

PROPERTY OF THE
AUA PAPAZIAN LIBRARY

ENGLISH DEPT.

AMERICAN UNIVERSITY OF ARMENIA

Yerevan

**Teachers' and Students' Perceptions
of Scoring Rubrics for Speaking Skills**

by

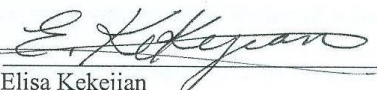
Narine Avtandilyan


2005

The thesis of Narine Avtandilyan is approved.

ENGLISH DEPT.


Marianne Celce-Murcia


Elisa Kekejian


Jo Lewkowitz, Committee Chair

American University of Armenia

2005

Abstract

In the recent shift in educational theory from transmission of knowledge towards transformation of knowledge, and to integration of knowledge with existing personal constructs and meanings, assessment has taken on new affective goals in which the personal growth of the learner is becoming increasingly important. In this context, the role of evaluation is to inform learners about their learning achievements, so that they can make informed plans for future study.

This paper investigates scoring rubrics for speaking skills and teachers' and students' perception of the rubrics. For this purpose, scoring rubrics for speaking skills and teachers' and students' perception of the rubrics are investigated by means of a student and a teacher questionnaire, as well as a teacher interview. The research was conducted in the Intensive English Program (IEP) of the Department of English Programs at the American University of Armenia. Forty-eight students and nine teachers participated in the study. Students were provided with a scoring rubric checklist and after doing an oral presentation for their class, they completed a survey on their use and perceptions of the rubrics. Students also used the rubrics to self-assess and peer evaluate their presentation. Nine teachers completed a questionnaire and four of them participated in the interviews.

The data analyses resulting from student and teacher questionnaires, and the teacher interviews, confirmed students' and teachers' positive attitudes towards the use of scoring rubrics.

Acknowledgements

I am deeply grateful to all the people who have contributed towards the completion of my Master's thesis. I would like to thank my supervisor and my professor, Associate Dean of English Programs at the American University of Armenia, Jo Lewkowicz, the Dean of the Department of English Programs, Marianne Celce-Murcia and the Assistant Dean of the Department of English Programs, Elisa Kekejian, for their insightful comments and encouragement. I owe a great deal to Jo Lewkowicz for her unconditional belief in my ability to complete this work and for reading and giving me feedback on the draft of this thesis. I have tried to consider all her suggestions. I felt Jo Lewkowicz' personal support as a high professional from the very moment I introduced her to my research topic and the first chapter of my thesis. To the extent that I have succeeded, I am grateful to Jo Lewkowicz, who always welcomed me and shared her knowledge and experience with me.

I also greatly appreciate the Assistant Dean of the Department of English Programs, Elisa Kekejian, for giving me permission and directions to carry out this research in the Intensive English Program (IEP) at AUA and the executive secretary Tamara Melkumyan, for her welcoming and supportive attitude, which allowed me to fully attain the goal I had set.

Special thanks are due to IEP instructors Rubina Gasparyan, Lilianna Edilian and Nelly Harutyunyan, who without hesitation, agreed not only to assist me in carrying out research in their classes, but also directly participated in this research. I am also grateful to IEP instructors Armine Khanzatyanyan, Caroline Van Antwerp, Farrell Payne, Lilianna Edilian, Melissa Brown and Talin Grigorian for completing questionnaires and for their (Farrell Payne, Lilianna Edilian, Melissa Brown and Talin Grigorian) participation in interviews, which contributed importantly to my future work. I appreciate the IEP (2004 year) students for their willingness to participate in the research.

My gratitude goes to all concerned.

Table of Contents

Abstract	iii
Acknowledgements	iv
Table of Contents	v
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1 Three Types of Scoring For Performance Assessment	4
2.2 Scoring Rubrics	5
2.2.1 'Instructional illuminators'	5
2.2.2 Rubric definition	5
2.2.3 Rubrics as an assessment tool	8
2.2.4 Involving students in the assessment process	8
2.2.5 The appeal of rubrics for teachers and students	9
2.2.6 Types of rubrics	10
2.2.7 Establishing reliability and validity	12
2.3 Assessment of Spoken Language	13
2.4 Establishing Criteria of Assessment	15
2.5 Use of Rubrics in Armenia	16
2.6 Research Question	17
CHAPTER 3: METHODS	18
3.1 Educational Context of the Study	18
3.2 Participants	18
3.2.1 The students	18
3.2.2 The teachers	19
3.3 Materials	19
3.3.1 Description of scoring rubric checklist	19
3.3.2 Description of student questionnaire	20
3.3.3 Description of teacher questionnaire	21
3.3.4 Description of teacher interview	21
3.4 Procedures	21
3.4.1 Procedure used with students	22

3.4.2 Procedure used with teachers	24
CHAPTER 4: DATA ANALYSIS AND RESULTS	26
4.1 Analyses of the Student Questionnaire	26
4.1.1 Personal information	26
4.1.2 Use of rubrics	26
4.1.3 Positive and negative aspects of using rubrics	28
4.1.4 Self-assessing speaking abilities	31
4.2 Analysis of the Instructor Questionnaire and Instructor Interview	33
4.2.1 Questionnaires	33
4.2.2 Interviews	35
4.2.3 Correlation of student's self-assessment grade and teacher grade	40
CHAPTER 5: DISCUSSION AND CONCLUSIONS	41
5.1 Teacher and Student Perception of Scoring Rubrics	41
5.2 Involvement of Students as Participants in the Assessment Process	43
5.3 Conclusions and Implications	45
5.4 Limitations of the Study and Future Directions	46
5.2.1. Conclusions and implications	45
5.2.2. Limitations of the study and future directions	46
REFERENCES	48
Appendix 1: Types of Scoring Instruments for Performance Assessments	52
Appendix 2: Scoring Rubrics	53
Appendix 3: Overall Band Scores For IELTS	54
Appendix 4: Scoring Rubric Checklist	55
Appendix 5: Student Questionnaire	56
Appendix 6: Teacher Questionnaire	60
Appendix 7: Sample Teacher Interview Typescript	63
Appendix 8: Profile of Survey Sample	70
Appendix 9: Frequencies and Frequency Tables of the Open-ended Questions of the Student Questionnaire	71
Appendix 10: Correlation of Student's Self-Assessment and Teacher's Evaluation Grades	74

Chapter 1: Introduction

“Traditionally assessment has not been viewed as a critical component of research on effective teaching. For a variety of reasons, the educational landscape has changed and assessment is now a critical concern”.

(Shepard, 2001)

During the past twenty-five years, along with traditional forms of assessment, a wide range of alternative assessment forms have become popular in foreign language learning and teaching. After decades of multiple-choice items being used as a dominant item format in large-scale standardized assessments, educators and test developers have realized the need to change the current practice of assessment to follow the trend of educational reform. As a result there has been a move towards more authentic, performance assessments (Frechtling, Hansen, Zhang & Nyre, 2002). According to the current *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999), performance assessment is defined as:

Product- and behavior-based measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied (p.179).

Performance assessment, also known as alternative or authentic assessment, is a form of testing that requires students to perform a task rather than select an answer from a ready-made list (<http://www.ed.gov/pubs/OR/ConsumerGuides/perfasse.html>).

Performance assessments can take on many different forms, which include written and oral demonstrations and activities that can be completed by either a group or an individual (Moskal, 2000). A factor that distinguishes performance assessments from other extended response activities is that they require students to demonstrate the application of knowledge to a particular context. Through observation or analysis of a student's response, the teacher can determine what the student knows, what the student does not know, and what

misconceptions the student holds with respect to the purpose of the assessment. Authentic, performance-based learning is a great way to make learning meaningful to students and to encourage them to be creative, innovative, and constructive. However, assessing student projects can sometimes be a problem because there is no clear answer or solution. For this reason, rubrics have become increasingly popular.

Various authors and researchers agree on scoring rubrics as a vital part of learner autonomy (Mertler, 2001; Moskal, 2000; Stix, 1997), providing the opportunity for learners to self- or peer assess their own progress and thus helping them to focus on their own learning. Popham (1997) sees them as “instructional illuminators” (cited in Tierney & Simon, 2004, without pagination). Mertler (2001) finds rubrics beneficial when evaluating student performances or products resulting from a performance task.

In view of this consensus of opinion regarding the desirability of the use of scoring rubrics, it appears reasonable to try to introduce such practices into EFL classes in Armenia. At tertiary level, an alternative approach to assessment does not present significant problems, since further education views students as adults who are studying of their own violation, and who can be expected to take on responsibility for learning.

The aim of this paper is to investigate scoring rubrics for speaking skills and teachers’ and students’ perceptions of the rubrics. This paper will proceed as follows. In the first part of the paper, I will first introduce three different types of scoring for performance assessment, and then I will explain the basic concepts of scoring rubrics. After that, I will discuss assessment of spoken language. I will report why it is necessary to establish criteria of assessment for measuring the quality of spoken performance. In the second part of the paper, I will show the context of the study and provide details of how the study was conducted and how the results were analyzed. In chapter 3, I will report findings of the study. In particular, I will analyze teacher and student questionnaires using descriptive statistics (percentages) in

order to investigate teacher and student likes and dislikes of scoring rubrics for speaking skills. I will present the results of a correlation to determine the relationship of the student's self-assessment grade with the teacher's evaluation grade of the presentation. In the final chapter I will conclude with a discussion of the overall picture that emerges from the findings. These findings, as well as their implications, will be discussed in detail and the conclusions and recommendations for further research will be presented.

Chapter 2: Literature Review

Introduction

In this literature review, I will first introduce three different types of scoring for performance assessment and then I will explain the basic concepts of scoring rubrics. After that, I will discuss assessment of spoken language. I will report why it is necessary to establish criteria of assessment for measuring the quality of spoken performance, especially in the Armenian situation, where educational policy has undergone certain changes, which have attempted to shift the main focus from the teacher to the student, thus actively engaging students in the learning process and developing learner autonomy.

2.1 Three Types of Scoring For Performance Assessment

Richardson (2001) introduces three different types of scoring for performance assessment (see, Appendix1):

- Checklists
- Rubrics
- Rating scales

The author specifies that checklists contain a list of behaviors or specific steps, which can be marked as correct/incorrect, appropriate/inappropriate. Checklists are simply lists of criteria that are checked off as they are accomplished. They are often used to make clear the specific directions or procedures that need to be followed or to spell out everything that needs to be included in an assignment or project. Checklists help to provide a structure for students

and they are useful tools to use when one wants to note the completion of a task, but they do not necessitate the assignment of a rating.

Rating scales are also used; they can be numerical, qualitative, or a combination of numerical and qualitative. Richardson (2001) sees rating scales as ways of attaching quality to various elements of the process or product.

As for the third type of assessment form, a rubric, the author sees it as “a means of scoring a performance assessment wherein multiple criteria are being assessed and quality of performance or product is important” (Richardson, 2001, p.8).

2.2 Scoring Rubrics

2.2.1. ‘Instructional illuminators’

There is an extensive literature on scoring rubrics, which discusses basic concepts of scoring rubrics (Mertler, 2001; Moskal, 2000; Simon & Forgette-Giroux, 2001; Tierney & Simon, 2004). Rubrics are performance-based assessments that evaluate student performance on any given task or set of tasks that ultimately leads to a final product, or learning outcome. Rubrics use specific criteria as a basis for evaluating or assessing student performances as indicated in narrative descriptions that are separated into levels of possible performance related to a given task. Starting with the highest level and progressing to the lowest, these levels of performance are used to assess the defined set of tasks as they relate to a final product or behavior. Each level describes degrees of proficiency and each level is assigned a value that rates the degree of proficiency or student performance.

In recent years, many educational researchers have noted the instructional benefits of scoring rubrics (for example, Andrade, 2000a; Arter & McTighe, 2001). Popham (1997) noted their potential as “instructional illuminators” (as cited in Tierney & Simon, 2004). Mertler (2001) finds rubrics beneficial as rubrics are typically the specific form of scoring used when evaluating student performances or products resulting from a performance task.

2.2.2. Rubrics: a definition

A rubric is a scoring tool that lists the criteria for a piece of work, or “what counts” (for example, purpose, organization, details, voice, and mechanics are often what count in a piece of writing). It also articulates gradations of quality for each criterion, from excellent to poor.

(Andrade, 2000b, without pagination)

Generally, rubrics specify the level of performance expected for several levels of quality. These levels of quality may be written as descriptive ratings (e.g., Excellent, Good, Needs Improvement) or as numerical scores (e.g., 4, 3, 2, 1) which are then added up to form a total score associated with a grade (e.g., A, B, C, etc). Brookhart (1999) claims that scoring rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of the products or processes of students' efforts. According to Moskal (2000), scoring rubrics provide specific ways to classify and judge student products or behaviors into categories that vary on a continuum from unacceptable to outstanding. They can be used to sort out varying degrees of achievement in any product or behavior (essays, research reports, portfolios, oral presentations, performances). Tierney & Simon (2004) support the use of scoring rubrics emphasizing the fact that they are useful in assessment for learning because detailed criteria are delineated and used to discriminate among levels of student performance.

