet 2 times per week with 1 hour class (total 20 hours). Because of summer period there were some changes in organization of the course, specifically, instead of 10 weeks the course material was covered within 7 weeks. Again students met 2 times per week, but with 1.5 hours class (again 20 hours). During the course the students had a mid term

as well as a final exam. Thus, additional four speaking assignments made their weeks much more overloaded. This may be a reason for not using rubrics by students properly which, in its turn, may lead to the difficulties in uncovering rubrics potential to the full.

Last but not least, while filling the anonymous questionnaire, only 70% of students stated that they always used rubrics to prepare their speaking assignments, and 30 % answered that they used the rubrics sometimes. However, we cannot be sure whether the students truly used the rubrics while preparing their assignments or not. Moreover, students may not consider all the criteria in the rubrics. (Huba & Freed, 2000; Stix, 1997; Taggart, Phifer, Nixon & Wood, 2001 as cited in Ros, 2011). The above mentioned arguments also can have influence on the results of the study.

. To answer the fourth research question about students' perceptions about effectiveness of rubrics the data from questionnaire was used. The students answered that using rubrics showed them their mistakes and helped them to understand what they need to do for the assignments. They stated that with the help of rubrics they knew how their work would be graded and it helped them to prepare better for the assignments, etc. Thus, we may conclude that Armenian EFL students had a positive attitude towards rubrics implementation.

5.2 Delimitations of the Study

One of the delimitations of the study is that the findings are too limited to be generalized to a wide population. The findings can be generalized to pre-intermediate level of EEC students, such as Com1, Com 2, Com 3, etc, whose age ranges from 12 to 16.

Another delimitation of this study is that a cumulative grade for speaking didn't let us find out precisely which aspect of speaking improved more.

5.3 Limitations of the Study

The study has several limitations and we need to be cautious about generalization. The first limitation is a quasi experimental design of the study, i.e. the participants were not chosen randomly.

The second limitation is the limited number of the participants. It should be mentioned that summer period when the course was organized had its negative influence on students' regular participations, specifically many students missed lessons as they were out of the city. Only 19 out of 29 students passed pre and post tests as well as prepared their four speaking assignments. Thus, data of only 19 students was analyzed. From this perspective summertime may be considered as a shortcoming that became a reason for a new limitation of the study, i.e., limited number of participants.

Another limitation of the study is time restriction (7 weeks). Because of time restriction the participants were assigned only four speaking tasks which cannot be considered enough to show significant improvement in favor of the experimental group.

The next limitation is the students' low level of English language proficiency and speaking skills. Low level of English proficiency was an obstacle for students to prepare presentations. Instead of assigned 4 minute long (the shortest) presentation, the majority of students was able to prepare at most 2.5 minute long presentation. However, longer presentations might

give an opportunity to have richer data to analyze and to have a clear picture about their speaking.

Another limitation of the study is that the pre and post tests do not include many tasks. Initially it was supposed to have three tasks in the pre and post tests. However, due to the results of the pilot test it was decided to remove one of the tasks as it was too difficult for the students. It should be also mentioned that limited time didn't allow the teacher to develop and pilot a new task instead of the removed one. According to Pitoniak et al. (2009, p.10), tests with more items or tasks will supply more reliable scores and it is desirable to provide students with multiple opportunities to show what they know and what they can do. Thus, the limited number of tasks in the tests is considered as a limitation of the research.

Finally, one of the limitations is the researcher's inability to assure the truthfulness of responses in questionnaire.

5.4 Suggestions for Further Research

To investigate the viability of a rubric approach in assessment and analyze whether rubrics may promote better inter-rater reliability it will be interesting to conduct a study to compare the inter-rater reliability of the scores when the raters are guided by rubrics and when they are not. Besides, in order to better investigate the effect of rubrics on the development of EFL students' speaking skills, further research are encouraged with involvement of a large number of participants and longer period of time for research. In that case the findings of the study will be much more valid and reliable and could be generalized to other populations. Besides, there is a

need to take into consideration all limitations mentioned above, as they also may have influence on the effectiveness of research.