“A rubric is a carefully designed ratings chart that is drawn up jointly by the teacher and students” (Stix, 1997, without pagination). Most rubrics consist of objectives, performance characteristics, and points or scores that indicate the degree to which the objectives were met. Rubrics should be introduced to the students at the very beginning of a project unit by, either presenting the rubric to the class or collaborating with the students to structure the rubric. This will allow students to understand the criteria for assessment before they start the project. Rubrics also help to make the evaluation of the project more objective and consistent.

McDaniel (1994, cited in <http://its.monmouth.edu/FacultyResourceCenter/rubrics.htm>) sees scoring rubrics as a set of categories which define and describe the important components of the work being completed, critiqued, or assessed. Each category contains a gradation of levels of completion or competence with a score assigned to each level and a clear description of what criteria need to be met to attain the score at each level. The characteristics of a typical scoring rubric are summed on the following website:

http://www.relearning.org/resources/PDF/rubric_sampler.pdf

1. It contains a scale of possible points to be assigned in scoring work, on a continuum of quality. High numbers usually are assigned to the best performances: scales typically use 4, 5 or 6 as the top score, down to 1 or 0 for the lowest scores in performance assessment.
2. It provides descriptors for each level of performance to enable more reliable and unbiased scoring.
3. It is either holistic or analytic. If holistic, a rubric has only one general descriptor for performance as a whole. If analytic, there are multiple rubrics corresponding to each independent dimension (such as syntax, focus, and voice in writing; or delivery,

communicative ability, content, organization in spoken assessment) of performance being scored.

4. It is generic, genre, or task specific. If generic, it can be used to judge a very broad performance, such as communication or problem solving. If genre specific, it applies to a more specific type of performance within the broad performance category (e.g. essay, speech, or narrative as forms of communication; open-ended problems or closed-ended problems as kinds of problems solved). Task specific is unique to a single task.
5. It may be longitudinal. It measures progress over time toward mastery of educational objectives such that we assess developmental change in sophistication or level of performance.

2.2.3. Rubrics as an assessment tool

An opinion expressed by Starr (2000) is that rubrics can help students and teachers define "quality". For a teacher, it serves as a target for which the student should strive in order to obtain a good grade. The rubric lays out a teacher's expectations on what an exemplary project contains. For a student, rubrics offer a platform from which the student may begin his or her project. With a rubric as a guide, students receive clarity of the assignment and do not feel so overwhelmed by the massiveness of a given project. Such scoring rubrics also make students sensitive to their own needs instead of relying entirely on their teacher's opinion.

Upbin (1999) aptly notes that the power of rubrics should not be underestimated. They are extremely effective for many classroom activities, including assessments of essays, oral presentations, group projects, and other assignments.

2.2.4. Involving students in the assessment process

Allowing students to assist in the creation of rubrics may be a good learning experience for them. They can be engaged in this process by showing them examples of the same task performed/project completed at different levels and discuss to what degree the different elements of the criteria were displayed. However, if students do not help to create the different rubrics, the teacher may be willing to share the rubrics with the students before they complete the task or project.

Reading or listening to a teacher's expectations is very different for a student than creating and accomplishing his or her own goals. The purpose of inviting students to develop their own evaluation structure is to improve their motivation, interest, and performance in the project. As students' overall participation in creation of rubrics increases, they are likely to excel in learning (Upbin, 1999). "Negotiable contracting is a new approach to involving students in the assessment process" (Stix, 1997, without pagination). Students are motivated intrinsically to design their own assessment tool. Once students have invested a significant amount of time, effort, and energy into a project, they naturally want to participate in deciding how it will be evaluated. The knowledge gained through experience in a particular field of study provides the foundation for creating a useful rubric.

As Wiggins (1993, p.52) suggests, the art of negotiable contracting consists of giving students shared ownership in their own learning. Students themselves can be involved in the assessment process through both peer and self-assessment. A major argument for involving students in self and peer-assessment is that it helps them to develop the ability to make judgments, in particular about themselves and their work. Self assessment, or self-evaluation, can be an important and effective tool in student growth. As for peer evaluation, it can be a useful and valuable tool in helping students to develop their critical skills and insight into the

evaluation process. By making a critical appraisal of another student's work or performance, students can begin to understand the requirements of the curriculum and the teacher.

2.2.5. The appeal of rubrics for teachers and students

Andrade (2000b) pinpoints a few basic reasons that rubrics appeal to teachers and students. They are powerful tools for both teaching and assessment. Rubrics describe to students in a consistent, fair and clear manner what is expected. Rubrics also can assist in improving students' performance in addition to assessing it. They give students important skills, allowing the production of quality work to become a habit for students and moving them to a fuller realization of their potential. The best argument for using rubrics is that they often result in improvement in the quality of students' work.

Teachers appreciate rubrics because rubrics lessen the amount of time they need to spend on evaluating student work. When a piece of work has been self- and peer-assessed according to a rubric, teachers have little to suggest in terms of improvements.

2.2.6. Types of rubrics

There are two major types of rubrics: analytic and holistic (see, for example, Mertler, 2001; Nitko, 2001), (Appendix 2). Analytical scales offer a separate scale for various dimensions, while holistic scales offer several dimensions together.

Discussing an analytic rubric, Richardson (2001) states the opinion that analytic scoring breaks down the objective or final product into component parts and each part is scored independently. In this case, the total score is the sum of the rating for all of the parts that are being evaluated. Mueller's (2003, without pagination) definition of an analytic rubric as a tool "which articulates levels of performance for each criterion so the teacher can assess

student performance on each criterion,” correlates with that of Moskal’s (2000, without pagination) definition, which states that if an analytic scoring rubric is created, then “each criterion is considered separately as the descriptions of the different score levels are developed”.

Sometimes a rubric is scored holistically, meaning there is one overall score instead of discrete dimensions. Holistic scales are used when one overall score is more important than sub-scores for specific categories. It is based on an overall impression of a student’s work as a whole, producing only a single score based on an established scale. The teacher scores separate, individual parts of the product or performance first, and then sums the individual scores to obtain a total score (Moskal, 2000). “The use of holistic rubrics is probably more appropriate when performance tasks require students to create some sort of response and where there is no definitive correct answer” (Nitko, 2001, cited in Mertler, 2001, without pagination). Holistic scoring is preferred when a consistent overall judgment is desired and when the skills being assessed are complex and highly interrelated. However, even though holistic scoring is often more efficient, analytical scoring systems generally provide more detailed information that may be useful in planning and improving instruction and communicating with students.

2.2.7. Establishing reliability and validity

Based on the work of the American Educational Research Association, American Psychological Association and National Council on Measurement in Education, Moskal & Leydens (2000) provide a framework for developing scoring rubrics and discuss the issues of validity and reliability. They also provide clear definitions of the terms “validity” and “reliability” and illustrate these definitions through examples. They examine three types of

evidence to support the validity of an assessment instrument: content, construct, and criterion and consider validity in the development of scoring rubrics. Moskal & Leydens (2000) point out that carefully designed scoring rubrics (analytic or holistic) of any type are likely to produce valid and reliable results. In addition, the reliability of assessment results depends on the scale's capability to direct to a regular and consistent analysis of student performance (Simon & Forgette-Giroux, 2001).

2.3 Assessment of Spoken Language

“Speaking skills are an important part of the curriculum in language teaching, and this makes them an important object of assessment as well” (Luoma, 2004, p.1). The focus of foreign language education is communication, and oral language is central to the teaching of foreign languages at all levels. For many years now, teachers have been moving away from teaching language in isolation in favor of teaching language through authentic tasks reflecting real-life situations. However, assessment practices have not always reflected how language is being taught and to date, the issues surrounding the assessment of second language learners' speaking abilities have not been explored in depth. Thus, oral language in the foreign language classroom is the most problematic of all the skills to assess (National Communication Association, 2005; Assessment Resource Library, 1998)

In relation to the assessment of speaking skills Weir (1993, p.41) observes that before designing oral assessment tasks there needs to be a clear idea of the purpose of assessment. This is essential because the same degree of detail is not required in every testing situation. The purpose of the test will determine the overall shape of the assessment criteria to be used.

The method used for assessing oral communication skills depends on the purpose of the assessment (Moskal, 2003). A method that is appropriate for giving feedback to students

who are learning a new skill is not appropriate for evaluating students at the end of a course. However, any assessment method should adhere to the measurement principles of reliability, validity, and fairness. The instrument must be accurate and consistent, it must represent the abilities we wish to measure, and it must operate in the same way with a wide range of students (Moskal, & Leydens, 2000; National Communication Association, 2005).

Assessing speaking is a process with many stages. At each stage, people act and interact to produce something for the next stage. While the assessment developers are the key players in the speaking assessment cycle, the examinees, interlocutors, raters and score users also have a role to play in the activities. The activity cycle of assessing speaking discussed by Luoma (2004) may be summarized as follows:

- Determining the need for a speaking assessment.
- Planning and developing the stage during which, in a shorter or longer process the developers define exactly what it is that needs to be assessed;
- Developing, trying out and revising the tasks, rating criteria and administering procedures that implement this intention.
- Setting up quality assurance procedures to help them monitor everything that happens in the assessment cycle.
- Using the assessment.

(Adopted from Luoma, 2004, p.5)

Oral language has always been a large component of the foreign language class, and class participation has always accounted for part of the oral grade assigned by teachers. Students have usually received broad guidelines about contributing and participating in the classroom activities, but they may not have been sure of the specific linguistic expectations. Teachers, then, have assigned students a grade based on their combined observations of the students, but because of the lack of time, these observations often have gone unrecorded.

Students may, on occasion, have disagreed and confronted the teacher when they felt that the assigned grade was not fair, since, in their eyes, they had met the requirements for participation. Traditionally, the more formal assessments of oral language have taken several forms. Teachers have graded the language excerpts and have considered them good, average, or poor based on their inner feeling. Teachers inherently know when something is good or bad, but they may not be as comfortable in giving students useful feedback. Or teachers have determined that mispronounced words, grammatical mistakes, wrong choice of words, and hesitations were assigned a point value which was deducted from the students' oral presentation, and, in doing, so they have devalued the content in favor of the form (Assessment Resource Library, 1998; Mantero, 2002; Shaaban, 2001).

Weir (1993) gives a detailed description of how we are to measure the quality of the output, which results from the spoken language tasks we adopt. Nunan (1993, cited in Luoma, 2004, p. 30) defines a communicative task as:

... a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is principally focused on meaning rather than form.

The relationship between a task and the criteria that can be applied to its product is an essential factor in taking decisions on what to include in a test of spoken production (Weir, 1993). Weir finds that tasks cannot be considered separately from the criteria that might be applied to the performances they result in. Having established suitable tasks and appropriate assessment criteria to accompany them, consideration needs to be given as to how best to apply the criteria to the samples of task performance.

2.4 Establishing Criteria of Assessment

“In order to measure the quality of spoken performance, we first need to establish criteria of assessment” (Weir, 1993, p. 41). In establishing a rating procedure, we need to consider the criteria by which performances at a given level will be recognized, and then to decide how many different levels of performance we wish to distinguish. Weir (1993) finds that the answers to these questions determine the basic framework or orientation for the rating procedure. Whether holistic scales or analytical scales are used, the important factors in developing effective rubrics is the use of clear criteria that will be used to rate a student's work and that the performance being evaluated is directly observable (McNamara, 2000). More importantly, students should be informed as to what criteria they are being held accountable.

A criterion is a principle or standard by which a thing is judged. To test oral language skills there need to be such criteria to act as guidelines for judgment. These should describe the various levels of performance in a way that can be tested both logically and consistently. “The use of rubrics is more likely to provide qualitative, meaningful, and stable appraisals than are traditional (pass/fail) scoring methods” (Simon and Forgette-Giroux (2001, without pagination). According to Simon and Forgette-Giroux (2001, without pagination), the assessment based on such a scale offers several advantages, which can be summed up as follows:

- It presents a continuum of performance levels, defined in terms of selected criteria, towards full attainment or development of the targeted skills.
- It provides qualitative information regarding the observed performance in relation to a desired one.
- Its application, at regular intervals, tracks the student's progress in his or her skill mastery.

Students must understand the goals we expect them to achieve in assignments, and importantly, the criteria we use to determine how well they have achieved those goals.

Several authors' (Andrade, 2000b; Stix, 1997; Upbin, 1999) views are consonant with the opinion that rubrics provide a readily accessible way of communicating and developing teachers' goals with students and the criteria teachers use to discern how well students have reached them.

2.5 Use of Rubrics in Armenia

In Armenian educational institutions rubrics are a new concept not only when related to the assessment of foreign language speaking skills, but when related to the other skills as well. However, the American University of Armenia and the British Council are the exceptions, since they implement scoring rubrics when assessing students' foreign language abilities. At the American University of Armenia, rubrics are used as an assessment tool for assessing students' speaking skills during their presentations, as well as for their writing. The British Council in Yerevan uses speaking scale/rubrics to assess speaking skills, as well as other skills of those taking the IELTS examination for immigration or academic purposes. The main reason underlying this test is assessing how well one can understand and use the kind of English one needs to study in an English speaking country. A Band Score for each of the four skills (listening, reading, writing and speaking), as well as an overall score is recorded on the Test Report Form (see Appendix 3).

2.6 Research Questions

The assessment of language learners is of growing importance in English language teaching and applied linguistics. While there has been a considerable body of work

undertaken on assessing the writing abilities of language learners, the issues surrounding the assessment of learners' speaking abilities are still emerging (Luoma, 2004).

To my knowledge, research on scoring rubrics for spoken assessment in Armenia has yet to been done, at least no such study has been reported officially. Consequently, research on scoring rubrics for spoken assessment is tremendously important for the local situation. Such research will help us to better understand oral assessment. Thus, it is the purpose of this study to investigate teachers' and students' use of and perception on scoring rubrics for speaking skills. Specifically, the questions, which guided this study, were:

(1) What are teachers' and students' perceptions of the use of rubrics as a speaking assessment aid?

(2) Can students use rubrics to self-assess accurately?

A better understanding of students' perceptions of educational innovations like rubrics can provide key insights into understanding the teaching-learning relationship. These results could also be used to design more effective assessment tools that take into consideration students' perceptions.

Chapter 3: Method

Introduction

This study set out to investigate teachers' and students' use of and perception of scoring rubrics for speaking skills within the tertiary context at one university in Armenia. This chapter introduces the context of the study and provides details of how the study was conducted and how the results were analyzed.

3.1 Educational Context of the Study

The research was conducted in the Intensive English program (IEP) at the American University of Armenia (AUA). The IEP is a six-month course, required for students whose TOEFL score is less than 570, and which focuses on the development of academic reading, writing, listening, and speaking skills and strategies. A two-month ESP (English for Specific Purposes) course is taught at the end of the IEP. Completion of the IEP is one of the admissions requirements and is the first step towards becoming a degree program student. However, applicants with a TOEFL score of 570 or higher are not required to enroll in IEP.

All the participants of this study, students and teachers, are from the IEP program. The study concentrated on assessment of oral skills within the program, specifically that of oral presentation.

3.2 Participants

3.2.1 The students:

The participants were 48 IEP students at AUA. All were intermediate, high intermediate or advanced level, male and female Armenian students. The students' first

language is Armenian and they all speak Russian as a second language. The students' ages range from 22 to 40. They have a BA in hard sciences or the humanities from higher educational institutions in Armenia and aim for an MA in Business Administration, Political Science, Engineering or Law. All the students took an institutional entrance TOEFL and were placed into groups according to their scores.

AUA students are highly motivated as studying and graduating from this educational institution provides opportunities for professional growth and career advancement.

3.2.2 Teachers:

The other participants of this study were nine IEP teachers at AUA. The teachers, who have used scoring rubrics before, were female Armenian and American professionals in the teaching field. All of them have either a Certificate or Master's degree in Teaching English as a Foreign/Second Language.

3.3 Materials

In this research, which investigates the teachers' and students' use of and perceptions on scoring rubrics for speaking skills, the following materials were used:

- Scoring Rubric Checklist (Adopted from the DEP, AUA; see Appendix 4)
- Student Questionnaire (see Appendix 5)
- Teacher Questionnaire (see Appendix 6)
- Taped Teacher Interview (see Appendix 7)

3.3.1 Description of scoring rubric checklist

The scoring rubric-checklist used in this study is a common institutional IEP evaluation form used to assess peer and student presentations. In this research the scoring

rubric checklist would clarify the concept of scoring rubrics, enable the students to understand the terminology, and prepare them to give meaningful answers to the questionnaire that was distributed later.

The scoring rubric checklist includes scoring rubrics for evaluating formal presentations. The criteria consist of scoring rubrics for “Delivery”, “Communicative ability”, “Content”, “Organization” and “Discussion”. A 4 point grading scale of 3=Very Good, 2=Satisfactory, 1=Weak, 0=Unacceptable, is used.

3.3.2 Description of student questionnaire

The student questionnaire consists of 21 questions related to the following areas: personal data (age, gender, and years of learning English, TOEFL score, and information about taking standardized tests), use of scoring rubrics, perception of scoring rubrics, and self-assessment of English speaking skills. The questionnaire includes closed-ended questions; there are only three questions requiring open-ended, short answers. The closed-ended questions are rated on a four, a five or a six point Likert-like scale and the answers are interpreted as quantitative data. The use of scoring rubrics section of the questionnaire, section A, includes questions 6 to 15. All of these questions ask the students about the use and helpfulness of rubrics while preparing for the presentation.

The three open-ended short questions of section B, questions 13 to 15, are related to the positive and negative aspects of using rubrics. These questions were categorized, coded and their frequencies were counted. Section C of the questionnaire refers to self-assessment of students' English speaking skills and includes questions 16 to 21. These questions are rated on a four, a five or a six-point Likert scale: each pre-determined response option is assigned a number (e.g., ‘Advanced High’=6, ‘Advanced Mid’=5, ‘Advanced Low’=4,

'Intermediate High =3, Intermediate Mid=2, Intermediate Low=1). To process the questions rated on a Likert scale, the values were ordered on a 'frequency' continuum as ordinal data.

3.3.3 Description of teacher questionnaire

The teacher questionnaire consists of 9 closed questions related to the development of scoring rubrics in foreign language assessment. The closed questions are rated on a five-point Likert-like scale and the answers are interpreted as quantitative data. Each questionnaire was given a unique identification code in order to retain the anonymity of respondents. To process the data the respondents' answers were converted to numbers by means of coding procedures (Dornyei, 2003, p. 97). Each item was coded and the answer was converted into numerical scores. The survey answers were analyzed via descriptive statistics.

3.3.4 Description of teacher interview

To better understand teacher perceptions of scoring rubrics for speaking skills, the teachers were interviewed. Four IEP teachers were willing to participate in questionnaire follow-up interviews. In this survey, interviews were included to make "data valid"¹. Interviews were conducted at the teachers' convenience at their offices. Each interview ranged from 30 minutes to one hour. Interviews included structured questions based on questionnaire statements. Furthermore, the teachers were asked several additional questions related to their confidence while assessing student's spoken skills and the professional training in spoken assessment they would like to get in the future. All interviews were audiotaped and later transcribed (see Appendix 7: Teacher sample interview). Interview data were analyzed via a constant comparison process (Farhady, 1995).

¹"Interviews probably provide the most valid sort of data about a phenomenon" (Farhady, 1995, p.220).

3.4 Procedures

A scoring rubric checklist, teacher and student questionnaires, taped teacher interviews were used as data collection instruments in this research. They may be regarded as separate observations conducted at different points in time.

3.4.1 Procedure used with students

a) Rubrics

The scoring rubric checklist that served as the focus of this study was given to the 48 IEP student-participants at the beginning of December, which was the last month of instruction.

The scoring rubric checklist was distributed to participants in order to better familiarize the students with the notion of scoring rubric use and to prepare the students for answering the questions in the student questionnaire. During the survey only the AUA/IEP institutional scoring rubric checklist was used, since the teachers preferred to use the form their groups were already familiar with. Students were instructed to use the rubrics at home as a guide while preparing for an upcoming regular presentation. They were informed that the same rubrics would be used by their teacher and themselves to self-assess or to do a peer-evaluation of their presentation during class.

After giving their oral presentation, the students either self-assessed or carried out peer-evaluation of their presentations using the scoring rubric checklist, and turned in that peer or self assessment prior to knowing their grade from the teacher. The teacher of the course did not have access to the student scoring rubric checklist responses or their self/peer assessments. Only 18 students self-assessed their presentations; thirty students did peer evaluation, since some of the teachers decided that peer evaluation is more effective than self-assessment for an oral presentation. For each presenter the whole class did peer

evaluation. To investigate the relationship between the student's self-assessment and the teacher's evaluation grade of the presentation a Bivariate correlation (Pearson) of scores was run.

b) Questionnaires

The student questionnaire was designed to be both closed-ended and open-ended, so that a detailed understanding of students' thoughts and also likes and dislikes could be determined.

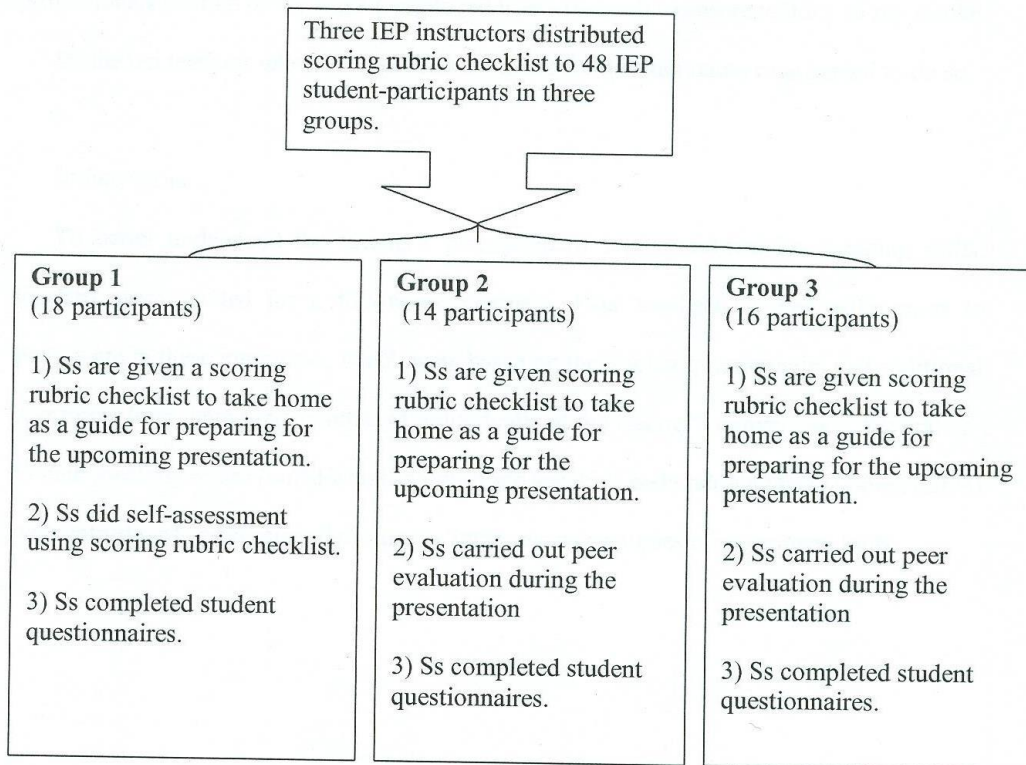
The student questionnaire was piloted with a small group of other students around the same level in order to make certain that the questions were clear and solicited the type of information I wanted to know relating to students' beliefs. The pilot was an important step since it led to changes in the overall format to make it more user-friendly. The font was made larger and participants were provided more room to write open-ended responses. In addition, the format of selected response statements was changed so that it was clearer when students should choose a response and what the choices were for each response. Also, the wording in the questions was made more standard throughout to help the reader understand the questions better.

The same day, when the participants self-assessed or carried out peer-evaluation of their presentations using the scoring rubric checklist, they also completed the student questionnaire. The number of useable questionnaires in this research totaled 48.

The open-ended questions of the student questionnaire provided qualitative data whereas the closed ended questions of the student questionnaire yielded interval scale quantitative data. The open-ended questions are worded in such a focused way that the question could be answered succinctly, with a 'short answer'- that is usually more than a phrase and less than a paragraph (Dornyei, 2003). The data from the open-ended questions

were categorized, coded, and their frequencies were counted. These “items were processed by means of some systematic ‘content analysis’, whereby the pool of diverse responses was reduced to a handful of key issues in a reliable manner” (Dornyei, 2003, p. 117). Each person’s response was taken in turn and distinct content elements, substantive statements, or key points was marked in them. Then based on these ideas and concepts highlighted in the texts, broader categories were formed to describe the content of the response in a way that allows for comparisons with other responses.

Figure 1: Flow Chart of procedure used with students



3.4.2. Procedure used with teachers

a) Questionnaires

In order to collect some baseline data and to understand where the teachers were with their use of and perceptions on the scoring rubrics for oral assessment, the teachers filled in a questionnaire.

The teacher questionnaire was reviewed by seven colleagues to get their perspective for any needed changes. Having the questionnaire reviewed by colleagues was also important since this gave me a professional perspective from other educators with similar and more experience in the classroom, in addition to perspectives of teachers from different levels and backgrounds. Piloting the questionnaires and getting insight through interactions with other professionals are two of the ways I employed in my research to ensure validity of my results.