Moreover, in order to have a clear picture about the potential of rubrics to help Armenian EFL students to improve their oral proficiency of the target language, it would be appropriate to do a longitudinal experiment including samples of populations from different schools, colleges and Universities.

Besides, it will be interesting to analyze the influence of rubrics on development of students' speaking skills separate components, such as pronunciation, fluency, grammar, and vocabulary. In this study, the students' performances were graded with the use of analytic rubrics that include different components of speaking. However, to analyze the results of the tests, the total scores were calculated and used.

5.5 Pedagogical implications

Cothran (2002) states that in order to design more effective tools it is important to take into consideration students' points of view. The questionnaire results reported learners' positive attitude towards rubrics for speaking assignments. That's why, though the findings of the research revealed that there is no impact of rubrics on the improvement of Armenian EFL students' speaking, I would like to recommend teachers to try using this tool

Besides, according to the findings of the research the use of rubrics may ensure the reliability which is viewed as an important issue in assessment. This is another reason to recommend teachers to use rubrics for reliable assessment of students' speaking.

References

- Aldenson, J., & Bachman, L. (2004). *Assessing speaking*. Cambridge University Press, New York
- Andrade, H. (2000). *Using Rubrics to Promote Thinking and Learning*. Retrieved April 19, 2012 from http://www.smallschoolsproject.org/PDFS/coho103/using_rubrics.pdf
- Andrade, H., & Ying, Du. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation*, 10(3), 1-11. Retrieved November 12, 2012 from <http://pareonline.net/pdf/v10n3.pdf>
- Brissenden, G. & Slater, T. (n.d.). *Assessment primer*. Retrieved November 25, 2012 from <http://www.wcer.wisc.edu/archive/cl1/flag/start/primerfull.htm>
- Brown, A. (2008). *Building Rubrics :A Step-By-Step*. Retrieved May 20, 2012 from http://www.linworth.com/pdf/lmc/reviews_and_articles/featured_articles/Brown_January2008.pdf
- Brown, D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY, Longman
- Celce-Murcia, M. & Olshtain, E. (2000). *Discourse and context in language teaching*. Cambridge: Cambridge University Press.
- Cothran, D. (n.d.). *Students' Use of and Perspectives on Rubrics*. Retrieved December 16, 2012 from <http://www.aare.edu.au/03pap/cot03119.pdf>
- Earl, L., & Katz, S. (2006). *Rethinking Classroom Assessment with Purpose in Mind*. Retrieved December 10, 2012 from

http://www.edu.gov.mb.ca/k12/assess/wncp/rethinking_assess_mb.pdf

- Florez, M (1999). *Improving Adult English Language Learners' Speaking Skills* . Retrieved June 1, 2012 from <http://www.marshalladulthoodeducation.org/pdf/briefs/Imprvg%20ELLspkg.Florez.pdf>
- Green, R., & M. Bowser. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, 1(2), 185–202. doi:10.1300/J111v45n01_10
- Halonen, J. (2003). *A Rubric for Learning, Teaching, and Assessing Scientific Inquiry in Psychology*. Retrieved March 18, 2012 from [http://virtual2.yosemite.cc.ca.us/mjcinstruction/CAI/Summer2005Institute/Rubrics/Rubric for Assessing Scientific Inquiry.pdf](http://virtual2.yosemite.cc.ca.us/mjcinstruction/CAI/Summer2005Institute/Rubrics/Rubric%20for%20Assessing%20Scientific%20Inquiry.pdf)
- Hattie, J., & Timperley. H. (2007). The power of Feedback. *Review of Educational Research*, 77 (1), 81-112. DOI: 10.3102/003465430298487
- Howell, R. (2011). Exploring the impact of grading rubrics on academic performance: Findings from a quasi-experimental, pre-post evaluation. *Journal on Excellence in College Teaching*, 22(2), 31-49. Retrieved December 10, 2012 from <http://www.cideronline.org/podcasts/pdf/17.pdf>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.