Of the ten teachers who were asked to complete the questionnaire, nine agreed to do so.

b) Interviews

To better understand the teacher's perception of scoring rubrics for speaking skills, teachers were invited for a follow-up interview. Four teachers showed willingness to participate in these interviews, which were based on the teacher questionnaire and additional questions. Interviews were conducted individually at the teacher's convenience. Several up-to-date speaking scales (holistic and analytic) and speaking tasks adopted from Luoma (2004) were introduced and discussed during the interview as examples of assessment tools.

Chapter 4: Results

Introduction

The teacher and student questionnaires were analyzed using descriptive statistics (percentages). Interview data were analyzed via a constant comparison process.

4.1 Analysis of the Student Questionnaire

4.1.1 Personal information

The first part of the questionnaire (questions 1 through 5) elicited personal information from the respondents. The majority of the students (58.3%) were between 20-25 years old. Most were female (77.1%). The most frequently occurring TOEFL score for all participants was 517 (see Appendix 8 for details).

4.1.2 Use of rubrics

The sixth question on the survey addressed the use of the rubrics as the students were preparing for the presentation. The results in Table 1a show that approximately 75% of the respondents used the rubrics sometimes or frequently as they were preparing for their presentation.

Tables 1-3 summarize the results of questions 6-12 of the student questionnaire.

Table 1a: Use of rubrics

Question	1 Very frequently	2 Sometimes	3 Seldom	4 Not really	5 Not at all	Missing
6. Used the rubrics when preparing for the presentation	11 (23%)	25 (52%)	9 (19%)	1 (2%)	2 (4%)	-

N = 48

For question seven of the questionnaire that asks about the help rubrics provided, the majority of respondents (83%) found that rubrics helped them very much or to some extent to prepare for their presentation (see Table 1b). It is worth pointing out here that only 8% of respondents felt they did not benefit from having the rubrics while preparing for their presentation. The respondents were asked to indicate the extent to which the use of rubrics ensures that the way they are graded is fair (question 8). A total of 83% percent of respondents indicated either very much or to some extent. Over 92% of respondents agreed that rubrics helped them to know what they were supposed to do (question 9). Ninety-six percent of respondents responded to question 10 positively, which asked whether the rubrics helped them understand how their teacher would assess their performance. And only 4% were not sure whether rubrics helped them to understand how the teacher assesses their performance. While 94% of these respondents felt that rubrics helped them know what they needed to work on, 6% thought that they were not sure (Table 2).

Table 1b: Use of rubrics

Question	1 Very much	2 To some extent	3 Not sure	4 Not really	5 Not at all
7. Rubrics helped to prepare better	12 (25%)	28 (58%)	4 (8%)	2 (4%)	2 (4%)
8. Rubrics ensure that the way Ss are graded is fair	12 (25%)	28 (58%)	7 (15%)	1 (2%)	-
9. Rubrics helped the student to know what the student was supposed to do	21 (44%)	23 (48%)	4 (8%)	-	-
10. Rubrics helped the student to understand how the teacher assesses Ss' performance	25 (52%)	21 (44%)	2 (4%)	-	-
11. Rubrics helped the student to know what the student needed to work on	23 (48%)	22 (46%)	3 (6%)	-	-

N = 48; No missing response

Table 2 presents analysis of the results to question 12, which asked students whether they knew what grade they would receive from the teacher. Surprisingly, the answers to this question indicated that the majority of the respondents (75 % of students) were not sure.

Table 2: Understanding the grading system

Question	1 Definitely	2 Not sure	3 Definitely not	Missing
12. I know what grade I will receive from the teacher	8 (17%)	36 (75%)	3 (6%)	1

N = 48

4.1.2 Positive and negative aspects of using rubrics

Students were asked to give short answers to the three questions related to their perceptions of positive and negative aspects of using the rubrics (see Appendix 9 for details).

In response to the open-ended question (13) on what students liked about having the rubrics while preparing for their presentation, recurrent themes were:

(Number in brackets notes frequency of responses.)

- Liked everything (6)
- Rubrics helped and guided in organizing the presentation (25)
- Rubrics clarified grading system (15)
- Disliked everything (1)

Typical comments in relation to having the rubrics while preparing for their presentation were:

“Rubrics helped me to organize and plan my presentation well, as I concentrated on general points important to the presentation.” [S 12]

“I knew the aspects that I must work on to get an organized presentation.” [S 20]

“Rubrics are like a guide providing information about general requirement of presentation.” [S 37]

“It completely clarifies what someone has to do to prepare a presentation, in what direction to work on and to which points to pay attention to.” [S 4]

“Through rubrics I knew what the teacher expected of me”. [S 41]

Among the most frequently cited comments about grading benefits while using rubrics were:

“What I liked more about rubrics is that it helps students to know in advance how the presentation will be graded...” [S 2]

“Rubrics clarify the process of assessment”. [S 44]

“Using rubrics makes it easy to evaluate peers, as well as to do self-assessment.” [S 36]

“I liked the criteria through which I was graded.” [S 23]

Though one of the students commented:

“...It is just a form, nothing more; a form which helps students to understand how the instructor will evaluate them.” [S 19]

When asked about whether there was anything students disliked about having the rubrics given to them (question 14: *What, if anything, did you dislike about having the rubrics given to you?*) specific comments made by the students related to the following:

- Liked everything (28)
- The problem of using the criteria (8)
- Evaluation by students (4)

It was noted that rubrics have “limitations”, since they require that students observe general points, as for example “to keep an eye contact”, “speak in a natural manner without memorization”. According to some respondents these can be restrictive during the evaluation

or self-assessment process, thus making self-evaluation through scoring rubrics rather difficult; an example of a comment was:

“...it’s difficult to evaluate somebody taking into account only given points in the rubrics.”[S 32].

There was also 1 opinion expressed that the use of rubrics sometimes does not assure that the way students are graded is fair.

The fifteenth question related to preference of being assessed either through scoring rubrics, called in the questionnaire “new way of assessment”, or through the way they have been assessed in the past prior to coming to AUA, called “traditional way of assessment” (see Appendix 9). Students commented here on two ways of assessment.

- New way better (32)
- Not sure (8)
- No difference (1)

The majority of participants indicated the “new way”, explaining that it gives not only a good understanding of the grading system the instructor uses, but also familiarizes students with the principles of doing a good presentation leading to the best grades. “In one list you can easily see all positive and negative aspects of your speaking skills during presentation”, indicated one of the respondents [S 8]. “Being graded this way helps to determine weak and strong areas of own speaking skills”, said another participant [S 11], and “...you can find out the reasons why you get this or that grade” [S 5]; “...and it is more detailed, consequently fairer” [S 39]. A fifth participant [S 44] noticed: “I think that this way of being graded is better than the traditional way used in the past, because it helps to realize the mistakes, and never repeat them in the future”. Another student indicated that the ‘new way’ is better as it provides the students with all the points they are going to be evaluated, “it makes the evaluation transparent” [S 16]. None of the respondents mentioned that they would like to be

assessed in a “traditional way”. Eight students were not sure whether being graded through scoring rubrics is better than through the way the student was graded in the past. They mentioned that being graded through scoring rubrics was not very familiar to them, but they thought maybe it would be better than the “traditional way” of assessment. One of the students answered: “I can’t answer because of lack of previous experience of being graded through rubrics.” [S 19]

4.14. Self-assessing speaking abilities

Tables 3-8 refer to questions 16-21 on the student questionnaire.

As can be seen from Table 3, some students (86%) consider that their entrance TOEFL score reflects their speaking ability in English either not at all or to some limited extent. Ninety-six percent of respondents found that their speaking skills had improved since they started to attend the IEP courses at AUA (see Table 4). When asked about whether the teacher gives feedback after the students complete a speaking task, according to Table 5, 94% of students indicated that she sometimes, usually or almost always does.

Table 3: TOEFL score and speaking ability in English

Question	1 Not at all	2 To some extent	3 Accurately	4 Very accurately	Missing
16. To what extent do you think your entrance TOEFL score reflects your speaking ability in English?	23 (48%)	18 (38%)	4 (8%)	1 (2%)	2

N = 48

Table 4: Influence of IEP on the speaking skills

Question	1 Not at all	2 To some extent	3 Well enough	4 A great deal	Missing
17. To what extent do you think your speaking skills have improved since the beginning of IEP?	1 (2%)	13 (27%)	23 (48%)	10 (21%)	1

N=48

Table 5: Teacher's feedback after a completed spoken task

Question	1 Almost never	2 Sometimes	3 Usually	4 Almost always	Missing
18. Does your teacher give you a feedback after you have completed any speaking task?	2 (4%)	5 (10%)	19 (40%)	21 (44%)	1

N=48

The results for self-assessing students' current speaking abilities of English language are presented in Table 6. About 61% of students rated their speaking abilities as advanced (high, mid or low), while 37% of the students rated their speaking abilities as either intermediate high or mid. Question 20 related to how the student improves his/her speaking skills in English (Table 7). While 40% of respondents were in favor of practicing on their own, 27% of students either take private lessons or enroll in a speaking course. Another 10% practice with a friend to improve their speaking skills. In total, 92% of the respondents indicated that they seek out opportunities to talk with native speakers of English to improve their speaking skills sometimes, usually, or almost always (see Table 8).

Table 6: Self-assessing speaking abilities

Question	1 Advanced high	2 Advanced mid	3 Advanced low	4 Intermediate high	5 Intermediate mid	Missing
19. How would you rate your own speaking abilities in English?	1 (2%)	18 (38%)	10 (21%)	16 (33%)	2 (4%)	1

N = 48

Table 7: Ways of improving speaking skills in English

Question	1 Take private lessons	2 Enroll in speaking course	3 Practice on my own	4 Practice with a friend	Other	Missing
20. What do you do to improve your speaking skills in English?	1 (2%)	12 (25%)	19 (40%)	5 (10%)	8 (17%)	3

N = 48

Table 8: Seeking out opportunities to talk with native speakers

Question	1 Almost never	2 Sometimes	3 Usually	4 Almost always	Missing
21. Do you seek out opportunities to talk with native speakers of English to improve your speaking skills?	3 (6%)	13 (27%)	18 (38%)	13 (27%)	1

N = 48

4.2 Analysis of the Teacher Questionnaire and Interview

The questionnaire and interview sought teachers' views in the assessment of students' speaking skills using scoring rubrics.

4.2.1 Questionnaires

To elicit teachers' perceptions of scoring rubrics for speaking skills, their views on several issues, such as discussing and developing scoring rubrics with students, determining assessment criteria, choosing and selecting criteria, using jargon-free language to describe criteria, making the assessment manageable, to mention only a few, were explored through the questionnaire.

Table 9 shows that 88.9% of the teachers either strongly agree or agree with discussing or developing the specific criteria with students before beginning the assessment. Furthermore, all the respondents (100%) strongly agreed or agreed with statement two: *Determining assessment criteria from curriculum objectives or student needs*. Reactions to statement 3 that is to preventing vagueness and increasing objectivity through choosing criteria that are easily observed were again favourable (88.9%), as it can be seen from Table 9, although there was one teacher who was uncertain about this. There was evidence that the teachers (100%) appreciated the opportunity to select criteria that students have had the opportunity to practice. The majority of teachers (88.9%) prefer not to use jargon while describing criteria. According to data based on statement 6, which is to make the assessment manageable, it became clear that 77.7% preferred a) keeping the number of criteria to less than, for example nine and 88.8% preferred b) limiting the number of students observed to a few at any one time. It is worth mentioning that 88.9% (see Table 9) either strongly agree or agree with using or adapting rating scales and rubrics (statement 7). It was clear from the results of the teacher questionnaire that the majority of teachers (77.8%) were in favor of

using numbered continuums to measure the degree to which students are successful at accomplishing a skill or activity, although 22.2% were uncertain about this. As for statement nine of Table 9, discussing the use of rubrics when the observation calls for a holistic rating scale, again the responses were strongly favorable (100%).