- Ko, S. (2004). Assessment, Feedback and Rubrics. Retrieved September 20, 2012 from <http://deoracle.org/online-pedagogy/assessment-feedback-rubrics/assessment-feedback-and-rubrics.html>
- Kohn, A. (2006) The Trouble with Rubrics. *The English Journal*. 95 (4), 12-15. Retrieved December 15, 2012 from http://www.csun.edu/~krowlands/Content/Academic_Resources/Composition/Responding/Kohn-Rubrics.pdf
- Luoma, S. (2004). Assessing Speaking. UK, Cambridge University Press
- McAlpine, M. (2002). Principles of assessment. University of Glasgow, *Robert Clark Center for Technological Education*. Retrieved November 10, 2012 from <http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf>
- Mertler, C. (2001). *Designing Scoring Rubrics for Your Classroom*. Retrieved April 10, 2012 from <http://pareonline.net/Articles.htm>
- Minoz, A., & Alvarez, M. (2009). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1) 1–17. DOI: 10.1177/0265532209347148
- Moskal, B. (2003). Recommendations for Developing Classroom Performance Assessments and Scoring Rubrics. Retrieved May 5, 2012 from <http://pareonline.net/Articles.htm>
- Mukminatien, N. (2000). The Advantages of Using an Analytic Scoring Procedure Speaking Assessment. Retrieved May 15, 2012 from <http://www.teflin.org/journal/index.php/teflin/article/viewFile/160/45>
- Nunan, D. (2003). *Practical English language teaching*. USA, NY, McGraw-Hill, Inc.
- Oakleaf, M. (2006). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the*

American Science for Information Science and Technology, 60(5), 969–983. Retrieved December 11, 2012 from <http://meganoakleaf.info/oakleafjasist2009.pdf>

Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education*, 21, 239–250.

Pallent, J. (2007). *SPSS survival manual*. New York: McGraw-Hill.

Perchemlides, N., & Coutant, C. (2004). Growing beyond grades. *Educational Leadership*, 62 (2), 53-56. Retrieved November 25, 2012 from http://www.educationleader.com/subtopicintro/read/ASCD/ASCD_315_1.pdf

Pinchok, N. & C. Brandt, (2009). *Connecting Formative Assessment Research to Practice*. Learning Point Associates. An Introductory Guide for Educators, Learning Point Associates. Retrieved November 20, 2012 from <http://www.learningpt.org/pdfs/FormativeAssessment.pdf>

Pitoniak, M., Young, J., Martiniello, M., King, T., Buteux, A., & Ginsburgh, M. (2009). Guidelines for the Assessment of English Language Learners. Retrieved December 10, 2012 from http://www.ets.org/s/about/pdf/ell_guidelines.pdf.

Reddy, M. (2007). Effect of rubrics on enhancement of student learning. *Educate*, 7 (1), 3-17. Retrieved June 20, 2012 from <http://www.educatejournal.org/index.php?journal=educate&page=article&op=viewFile&path%5B%5D=117&path%5B%5D=148>

Reitmeier, C., Svendsen, K., & Vrchota, D. (2004). Improving oral communication skills of students in food science courses. *Journal of Food Science Education*, 3, 15–20. Retrieved November 10, 2012 from

<http://onlinelibrary.wiley.com/doi/10.1111/j.1541-329.2004.tb00036.x/pdf>

Richards, J. (2008). *Teaching Listening and Speaking From Theory to Practice*.

Retrieved November 25, 2012 from

http://www.cambridge.org/other_files/downloads/esl/booklets/Richards-Teaching-Listening-Speaking.pdf

Richards, J., & Rodgers, T. (2002). *Approaches and methods in language teaching*. Cambridge, Cambridge University Press

Ros, R. (2011). Analysis and validation of a rubric to assess oral presentation skills in university contexts. *Education and Philosophy*, 9 (3), 1043-1062. Retrieved December 10, 2012 from

http://www.investigacionpsicopedagogica.org/revista/articulos/25/ingles/Art_25_639.pdf

Sari, V. (2008). *Improving students` speaking mastery using information gap at the second year of SMP N 3 Kebakkramat Ebakkrama Tkaranganyanr in 2007/2008 academic Year* (Unpublished research paper).

Muhammadiyah University of Surakarta, Indonesia. Retrieved December 10, 2012 from

<http://ru.scribd.com/doc/52764640/9/Principles-of-Teaching-Speaking>

Schafer, W. (2001). Effects of Teacher Knowledge of Rubrics on Student Achievement in Four Content Areas. Retrieved March 6, 2010 from <http://www.education.umd.edu/Depts/EDMS/MARCES/mdarch/pdf/M029774.pdf>

Schneider, F. (2006). Rubrics for teacher education in community college.

The Community College Enterprise, 12 (1), 39–55. Retrieved January 7, 2013 from <http://www.schoolcraft.edu/pdfs/cce/12.1.39-55.pdf>

- Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research and Evaluation*, 7(18), 1-7. Retrieved April 26, 2012, from <http://pareonline.net/getvn.asp?v=7&n=18>
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research and Evaluation*, 9(2), [Electronic Document]. Retrieved April 26, 2011, from <http://PAREonline.net/getvn.asp?v=9&n=2>
- Tuan, L. (2012) Teaching and Assessing Speaking Performance through Analytic Scoring Approach, *Theory and Practice in Language Studies*. Academy Publisher, 2(4), 673-679. Retrieved May 25, 2012 from <http://ojs.academypublisher.com/index.php/tpls/article/view/tpls0204673679/4553>
- Ur, P. (2006). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- Walser, M. (2011). Using a Standard Rubric to Promote High Standards, Fairness, Student Motivation, and Assessment for Learning. *Mountain Rise*, 6 (3), 1-13. Retrieved November 20, 2012 from <http://mountainrise.wcu.edu/index.php/MtnRise/article/view/145/118>
- Wenjie, C. (2009). Using communicative language teaching (CLT) to improve speaking ability of Chinese non-English major students. Retrieved December 12, 2012 from <http://minds.wisconsin.edu/bitstream/handle/1793/34646/Cai,%20Wenjie.pdf?sequence=1>

Widiati, U., & Cahyono, B. (2006). The teaching of EFL speaking in the Indonesian context: the state of the art. *Bahasa Dan Seni*, 34(2), 269-291. Retrieved September 16, 2012 from

<http://jurnal.pdii.lipi.go.id/admin/jurnal/34206269291.pdf>

Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Practice*. San Francisco: Jossey-Bass Retrieved September 20, 2012 from <http://www.edrev.info/reviews/rev.htm>

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, 7 (1), 3-14.

Retrieved November 10, 2012 from

http://www.uncw.edu/cte/et/articles/Vol7_1/Volume0701.pdf

Appendix A

Speaking Proficiency Pre-Test

Timing: 8-12 minutes per pair of students

Test structure: 2 tasks

Task 1 Short exchange with the interlocutor

Timing: 2-3 minutes for each student

Questions:

1. Hello. How are you?
2. What's your name? How do you spell that?
3. Where do you live?
4. What subjects do you like best at school?
5. Tell me about your school. or What do you like about your school?
6. How old are you?
7. When's your birthday?
8. How do you usually celebrate your birthday?
9. Where did you go on holiday last year?
10. Tell me something about your hobbies?
11. What is your favourite sport?
12. Tell me about your family.
13. Tell me about your room
14. What do you enjoy doing in your free time?