Table 9: Teacher Questionnaire Analysis

When undertaking assessment it is advisable to	1 Strongly Agree	2 Agree	3 Uncertain	4 Disagree	5 Strongly disagree
1. Discuss or develop the specific criteria with students before beginning the assessment	7 (77.8%)	1 (11.1%)	1 (11.1%)	0%	0%
2. Determine specific assessment criteria from curriculum objectives, components of a particular activity, or student needs	7 (77.8%)	2 (22.2%)	0%	0%	0%
3. Choose criteria that are easily observed	5 (55.6%)	3 (33.3%)	1 (11.1%)	0%	0%
4. Select criteria that students have had the opportunity to practice	4 (44.4%)	5 (55.6%)	0%	0%	0%
5. Use jargon-free language to describe criteria	7 (77.8%)	1 (11.1%)	1 (11.1%)	0%	0%
6. Make the assessment manageable by a) keeping the number of criteria to less than, for example nine (as it is in the IELTS rating scale); b) limiting the number of students observed to a few at any one time	3 (33.3%)	4 (44.4%)	2 (22.2%)	0%	0%
	4 (44.4%)	4 (44.4%)	1 (11.1%)	0%	0%
7. Use or adapt rating scales and rubrics from other sources	2 (22.2%)	6 (66.7%)	1 (11.1%)	0%	0%
8. Use numbered continuums to measure the degree to which students are successful at accomplishing a skill or activity	2 (22.2%)	5 (55.6%)	2 (22.2%)	0%	0%
9. Use rubrics when the observation calls for a holistic rating scale	5 (55.6%)	4 (44.4%)	0%	0%	0%

N=48; No missing responses

4.2.2. Interviews

To gain more in-depth understanding of the teachers' responses to the questionnaires, the teachers were interviewed. In the semi-structured interview, the order of questions in the questionnaire was followed, but the teachers were encouraged to explain their views. Generally speaking, the teachers participating in the study appeared to have positive attitudes towards scoring rubrics for assessing speaking skills. Various perspectives emerged during the interviews.

Question 1:

When undertaking assessment it is advisable to...

Discuss or develop the specific criteria with students before beginning the assessment.

There was evidence that teachers sometimes appreciated the opportunity to discuss or develop the specific criteria with students before beginning the assessment. Most teachers (3) reported that in practice they give students developed criteria and then discuss these criteria together. They think it is necessary for the students to know what the teacher expects from them, how they are going to be evaluated.

Question 2:

Determine specific assessment criteria from curriculum objectives, components of a particular activity, or student needs.

A typical comment in relation to the determining specific assessment criteria was:

“When students come to AUA, we usually speak about their objectives at this institution; the objectives and aims that the institution usually pursues ...And within the framework of the discussion of the objectives of their studies at AUA, we speak about the necessity for them; to know how they're going to be evaluated, and criteria according to which they'll be evaluated later. And that we also proceed from the student's needs definitely because some students generally when they come, they are not prepared for making presentations...” [T 3]

Question 3:

Choose criteria that are easily observed.

Remarks from teachers indicated that it is very important to have specific criteria. According to them, for students, who come from an educational background where they are not used to doing presentations, it is extremely important to use criteria to prevent vagueness and increase objectivity by placing and emphasizing in the rubrics the major points.

“...if I put in my rubric a thesis statement as the first one ... I emphasize the importance of this criterion, the importance of starting the speech with a clearly stated thesis statement. ” [T 3]

“I think it’s extremely important to use criteria to prevent ... vagueness and mostly to be objective. Because sometimes I have and this is why rubrics are very important, because you think, “Oh, I can tell he is a good speaker, unfortunately John Carry is not as a good speaker” ... whatever ...you can see that in judging students it is very broad to have sort of she is good, she is bad. We need to look exactly”. [T 4]

Question 4:

Select criteria that students have had the opportunity to practice.

The teachers showed a positive attitude to selecting criteria that students have had the opportunity to practice.

Among the notable citations is:

“...there are no secrets...in any test that I give either oral or written. And if I’m going to give a different kind of test, then I prepare them to take it before they have to take it. ...I don’t want the students coming to me later and say “I didn’t understand the instructions”. To me that’s a waste of their and my time. So, I’ll make sure that everything is understood, we do a model test ...and they understand how to do it. ...And if I can make it easier and I can make it, so it also teaches too them, that it’s productive so much the better”. [T 2]

Question 5:

Use jargon-free language to describe criteria so that data can be used effectively when giving feedback to Ss.

The teachers generally stated that they would not use jargon terms to describe criteria especially depending on the class and the students of the course. A typical comment is:

“The teacher shouldn’t use linguistic jargon with students. It should be free, naturally of jargon, because as it is student has a problem; why double this problem by making it difficult for him by making the learning process difficult”. [T 3]

Question 6:

Make the assessment manageable by:

a) keeping the number of criteria to less than, for example nine (as it is in the IELTS rating scale);

Interviews showed, particularly in relation to the making the assessment manageable, that some of the teachers think a teacher should never try to look at too many things at once. Some of the interviewees find that nine criteria may not be enough or it may be too many. As it was stated “... it all depends on the focus”. [T 1]

b) limiting the number of students observed to a few at any one time.

In general, teachers commented that while assessing students’ speaking skills in any form, they do not like to assess more than 4 students at one time, since it becomes very tiring for students and for the teacher, too.

Question 7:

Use or adapt rating scales and rubrics from other sources, such as those in the literature.

Reactions to the adapting rating scales and rubrics were favourable. A typical comment related to this issue was:

“... with the emphasis on adapt. There are things that I choose to ignore in a specific time. There are things that I give more weight to in a specific instance”. [T 4]

Question 8:

Use numbered continuums to measure the degree to which students are successful at accomplishing a skill or activity.

The majority of teachers (3) were negative about numbers, though they use them, since the institutions require grades.

“...because, let’s say the course or institution or even a program requires... you should follow certain scales or ratings depending on what is required of us. But I personally, don’t like, I prefer more holistic approach to grading and I don’t really like numbers, you know, because language is kind of complicated and you cannot assess everything at the same time and give it just one number. ...” [T 1]

“I hate putting numbers on people; “You’re 2; you’re 5...” because of this physiological baggage that goes... And then I don’t automatically know if this is necessary. I mean we do it. It’s necessary I guess from a practical point of view. But I don’t quite know it is necessary... Anyway, we have to use them and I have mixed feelings about them”. [T 4]

“...I don’t tell the students: “You got a 97 or a 96. I never give them a number. I say: “You have an A” and an A means you have successfully completed A, B, C, D ...As I said, this, this and this. But to the college and university, they want to see the numerical numbers.” [T 2]

But there was also a different point of view about grades:

“...students should know and the students want to know whether they’re improving or not. How can they show this if not by continuum, by giving them a kind of grading? That shows them a narrow-marrow of their own progress through this grading system”. [T 3]

Question 9:

Use rubrics when the observation calls for a holistic rating scale, since rubrics describe the attributes of student knowledge, or on a numbered continuum of possibilities.

Almost all of the teachers showed preference for a holistic scale.

The interviews showed, particularly in relation to the additional questions (see Appendix 6), that though the instructors feel confident, while assessing student’s speaking skills, they would willingly take a course or participate in some type of program, in a seminar or in a workshop. They think they would benefit from getting training in the field of spoken language testing. A typical comment is: “I won’t have any problems in learning new things”.

Typical comments in relation to the possibility of increasing public understanding and support of spoken language testing were: first of all it is necessary to increase teachers' understanding in the field of spoken assessment. "...I don't think public knows very much about second language teaching in general". [T 3]

In addition, the participants showed willingness to cooperate with other groups interested in spoken language testing.

4.3 Correlation of the student self-assessment and teacher evaluation grade for the presentation

As it was mentioned in chapter 3 (Method), only 18 students out of 48 students did self-assessment, the others (30 students) did peer assessment. The correlation study that investigated the relationship between the students' self-assessment and teacher's evaluation grade yielded the following results. The observed correlation coefficient (Pearson's r) is 0.44, which is very low, and the overlap between the two measures is 20% (for more details see Appendix 10). While analyzing the data, it was clear that 7 students' (39%) self-assessment grades were underestimated by -1 compared with the teacher's evaluation grade.

Table 10: Students' grades compared with those of the teacher

Grade	N	-1	Total
Underestimated	7 (39%)	7	N=18
Exact	11 (16%)	0	-
Overestimated	0	0	Missing response

Chapter 5: Discussion and Conclusions

Introduction

This study set out to address the following questions.

- 1) What are teachers and students perception of the use of rubrics as a speaking and assessment aid?
- 2) Can students use rubrics to self-assess accurately?

The overall picture that emerges from the findings is that teachers and students have a positive perception of the use of rubrics as a speaking and assessment aid; yet some of the students were unable to self-assess accurately using rubrics. These findings, as well as their implications, will be discussed in detail and the conclusions and recommendations for further research will be presented.

5.1 Teacher and Student Perception of Scoring Rubrics

The percentage study revealed that about 94 % of students used the rubrics to prepare for their presentation and 83% think that the rubrics helped them to prepare better (see Chapter 4, Table 1a & Table 1b). The results of this study are consistent with some of the research findings and discussions in the literature on scoring rubrics (Chapter 2). Many researchers (Andrade, 2000b; McDaniel, 1994; Moskal, 2000; Tierney & Simon, 2004) argue that a rubric is a scoring guide that seeks to evaluate a student's performance based on the sum of a full range of criteria rather than a single numerical score. It is a working guide for students and teachers, usually handed out before the assignment begins in order to get students to think about the criteria on which their work will be judged (Stix, 1997). The

participants (95.9%) liked rubrics mostly for the way in which rubrics were able to guide them while preparing for the presentation (see Chapter 4).

As Andrade (2000b) states, when rubrics are used to lead self- and peer-assessment, students become more competent to notice and correct problems in their own works and in one another's work, too. However, a few students (16.7%) felt that rubrics have some limitations because of their limited criteria (see Chapter 4). In general, students (66.7%) preferred the new way of assessment through scoring rubrics (see Chapter 4).

Andrade argues that rubrics appeal to teachers and students for several reasons (2000b). First of all they are powerful tools for both teaching and assessment, which describe to students what is expected. Rubrics also can assist in improving student performance in addition to assessing it. The best argument for using rubrics is that they often result in improvement in the quality of student work. Teachers of this study confirmed that rubrics help students to be aware of their own progress, which is a path towards autonomous learning. According to the teachers (see Appendix 7), when students do self-assessment from the very beginning of the course, they become intellectually and morally independent in their learning.

As the interviews revealed, teachers have to overcome many things while assessing students' oral performances. However, when they use rubrics to evaluate the students' oral speech, they try to be objective and to follow the rubrics. An opinion expressed by Starr (2000) is that rubrics offer teachers objective methods to evaluate students' works. Besides Andrade (2000b) mentions that teachers appreciate rubrics because they lessen the amount of time teachers spend evaluating student work. Rubrics provide students with more instructive comment about their strengths and areas in need of improvement.

Other data from this study, completed teacher questionnaires, and teacher interviews, indicate high levels of agreement among the surveyed teachers concerning their perception of

and use of scoring rubrics, which carry implications for their identity as committed professionals. There was overwhelming agreement that the teacher should discuss or develop the specific criteria with students before beginning the assessment. They should also determine specific assessment criteria from curriculum objectives, components of a particular activity, or student needs and select criteria that students have had the opportunity to practice. On the other hand, considerably less certainty and more disagreement were reported regarding using numbered continuums to measure the degree to which students are successful at accomplishing a skill or activity. A thought-provoking finding in this study is that though the teachers feel confident, while assessing student's speaking skills, they would not mind if there were an opportunity to participate in some type of program, where they could get training in the field of spoken language testing.

5.2 Involvement of Students as Participants in the Assessment Process

The correlation study that investigated the relationship between the student's self-assessment grade of the presentation and the teacher's evaluation grade of the same presentation show that some of the students underestimated their speaking abilities of English while using scoring rubrics during the presentation. This may be for the following reasons. Maybe they had limited experience of assessing themselves or their peers through scoring rubrics in their educational institutions before coming to AUA. No assessment tool is effective if it is not used on a regular basis. Rubrics are most effective when we practice using them with our students over and over again. "Negotiable contracting is a new approach to involving students in the assessment process" (Stix, 1997). Students themselves can be involved in the assessment process through both peer and self-assessment. As students become familiar with rubrics, they can assist in the rubric design process, since "this

involvement empowers the students and as a result, their learning becomes more focused and self-directed” (Upbin, 1999, p.3). As discussed earlier, there is evidence that the art of negotiable contracting consists of giving students shared ownership in their own learning (Wiggins, 1993).

The respondents’ attitude and approach to the scoring rubrics and self-assessment procedures may also have influenced their answers. It is also possible that the students might have assessed themselves not in absolute terms but relative to their peers. The students might have assessed themselves based on their effort and participation in class. Maybe some of them thought that their self-assessment grade might influence their teacher’s evaluation of them. The other reason that some of the students did not self-assess themselves correctly might be because of the problem within the criteria.

The results of this study should encourage EFL teachers to adapt Upbin’s main pragmatic argument, in terms of looking for a certain level of readiness for active student involvement in the assessment process through both peer and self-assessment in order to give students shared ownership in their own learning. Richardson (2001) observes that teachers who involve their students in the creation of rubrics that will be used to score student work have found the student participation very helpful. He argues that the discussion guiding to the final rubric also helps students make clear the expectations for their work.

5.2 Conclusions and Implications

The study shows that in general students and teachers have positive attitudes towards the use of scoring rubrics. The analysis of the results of the teacher questionnaire and interview revealed the possible ways in which many aspects of this research may be further investigated. Since this is the first research related to the issues of scoring rubrics for

speaking skills done in Armenia, it can be regarded as an exploratory study on which further research can be based.