Task 2. Describing a picture

Timing: 2-3 minutes for each student

Materials: Picture 1 and Picture 2 (one for each student). *See next page*

Picture 1



Picture 2



Appendix B

Speaking Proficiency Post-Test

Timing: 8-12 minutes per pair of students
Test structure: 2 tasks

Task 1 Short exchange with the interlocutor

Timing: 2-3 minutes for each student

Questions:

1. Hello. How are you?
2. Who is your best friend? How do you spell his/her name?
3. Tell me something about your best friend(s).
4. How do you spend your time with your best friend?
5. What did you do yesterday evening?
6. How many brothers and sisters have you got?
7. What food do you like best?
8. What is your favorite season?
9. (follow up question) Why do like summer/spring/autumn/ winter?
10. What days do you have English lessons?
11. Do you enjoy studying English?
12. What is your favorite book?
13. Do you like watching TV? What programs do you like best?
14. Do you like travelling? Follow up Why ? or Why not?

Task 2. Describing a picture

Timing: 2-3 minutes for each student

Materials: Picture 1 and Picture 2 (one for each student). See next page

Picture 1



Picture 2



Appendix C

Questionnaire

Dear Students,

Please read and fill in the questionnaire about using rubrics during the course.

Q	Choose one of the options	Always	Sometim es	Rarely	Never
1.	Did you use scoring rubrics before this course?				
2.	You had four speaking tasks during the course. Did you use the rubrics while preparing for these tasks?				

Q	Please choose SA (<i>strongly agree</i>), A (<i>agree</i>), D (<i>disagree</i>), SD (<i>strongly disagree</i>), NS (<i>not sure</i>) which best describe your opinion	SA	A	D	SD	NS
3	Rubrics help me to prepare better for speaking tasks					
4	Rubrics show me what I need to do for the tasks					
5	Rubrics show me how my work will be graded					
6	Use of rubrics is useless for me					
7	Rubrics show me my achievements					
8	Rubrics give good feedback on my work					
9	Use of rubrics is waste of time for me					
10	Rubrics show me my strengths and weaknesses					

Q. 11-12. Please describe your opinion in two or three sentences (you can write in Armenian as well)

11. What do you **like** best about rubrics used for speaking tasks?

12. What would you like **to change** in rubrics for speaking tasks?

THANK YOU!

ՀԱՐՑԱԹԵՐԹԻԿ

Հարգելի ուսանողներ, խնդրում եմք կարդալ և լրացնել սույն հարցաթերթերը գնահատման չափանիշները(rubrics) դասընթացի շրջանակներում կիրարկման վերաբերյալ:

Նշեք Ձեզ առավելագույն համապատասխանող տարբերակը	Միշտ	Երբեմն	Հազվա-դեպ	Երբեք
1. Դուք կիրառել եք գնահատման չափանիշները (rubrics)մինչ սույն դասընթացը				
2. Այս դասընթացի շրջանակներում դուք պատրաստել եք չորս բանավոր առաջադրանք: Կիրառել եք արդյոք գնահատման չափանիշները բանավոր առաջադրանքները պատրաստելիս				

Նշեք Ձեզ առավելագույն համապատասխանող տարբերակը	Ամբողջովին համաձայն եմ	Համաձայն եմ	Համաձայն չեմ	Բացարձակ համաձայն չեմ	Համոզված չեմ
3.Գնահատման չափանիշները (rubrics) օգնում են ինձ ավելիլավ պատրաստել իմ բանավոր առաջադրանքները					
4. Գնահատման չափանիշները (rubrics) ցույց են տալիս ինձ թե ինչ է հարակվոր բանավոր առաջադրանքը պատրաստելու համար					
5. Գնահատման չափանիշները (rubrics) ցույց են տալիս թե ինչպես է գնահատվելու իմ աշխատանքը					
6. Գնահատման չափանիշների (rubrics) կիրառումը ինձ համար անօգուտ է					
7. Գնահատման չափանիշները (rubrics) ցույց են տալիս ինձ իմ ձեռքներումները					
8. Գնահատման չափանիշները (rubrics) տալիս են ինձ լավ արձագանք իմ աշխատանքի վերաբերյալ					
9.Գնահատման չափանիշների (rubrics) օգտագործումը ժամանակի վատնում է ինձ համար					
10. Գնահատման չափանիշները (rubrics) ցույց են տալիս ինձ իմ ուժեղ և թույլ կողմերը					