In concluding this study, I wish to highlight three salient issues, which have emerged. These issues carry implications for teaching and learning.

Firstly, it is apparent that the rubrics are powerful tools for both teaching and assessment, which can improve student performance.

Secondly, as it was revealed through the analysis of the open-ended questions of the student questionnaires, rubrics are useful since they help students become more thoughtful judges of the quality of their own and others' work. As the interviewed teachers stated, peer-assessment, and especially self-assessment, is likely to increase students' sense of responsibility for their own work.

Lastly, findings show that teachers are not against teacher training in assessment. Teacher training should therefore give due emphasis to aspects of language assessment. Such professional preparation would help equip teachers with the competence they need to effectively and confidently carry out their duties. Competence and confidence will invariably enhance perceptions of their professionalism.

This research showed that scoring rubrics may be a positive addition to a teacher's methods as they can provide a meaningful way to plan for and interpret student learning; and that the students need training in using scoring rubrics while preparing for a specific speaking task.

A practical implementation of this study may be familiarizing the English language teaching community with these findings by means of workshops conducted in the Association of English Language Teachers of Armenia (AELTA) within the framework of the Teacher Development Program.

5.3 Limitations of the Study and Future Directions

There are some limitations to the present study that should be considered. First, the study is not large-scale: it included only 48 students. Thus, the findings of this research are limited to AUA and its results cannot be generalized to other educational institutions. AUA provides a unique learning environment due to the facts that the language of instruction is English, modern technology is widely used in the language learning process, and the IEP students already have BA's from different educational institutions in Armenia.

Second, students may not be familiar with scoring rubric checklists and questionnaires, as such scoring rubric checklists and scoring rubrics are not commonly used in the Armenian educational system.

Third, the student questionnaire used in this research may have resulted in certain limitations: the student questionnaire was long and towards the end the respondents may have got tired and may not have given equal attention to all the questions.

Finally, regarding the teacher interview, I could have planned to elicit necessary, and at the same time better focused information. Now I realize that attempts should have been made to detect and identify unclear responses and to design an appropriate way of clarifying the responses. I should also have avoided any sort of leading questions, guiding to desired responses. Further, I could have been more consistent in eliciting information from the participants.

In future research, it would be interesting to investigate the possible involvement of students as participants in creation of scoring rubrics.

Another interesting direction of further research might be to investigate the possible impact of student self-assessment through scoring rubrics on teaching techniques, classroom activities and curriculum development. This will enable the researcher to show the positive role of scoring rubrics in providing students with standards and expectations they can use to

evaluate their performance while completing an assignment. Such scoring rubrics also make students sensitive to their own needs instead of relying entirely on their teacher's opinion (Mertler, 2001; Moskal, 2000 & 2003; Simon & Forgette-Giroux, 2001; Tierney & Simon, 2004).

There is clearly room for more extensive research in the local setting and other settings, including similar studies on a larger scale. Accordingly, my discussion of the present findings remains essentially exploratory in spirit. In spite of this I consider that my study helped us to better understand the criteria and the use of scoring rubrics in the field of the oral assessment.

References

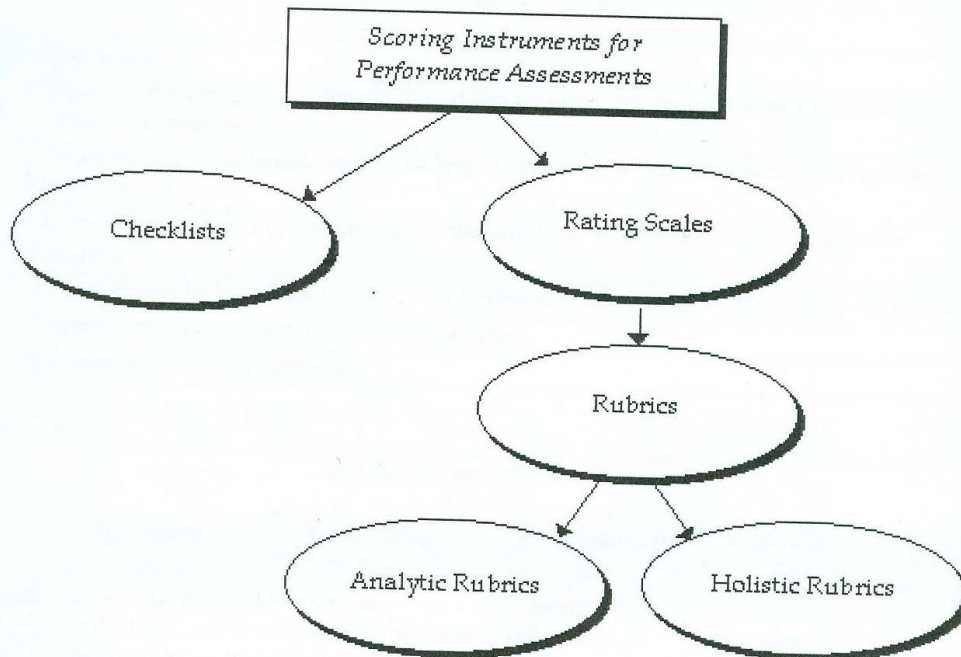
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). Standards for Educational and Psychological Testing. Washington, DC: Author.
- Andrade, H. G. (2000a). Using rubrics to promote thinking and learning. *Educational Leadership*, 57, 5. Retrieved January 22, 2005, from <http://www.education.tas.gov.au/ocll/currccons/profreadings/andrade.htm>
- Andrade, H. G. (2000b). Understanding rubrics. Retrieved December 18, 2004, from <http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>
- Arter, J. & McTighe, J. (2001). Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Thousand Oaks, CA: Corwin Press/Sage Publications.
- Assessment Resource Library. (1998) Bibliography of Assessment Alternatives: Oral Communication. Retrieved February, 2005 from <http://www.nwrel.org/assessment/pdfLibrary/oral.pdf>
- Brookhart, S. M. (1999). The art and science of classroom assessment: The missing part of pedagogy. ASHE-ERIC Higher Education Report, 27 (1). Washington, DC: The George Washington University, Graduate School of Education, Human Development.
- Dornyei, Z. (2003). Questionnaires in second language research: Construction, Administration, and Processing. Mahwah, New Jersey & London. Lawrence Erlbaum Associates, Publishers.
- Farhady, H. (1995) Research methods in applied linguistics. Teheran. Payamnour Univ.
- Frechtling, J., Hansen, J., Zhang, X. & Nyre, G. (2002). ABC on Scoring Rubrics Development for Large Scale Performance Assessment in Mathematics and Science. Retrieved March 8, 2005, from <http://lsc-net.terc.edu/do.cfm/paper/8315>

- <http://www.ed.gov/pubs/OR/ConsumerGuides/perfasse.html>
- http://www.relearning.org/resources/PDF/rubric_sampler.pdf
- Luoma, S. (2004). *Assessing speaking*. Cambridge. Cambridge University Press.
- Mantero, M. (2002). *Evaluating Classroom Communication: In Support of Emergent and Authentic Frameworks in Second Language Assessment*. *Electronic Journal: Practical Assessment, Research & Evaluation*, 8(8). Retrieved May, 2005 from <http://pareonline.net/Articles.htm>
- McDaniel (1994). *Rubrics*. Monmouth University. Retrieved May 30, 2005, from <http://its.monmouth.edu/FacultyResourceCenter/rubrics.htm>
- McNamara, T. (2000). *Language testing*. NY. Oxford University Press.
- Mertler, C. A. (2001). *Designing scoring rubrics for your classroom*. *Electronic Journal: Practical Assessment, Research & Evaluation*, 7(25). Retrieved March 4, 2005 from <http://pareonline.net/Articles.htm>
- Moskal, B. M. (2000). *Scoring rubrics: what, when and how?* *Electronic Journal: Practical Assessment, Research & Evaluation*, 7(3). Retrieved February 7, 2005 from <http://pareonline.net/Articles.htm>
- Moskal, B. M. & Leydens, J. (2000). *Scoring rubric development: validity and reliability*. *Electronic Journal: Practical Assessment, Research & Evaluation*, 7(10). Retrieved February 26, 2005 from <http://pareonline.net/Articles.htm>
- Moskal, B. M. (2003). *Recommendations for Developing Classroom Performance Assessments and Scoring Rubrics*. *Electronic Journal: Practical Assessment, Research & Evaluation*, 8(14). Retrieved May 13, 2005 from <http://pareonline.net/Articles.htm>
- Mueller, J. (2003). *Mueller's Glossary of authentic assessment terms*. Retrieved January 8, 2005, from <http://jonathan.mueller.faculty.noctrl.edu/toolbox/glossary.htm>

- National Communication Association (2005). Criteria for Assessment: Criteria for Assessment of Oral Communication. Retrieved May, 2005 from <http://www.natcom.org/Instruction/assessment/Assessment/CriteriaAssessment.htm>
- Nitko, A. J. (2001). Educational assessment of students (3rd ed.). Upper Saddle River, NJ: Merrill.
- Nunan, D. (1993). Task-based syllabus design: selecting, grading and sequencing tasks. In S. Popham, J. (1997). What's wrong and what's right with rubrics. Educational Leadership, 55, 72-75.
- Richardson, E. (2001). Scoring performance assessments: checklists, rating scales and rubrics. Retrieved January 18, 2005, from <http://www.web.utk.edu/~mccay/apdm/scoring/scoring.pdf>
- Shaaban, K. (2001). Assessment of Young Learners. Forum. 39-4. Retrieved April 19, 2005, from <http://exchanges.state.gov/forum/vols/vol39/no4/p16.htm>
- Shepard, L.A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), Handbook of research on teaching (4th ed), (pp.1066-1101). Washington, D.C.: American Educational Research Association. <http://www.colorado.edu/education/faculty/lorrieshepard/assessment.html>
- Simon, M. & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. Electronic Journal: Practical Assessment, Research & Evaluation, 7(18). Retrieved October 20, 2004, from <http://pareonline.net/Articles.htm>
- Starr, L. (2000). Creating rubrics: tools you can use. Retrieved December 19, 2004, from http://www.education-world.com/a_curr/curr248.shtml
- Stix, A (1997). Creating rubrics through negotiable contracting and assessment. Retrieved January 12, 2005, from http://www.interactiveclassroom.com/articles_006.htm

- Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Electronic Journal: Practical Assessment* Practical Assessment, Research & Evaluation, 9(2). Retrieved February 27, 2005 from <http://pareonline.net/Articles.htm>
- UCELS (2003). IELTS bulletin. British Council, Information for Candidates.
- Upbin, B. (1999). Instant Feedback In The Classroom. *Forbes Magazine*, Retrieved January 14, 2005 from <http://kancrn.kckps.k12.ks.us/Harmon/breighm/rubrics.html>
- Weir, C. (1993). *Understanding and developing language tests*. UK, London. Prentice Hall International Ltd.
- Wiggins, G. (1993). *Assessment: authenticity, context and validity: Assessing Student Performances*. *Phi Delta Kappan*, 75, 3, pp. 200-214. San Francisco: Jossey-Bass Publishers.

Appendix 1: Types of scoring instruments for performance assessments
Figure 1:



(Adopted from Mertler, 2001, without pagination).

Checklists

Applies to basic speech skills students might be expected to demonstrate in making a three-minute presentation to the class.

The student:

- _____ maintains eye contact with the audience
- _____ speaks loudly enough to be heard in all parts of the room
- _____ enunciates clearly
- _____ stands up straight (does not shift from foot to foot)
- _____ does not go over the allotted time
- _____ has notes
- _____ uses notes sparingly

Rating Scales

How often does the student... Never Seldom Sometimes Usually Always
 Turn in lessons on time

(Adopted from Richardson, 2001-2003)

Appendix 2: Scoring Rubrics

Table 1:

<i>Template for Holistic Rubrics</i>	
Score	Description
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

Table 2:

<i>Template for analytic rubrics</i>					
	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Criteria #1	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #2	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #3	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #4	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	

(Adopted from Mertler, 2001, without pagination).

Appendix 3: Overall Band Scores for IELTS

Band 9 - Expert User
Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.
Band 8 - Very Good User
Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.
Band 7 - Good User
Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
Band 6 - Competent User
Has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.
Band 5 - Modest User
Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.
Band 4 - Limited User
Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.
Band 3 - Extremely Limited User
Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.
Band 2 - Intermittent User
No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty in understanding spoken and written English.
Band 1 - Non User
Essentially has no ability to use the language beyond possibly a few isolated words.
Band 0 - Did not attempt the test
No assessable information provided.