11. Ինչը Ձեզ **դուր եկավ** գնահատման չափանիշները (rubrics) բանավոր առաջադրանքների համար կիրառելու ընթացքում

12. Ինչ կցանկանայիք **փոխել** գնահատման չափանիշներում (rubrics)

Շնորհակալություն մասնակցելու համար

Appendix D

Rubric for pre and post tests

Name _____

Total score _____/20

	4 (Excellent)	3 (Good)	2 (Fair)	1 (Poor)
Pronunciation	No consistent mispronunciation; no phonemic errors good intonation and juncture	Some identifiable deviations in pronunciation, but with no phonemic errors. Non-native accent evident with occasional mispronunciations that do not interfere with understanding.	Identifiable deviations in pronunciation with some phonemic errors. Non-native accent requires careful listening, and mispronunciations lead to occasional misunderstanding.	Frequent pronunciation errors with a heavy non-native accent. Many phonemic errors that make understanding difficult.
Fluency	Speech is effortless and smooth	Speech is mostly smooth but with some hesitation and unevenness caused primarily by rephrasing and groping for words	Speech is slow and often hesitant Sentences may be left uncompleted, but speaker is able to continue, however haltingly	Speech is very slow and exceedingly halting. Difficult for a listener to perceive continuity in utterances and speaker may not be able to continue.
Grammar	Very strong command of grammatical structure and some evidence of difficult, grammatical patterns. No errors.	Good command of grammatical structures but with imperfect control of some patterns. Limited number of errors that are not serious and do not impede comprehension.	Fair control of most basic syntactic patterns. Speaker always conveys meaning in simple sentences. Some important grammatical patterns are uncontrolled and errors may occasionally impede comprehension.	Any accuracy is limited to set or memorized expressions; limited control of even basic syntactic patterns. Frequent errors impede comprehension.
Vocabulary	Very good range of vocabulary with evidence. Strong command of idiomatic expressions. Uses new vocabulary appropriately	Good range of vocabulary with limited evidence. Limited evidence of idiomatic expressions. Sometimes use the vocabulary taught in the class, mainly use the vocabulary appropriately	Adequate range of vocabulary. No evidence of idiomatic expressions. Sometimes use the vocabulary taught in the class, but use the vocabulary inappropriately	Limited range of vocabulary. No command of the vocabulary taught in the class
Q&A	Understands the entire message very well and answers appropriately	Understands the entire message, however, needs explanations of the questions and then answers appropriately.	Has difficulties with understanding questions and sometimes answers inappropriately.	Does not understand the message. Teacher has to repeat or change questions, speak very slowly. Always answers appropriately

Retrieved from www.nclrc.org/essential/assessing/analyticscalepop.htm and adopted

Appendix E
Rubric for Speaking Assignments

Name _____ Total score _____ /24

	 4 (Excellent)	 3 (Good)	 2 (Fair)	 1 (needs improvement)
Pronunciation and voice	<ul style="list-style-type: none"> Doesn't have pronunciation errors. Speaks in a clear, loud voice 	<ul style="list-style-type: none"> Has some pronunciation errors that do not lead to misunderstanding. Speaks in a clear voice most of the time 	<ul style="list-style-type: none"> Has some pronunciation errors that lead to misunderstanding. Often speaks in an unclear voice 	<ul style="list-style-type: none"> Many pronunciation errors that make understanding difficult. Speaks in a voice that is hard to hear
Fluency	Speech is smooth	Speech is smooth but with some hesitation	Speech is slow and often hesitant. Sentences may be left uncompleted, but speaker is able to continue	Speech is very slow and speaker is often not able to continue his/her ideas.
Grammar	No grammar errors Uses the grammar taught in the class correctly.	Some errors that are not serious and do not make understanding difficult. Mainly uses the grammar taught in the class correctly	Some errors that make understanding difficult Often has errors of grammar taught in the class	Many errors that make understanding difficult. Has many errors of grammar taught in the class
Vocabulary	Very good range of vocabulary and always uses new vocabulary taught in the class appropriately	Good range of vocabulary and uses the vocabulary taught in the class, mainly uses the vocabulary appropriately	Sometimes uses the vocabulary taught in the class, often uses the vocabulary inappropriately	Limited range of vocabulary No command of the vocabulary taught in the class
Q&A (question and answer)	Understands questions very well and answers them appropriately	Understands questions, however, sometimes needs explanations of the questions and then answers them appropriately	Has difficulties with understanding questions and sometimes answers inappropriately	Does not understand the questions and always answers inappropriately
Task	Excellent completion all required elements are present: - Speaks more than 3 min. - Speaks to the point all the time - Doesn't read the notes	Good completion of the task: - Speaks 3-2 min. - Speak to the point most of the time - Almost doesn't read the notes	Partial completion of the task: -Speaks 2-1 min. - Often doesn't speak to the point -Often reads the notes	Poor completion of the task. Missing required elements: - Speaks less than 1 min - Does not speak to the point at all - Reads notes and throughout the presentation

Retrieved from www.nclrc.org/essential/assessing/analyticscalepop.htm and adopted

Appendix F

Rubric for Speaking Tasks (example)

Name _____

Total score 20 /24

	 4 (Excellent)	 3 (Good)	 2 (Fair)	 1 (needs improvement)
Pronunciation and voice	<ul style="list-style-type: none"> Doesn't have pronunciation errors. Speaks in a clear, loud voice 	<ul style="list-style-type: none"> Has some pronunciation errors that do not lead to misunderstanding. Speaks in a clear voice most of the time 	<ul style="list-style-type: none"> Has some pronunciation errors that lead to misunderstanding. Often speaks in an unclear voice 	<ul style="list-style-type: none"> Many pronunciation errors that make understanding difficult. Speaks in a voice that is hard to hear
Fluency	Speech is smooth	Speech is smooth but with some hesitation	Speech is slow and often hesitant. Sentences may be left uncompleted, but speaker is able to continue	Speech is very slow and speaker is often not able to continue his/her ideas.
Grammar	No grammar errors Uses the grammar taught in the class correctly.	Some errors that are not serious and do not make understanding difficult. Mainly uses the grammar taught in the class correctly	Some errors that make understanding difficult Often has errors of grammar taught in the class	Many errors that make understanding difficult. Has many errors of grammar taught in the class
Vocabulary	Very good range of vocabulary and always uses new vocabulary taught in the class appropriately	Good range of vocabulary and uses the vocabulary taught in the class, mainly uses the vocabulary appropriately	Sometimes uses the vocabulary taught in the class, often uses the vocabulary inappropriately	Limited range of vocabulary No command of the vocabulary taught in the class
Q&A (question and answer)	Understands questions very well and answers them appropriately	Understands questions, however, sometimes needs explanations of the questions and then answers them appropriately	Has difficulties with understanding questions and sometimes answers inappropriately	Does not understand the questions and always answers inappropriately
Task	Excellent completion all required elements are present: - Speaks more than 3 min. - Speaks to the point all the time - Doesn't read the notes	Good completion of the task: - Speaks 3-2 min. - Speak to the point most of the time - Almost doesn't read the notes	Partial completion of the task -Speaks 2-1 min. - Often doesn't speak to the point -Often reads the presentation notes	Poor completion of the task. Missing required elements: - Speaks less than 1 min - Does not speak to the point at all - Reads notes throughout the presentation

Commented [A2]: .In front of **frant** , purchase ['pɜ:ʃəs], son [sʌn],United States [steits], aunt [ɑ:nt]

Commented [A1]: Excellent. Total score for this section is 3.5

Commented [A3]: good

Commented [A4]: Total score for grammar .2,5

Commented [A5]: didn't live (not wasn't live), to be married to smb (not in), **didn't see** (not didn't saw- Past Simple , please be attentive)

Commented [A6]: good !

Commented [A7]: Excellent

Commented [A8]: Excellent , you presentation lasted more than 5 min.!

Commented [A9]: Great job !