(British Council, Information for candidates, June 2003)

Appendix 4: Scoring Rubric Checklist

FORMAL PRESENTATION EVALUATION FORM

Student _____ Teacher _____

Topic _____ Grade _____

3=Very Good 2=Satisfactory 1=Weak 0=Unacceptable

Delivery

- _____ Maintained eye contact with listeners in all parts of the room
- _____ Spoke loudly and clearly
- _____ Spoke in a natural manner (did not read or memorize)
- _____ Used effective posture, movement, and gestures
- _____ Used appropriate tone

Communicative ability

- _____ Spoke clearly
- _____ Spoke fluently, without too much hesitation or repetition
- _____ Used reasonably accurate grammar and vocabulary
- _____ Used appropriate register

Content

- _____ Fulfilled assignment
- _____ Met time limit
- _____ Developed topic with sufficient reasons, examples, and details
- _____ Referred to sources appropriately

Organization

- _____ Introduced topic effectively
- _____ Organized ideas logically
- _____ Made clear transitions
- _____ Concluded presentation effectively

Discussion

- _____ Responded appropriately to questions
-

Comments...

(Adopted from DEP, AUA, 2004)

Appendix 5: Student Questionnaire

Dear Student,

With the help of this questionnaire, I would like to find out what you think about the way your speaking skills in English are assessed.

Please read each statement attentively before responding to each of them.

The information provided by you will be used only for the purpose of this research and will remain confidential.

Thank you for your cooperation.

Personal Information

1. Your age 20-25, 26-30, 31-40, 41-50, above 50
2. Gender male female
3. How many years have you been learning English? (Including school years and private lessons): 0-1____, 1-3____, 3-5____, 5-10____, more than 10____
4. TOEFL Score _____
5. Have you ever taken IELTS, FCE, BEC or EST? Yes No
If yes, which test and when? _____

Section A. Use of Rubrics

Directions:

Please check () the appropriate box for each item.

6. I used the rubrics when preparing for the presentation.

1. Very frequently 2. Sometimes 3. Seldom 4. Not really 5. Not at all

7. The rubrics helped me to prepare better.

1. Very much 2. To some extent 3. Not sure 4. Not really 5. Not at all

8. The use of rubrics ensures that the way we are graded is fair.

1. Very much 2. To some extent 3. Not sure 4. Not really 5. Not at all

9. The rubrics helped me know what I was supposed to do.

1. Very much 2. To some extent 3. Not sure 4. Not really 5. Not at all

10. The rubrics helped me understand how the teacher would assess my performance.

1. To a very large extent 2. To some extent 3. Not sure 4. Not really 5. Not at all

11. The rubrics helped me know what I needed to work on.

1. Very much 2. To some extent 3. Not sure 4. Not really 5. Not at all

12. I know what grade I will receive from the teacher.

1. Definitely 2. Not sure 3. Definitely not

Section B. *Positive and negative aspects of using rubrics*

Directions:

Please use the spaces provided to write short answers to the three questions below related to the positive and negative aspects of using the rubric. You may write your answers in Armenian, Russian or English.

13. What, if anything, did you like about having the rubrics given to you?

14. What, if anything, did you dislike about having the rubrics given to you?

15. Do you think that being graded this way is better than the way you have been graded in the past?

Section C. Positive and negative aspects of using rubrics

Directions: Please circle the appropriate choice for each item.

16. To what extent do you think your entrance TOEFL score reflects your speaking ability in English?

1. Not at all
2. To some extent
3. Accurately
4. Very accurately

17. To what extent do you think your speaking skills have improved since the beginning of IEP?

1. Not at all
2. To some extent
3. Well enough
4. A great deal

18. Does your teacher give you a feedback after you have completed any speaking task?

1. Almost never
2. Sometimes
3. Usually
4. Almost always

19. How would you rate your own speaking abilities in English?

I. **Advanced High** _____

(Speakers at the **Advanced-High level** perform all **Advanced- level** tasks with linguistic ease, confidence and competence)

II. Advanced Mid ____

(Speakers at the **Advanced-Mid level** are able handle with ease and confidence a large number of communicative tasks)

III. Advanced Low ____

(Speakers at the **Advanced Low** level are able to handle a variety of communicative tasks, although somewhat pausing at times)

IV. Intermediate High ____

(**Intermediate High** speakers are able to converse with ease and confidence when dealing with most routine tasks and social situations of the **Intermediate level**)

V. Intermediate Mid ____

(Speakers at the **Intermediate-Mid level** are able to handle successfully a variety of uncomplicated communicative tasks in straightforward social situations)

VI. Intermediate Low ____

(Speakers at the **Intermediate Low level** are able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations)

20. What do you do to improve your speaking skills in English?

1. Take private lessons
2. Enroll in speaking courses
3. Practice on my own
4. Practice with a friend
5. Other (please specify) _____

21. Do you seek out opportunities to talk with native speakers of English to improve your speaking skills?

1. Almost never
2. Sometimes
3. Usually
4. Almost always

Name (optional) _____
completing the questionnaire.

Thank you very much for

Appendix 6: Teacher Questionnaire

Dear Teacher,

I would like to ask you to help me by answering the following questions concerning **scoring rubrics for speaking skills**. This survey is conducted by a student of MA TEFL, Department of English Programs at American University of Armenia to better understand current assessment practices of spoken English. This is not a test so there are no “right” or “wrong” answers and you do not have to write your name on this questionnaire. I am interested in your personal opinions. Please give your answers sincerely as only this will guarantee the success of the investigation.

Please read each statement attentively before responding to it.

The information provided by you will be used only for the purpose of this research and will remain confidential.

Thank you for your cooperation.

For your information;

What are Rating Scales and Rubrics?

Rating scales record the extent to which specific criteria have been achieved by the student or are present in the student's work. **Rating scales** also record the quality of the student's performance at a given time or within a given process. **Rating scales** are similar to checklists, and teachers can often convert checklists into rating scales by assigning number values to the various criteria listed. They can be designed as number lines or as holistic scales or rubrics.

Rubrics provide a set of scoring guidelines for evaluating student work. **Rubrics** answer the questions: By what criteria should performance be judged? Where should we look and what should we look for to judge performance success? What does the range in the quality of performance look like? How do we determine validity, reliability, and fairness? What score should be given and what does that score mean? How should the different levels of quality be described and distinguished from one another?

Directions: Following are a number of statements. I would like you to indicate your opinion after each statement by putting an 'X' in the box that best indicates the extent to which you agree or disagree with the statement. Thank you very much for your help.

	Strongly Agree	Agree	Uncertain	Disagree	Strongly disagree
	5	4	3	2	1
When undertaking assessment it is advisable to					
1. Discuss or develop the specific criteria with students before beginning the assessment					
2. Determine specific assessment criteria from curriculum objectives, components of a particular activity, or student needs					
3. Choose criteria that are easily observed in order to prevent vagueness and increase objectivity					
4. Select criteria that students have had the opportunity to practice and remember that these criteria may differ from student to student, depending upon their strengths and needs.					
5. Use jargon-free language to describe criteria so that data can be used effectively when giving feedback to Ss					
6. Make the assessment manageable by					
a) keeping the number of criteria to less than, for example, nine (as it is in the IELTS rating scale);					
b) limiting the number of students observed to a few at any one time					
7. Use or adapt rating scales and rubrics from other sources, such as those in the literature					
8. Use numbered continuums ¹ to measure the degree to which students are successful at accomplishing a skill or activity					
9. Use rubrics when the observation calls for a holistic rating scale, since rubrics describe the attributes of student knowledge or achievements on a numbered continuum of possibilities					

¹ A scale of possible points to be assigned in scoring work, on a continuum of quality. High numbers usually are assigned to the best performances: scales typically use 4, 5 or 6 as the top score, down to 1 or 0 for the lowest scores in performance assessment.

Dear Teacher,

To better understand your perceptions of scoring rubrics for speaking skills, I would like to invite you to interview. Please indicate if you are willing to be interviewed by checking the box below, if you are not, please leave the section below blank.

I am willing to participate in the interview.

Name _____

Thank you.

Teacher Interview Questions

Greeting. Please say what you really think.

Do you mind if I record this interview? (Based on the questionnaire questions).

Additional questions

1. Do you feel confident when assessing Ss oral performances?
2. Do you think you would benefit from getting training in the field of spoken language testing?
3. How do you think you could be prepared professionally?
4. What kind of other professional services would you like to be provided in the field of spoken language testing?
5. Do you think it is possible to increase public understanding and support of spoken language testing? How?
6. Would you like to cooperate with other groups interested in spoken language testing?

Appendix 7 : Teacher Sample Interview Tape Script

Interview #3

Interviewer: This interview is based on the questionnaire questions. In the questionnaire you've strongly agreed with developing criteria discussing or developing the specific criteria with students before beginning the assessment. Could you please specify?

Interviewee: In practice actually I give them the developed criteria and then go over it, discuss it together. I think it's necessary for the students to know what is it expected from them by the teacher, how they're going to be evaluated.

Interviewer: With the second statement you have again strongly agreed, which is determining specific assessment criteria from curriculum objectives, components of a particular activity, or student needs.

Interviewee: When students come to AUA, we usually speak about their objectives at this institution; the objectives and aims that the institution usually pursues. And in the framework of this discussion, within the framework of the discussion of the objectives of their studies at AUA, we speak about the necessity for them; to know how they're going to be evaluated, and criteria according to which they'll be evaluated later. And that we also proceed from the student's needs definitely because some students generally when they come, they are not prepared for making presentations. We are speaking about presentation in particular. If we speak about presentation as part of oral speech or for example prepared speech, or something else, they must know what other requirements for them to follow. They must know that if they know what is required from them they will know how to get ready for the task. For instance, if I give them one minute speech, the speech should have the thesis statement, organization, and other things. That's also spoken language, which is determined by certain rules. They have to follow these rules. And they should know the rules and they should know that they are going to be evaluated according to the criteria, necessary for the evaluation of the speech.

Interviewer: So, as far as I understood, you strongly agree that scoring rubrics are necessary especially in the speaking tasks.

Interviewee: Yes. They are necessary, they are useful. Especially if it concerns speeches. Of course in prompt to speeches it is more difficult to put on the rubrics, but again it is possible. But speeches which should be prepared for them rubrics are necessary and I think that these kinds of speeches, which are prepared at that moment, will later contribute to the development of the oral speech in prompt to speeches of the students in future.

Interviewer: So, you think that these rubrics will help them to determine what criteria the instructor waits from them.

Interviewee: That's right. The instructor waits from them criteria according to which they will be evaluated.

Interviewer: Expectations of the teacher?

Interviewee: Expectations of the teacher from them while they speak. And that will help them to organize. I think this is not only speaking. The organization of thoughts, in fact, will also help them later in writing, in anything. I think it's all the skills are integrated, as we know. The development of criteria that we give them helps them to develop skills, which will later be useful in writing, speaking...

Interviewer: Very nice. You've strongly agreed with statement 3, which is choosing criteria that are easily observed in order to prevent vagueness and increase objectivity. What can you say about this statement?

Interviewee: The thing is that this is very important criteria. Our students, who come from an educational background, where they are not used to coming with the thesis statement, with the idea of the thesis statement supporting ideas and so on. It's very important to teach them to start their speech with thesis statement. I think, here, if I put in my rubric a thesis statement as the first one, so they know by putting, placing, it in my rubric as number one, I emphasized the importance of this criteria., the importance of starting the speech with a clearly stated thesis statement.

Interviewer: A few days ago I asked you to distribute to you students one form of the IEP Presentation Evaluation form including scoring rubrics for the purpose of self- assessment mentioned in my study. But you preferred the other one IEP uses.

Interviewee: The thing is that actually students were used to these rubrics. I saw that these rubrics answer the needs of the demands that I as a teacher require from my students. There are more elaborate, whether the thesis statement is clearly stated, speed of delivery and so on. It's kind of more detailed. I guess both of the rubrics are the same. I can't say the one is better than the other. It's just I preferred the more detailed one.

Interviewer: Statement 4: Select criteria that students have had the opportunity to practice. What do you think of this?

Interviewee: The thing is that if the students actually during prepared speeches, which is a step towards in prompt to speech, during these prepared speeches with the students prepared at home, they are being evaluated by their peers and when the student selects one criteria that needs improvement for that student, the student knows his drawbacks. He realizes his weak points. It is not that deliberately, it is then naturally by pointing out the weakness during the presentation through his rubrics. If the student, let's say had very monotonous voice and they mention that by selecting this as a weak point for that student will be formed as a next presentation as a point to pay attention to.

Interviewer: So he will have an opportunity to practice and use it in his further education?

Interviewee: That's right. That's his already problem to practice it or to bear in mind change his tone of voice or indicate whatever and improve what needs to be improved. His peers think so and I will usually join them. We are unanimous and then the selection of the criteria, of the weak point in a student's presentation is that by selecting this criteria naturally, which should be later on be improved. And of course, this selection of criteria will be different. For one it'll be the tone of the voice, for the other will be delivery, for the third will be mistakes and grammar, or vagueness, clarity, anything. Everyone in these criteria will be individual for every student.

Interviewer: That's why you strongly agreed with this statement. What about statement 5; use jargon-free language to describe criteria so that data can be used effectively when giving feedback to Ss.?

Interviewee: In general adhere to the principle that the teacher shouldn't use linguistic jargon with students. It should be free, naturally of jargon, because as it is student has a problem; why double this problem by making it difficult for him by making the learning process difficult. I think that for me it's not a question to be discussed rather than use linguistic jargon in explaining anything to my students.

Interviewer: And what about making the assessment manageable by a) keeping the number of criteria to less than, for example nine (as it is in the IELTS rating scale); b) limiting the number of students observed to a few at any one time.

Interviewee: It's difficult to listen to anybody, especially to the students whose speech is full of flows. And I don't think it's very useful for these students to hear each other's speech; more than four is enough. It's the limit for students for one day. Yes, and number of students should be limited to a few. Not more than four. I mean four is the upper limit. Five is already too much, since students get tired and not all speeches are interesting. Some of them might be boring; the students might fail to control the audience. We don't deal with professional speech makers, orators.

Interviewer: Your groups are doing presentations group by group?

Interviewee: Yes, not to get exhausted. For example, to wait until their turn comes. I usually say to my students: "This is the time allotted for your presentation. When will you do it within this time limit?" They say when they can be ready on a presentation. And usually I don't have more than four in a day.

Interviewer: What about keeping the number of criteria for less, for example than nine?

Interviewee: I agree with this question, because if it's more, if it's too many, it's again makes things more complicated. If you chose the most important points, if you concentrate on the most important it can't be too many. If it's too many, that means you involve also additional things, not the only important ones. And I think if we limit ourselves, refrain ourselves to the most important, than it won't be more than nine. And then it will be easier for the students to understand what is the most important, what is it that they should follow. But when it's too many they might just not pay attention to them.

Interviewer: They'll lose the thread of their train, you mean?

Interviewee: Right.

Interviewer: You prefer analytic scale or holistic scale?

Interviewee: I think holistic.

Interviewer: The majority of lecturers gave preference to a holistic scale.

Interviewee: I do, too, prefer holistic scale.

Interviewer: But do you think that sometimes it depends on the circumstances what kind of scale to use.

Interviewee: I agree with you. It depends on the circumstances. On my own needs I use holistic scale. I think holistic scale answers the needs of what I'm doing in the classroom.

Interviewer: And the last statement; using numbered continuums to measure the degree to which students are successful at accomplishing a skill or activity. I didn't mean math scale by this continuum, where you can put the points, but something like ranging numbers from 8 to 12. I meant this numbered continuum.

Interviewee: If that's what you mean, I think I agree with this, because students should know and the students want to know whether they're improving or not. How can they show this if not by continuum, by giving them a kind of grading? That's shows them a narrow-marrow of their own progress thorough this grading system.

Interviewer: And while you're giving them numbers, do you discuss with them what these numbers mean? For example, A; this means..., or C this and this...

Interviewee: Yes, of course. Not only discuss; it's written. They know that A means this.

Interviewer: Do you give them certain rubrics and explain them to students?

Interviewee: Yes. It's not only me, but all the instructors use them. I have it, here it is. I guess this is what you mean. What is excellent, what is good? For instance, this I give them as evaluation, as self-evaluation after they complete the summer course. They self-evaluate themselves. And as you can see, they don't all evaluate themselves "excellent"; "good/3" and 3 means adequate. And under adequate there comes the rubrics, the explanation what it means.

Interviewer: Do you have such rubrics for speaking or you use it for all skills?

Interviewee: Actually, this is for listening, the other list for reading, writing. There must be for speaking, too. Yes, here it is. But it isn't printed out.

Interviewer: So, all the IEP instructors use these rubrics?

Interviewee: Yes.

Interviewer: I'd like to give you some additional questions concerning our scoring rubrics. Do you feel confident when assessing students?

Interviewee: It's very difficult. It's the most difficult thing to assess students' oral performance, because every teacher might have his or her preferences. Some teachers do not pay attention to grammar mistakes. They pay attention to fluency or some other teachers pay attention to organization and less attention to other things. So it's very subjective. But I try to be objective, because I know that I pay attention to grammar mistakes. So I'm aware of it and

I try to refrain myself. Not to allow this to emphasize my grading: to be objective and to follow the rubrics when I evaluate the students' oral speech. I go over this.

Interviewer: Do you think you use these rubrics, because you have to use them, or do you use them because you like them?

Interviewee: The thing is that these rubrics help students to be aware of their own progress. It's a way, a path towards autonomous learning. And if they start doing it from the very beginning, by self-assessing themselves they become intellectually, morally independent in their learning, more reasonable for their learning and this is one of our goals to make them more independent ones. You can never feel confident but you are 100% subjective. No. There are so many things that cop up in this evaluation. But I try to be objective in my assessment of students' oral performance. And oral performance presents more difficulty for assessment than let's say writing.

Interviewer: Do you think you would benefit from getting training in the field of spoken language testing?

Interviewee: Definitely. I would benefit. Generally all kinds of training are beneficial for the teachers. And I think teacher should go through this sort of training from time to time. And that will help them to be aware of the latest developments. Then to know how they're doing to self-assess themselves. These trainings are very necessary.

Interviewer: And how could you be prepared professionally; may be take courses, seminars, locally?

Interviewee: I don't think locally. Maybe only within the British council or AUA, because I don't see any other ways of professionally preparing or developing in Armenia, except these two places. If they organize this type of training, that would be wonderful. I don't think that the teacher training that we have in Yerevan University of education gives me anything that will help me develop professionally. But abroad will do of course.

Interviewer: I talked to some of the instructors, but they think this is not realistic, because it's pretty expensive.

Interviewee: It is expensive, that's right. In the past during the Soviet period teacher from republics were sent to Moscow once every 5 years. That was called professional training development. But now...

Interviewer: Do you think it is possible to increase public understanding and support of spoken language testing? And how?

Interviewee: First of all we must illuminate people who are in our profession-teachers. They should realize this important, because public it's a professional quest. I don't think public knows very much about second language teaching in general.

Interviewer: What about students? I consider students as part of the public. To raise their understanding about testing. For example, it seems to me that IEP students don't think that while testing you do something seriously. They think we are doing it to help them pass to the first year degree program.

Interviewee: Yes, it's possible if you explain it to them. But it's not a question for students, but teachers, professionals. As the teacher how many of them think that testing should be regarded as a profession and find out how many of them think that it is. I don't think that they will generally. We should first of all raise teacher's awareness of the necessity to regard testing as a profession.

Interviewer: And then teachers on their turn will raise public understanding.

Interviewee: If they are aware of its necessity, they will willy-nilly do that, because it will be reflected in their work, in their approach to teaching and the public will understand.

Interviewer: And the last question; would you like to cooperate with other groups interested in spoken language testing?

Interviewee: Yes.

Interviewer: Have you cooperated with other groups?

Interviewee: No. I don't know any other groups. No, not to my knowledge.

Interviewer: Maybe school teachers should get training at AUA, or maybe the instructors of AUA should train teachers, school teachers.

Interviewee: Yes, it would be nice, not only AUA, but British council, too.

Interviewer: The last thing. I asked you to get the students self-assess themselves through scoring rubrics, but you told them to do peer evaluation. Could you please explain why?

Interviewee: The thing is that when the student prepares the speech at home, he does his best to present it taking into account criteria according to which he is going to be evaluated. He knows the rubrics in advance. So I think the student has already prepared for the speech and has done his best. When the student is being evaluated by his peers and then he becomes aware of the drawbacks that his peers noticed.

Interviewer: Feedback that peers give?

Interviewee: The feedback that they give is the reflection of their opinion of the presentation. And that will help the presenter to be aware of his own weaknesses. If I ask the student to self-evaluate himself, the student has already prepared for this presentation, for the speech, he has done his best. So the only thing he must do is; to what extent he managed to fulfill what he had planned to do. That's also one way of learning. But when the student hears the weak points that his peers heard in his presentation and then the discussion that follows after the presentation. I think that it already helps him to be aware of these weaknesses.

Interviewer: So you feel that students will benefit more from peer-evaluation rather than self-assessment?

Interviewee: That's right. Self-assessment is also possible. But I think peer evaluation is more useful.

Interviewer: Peer evaluation through rubrics?

Interviewee: Definitely through completing the rubrics. And because a person may not notice his own weakness, for example that your voice is monotonous or that he's looking at one person all the time.

Interviewer: No eye contact?

Interviewee: Right. From outside this is more obvious. I think it will be more beneficial for the presenter.

Interviewer: Oh, I see. Thank you very much for interview.

Appendix 8 : Profile of Survey Sample

Table 1: Distribution of respondents by age

Age	Number	Percentage
20-25	28	58.3%
26-30	12	25%
31-40	7	14.6%
41-50	-	-
above 50	-	-
Missing	1	2.1%
Total	48	100%

Table 2: Distribution of respondents by gender

Gender	Number	Percentage
Male	11	22.9%
Female	37	77.1%
Total	48	100%

Table 3: Distribution of respondents by language training

Years of language training	Number	Percentage
0-1	1	2.1%
1-3	6	12.5%
3-5	9	18.8%
5-10	13	27.1%
More than 10	17	35.4%
Missing	2	4.1%
Total	48	100%

Table 4: Descriptive Statistics of respondents' TOEFL scores

	Number	Minimum	Maximum	Mean	Std. Deviation
TOEFL	43	423	560	517.02	29.755

Table 5: Distribution of respondents by IELTS, FCE, BEC, or EST

Took IELTS, FCE, BEC, or EST	Number	Percentage
Yes	1	2.1%
No	46	95.8%
Missing	1	2.1%
Total	48	100.0

Appendix 9: Frequencies and Frequency tables of the open-ended questions of the student questionnaire

Frequency

Statistics

		THIRTEEN	FOURTEEN	FIFTEEN
N	Valid	47	40	41
	Missing	1	8	7

Frequency Table

THIRTEEN

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Liked everything	6	12.5	12.8	12.8
	Good guidance	25	52.1	53.2	66.0
	Grading benefits	15	31.3	31.9	97.9
	Disliked evrything	1	2.1	2.1	100.0
	Total	47	97.9	100.0	
Missing	9	1	2.1		
	Total	48	100.0		

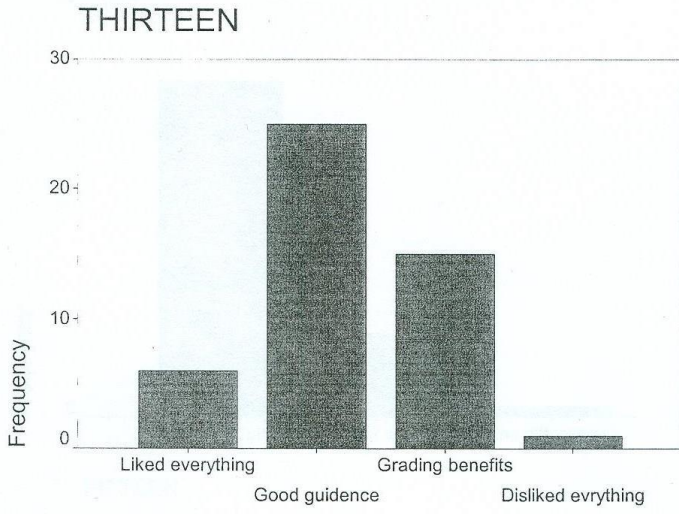
FOURTEEN

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	liked everything	28	58.3	70.0	70.0
	limitations	8	16.7	20.0	90.0
	evaluation by Ss	4	8.3	10.0	100.0
	Total	40	83.3	100.0	
Missing	9	8	16.7		
	Total	48	100.0		

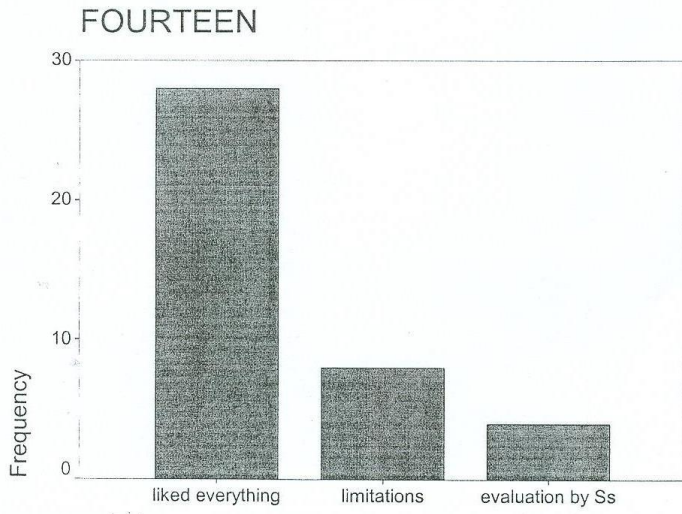
FIFTEEN

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	new way better	32	66.7	78.0	78.0
	not sure	8	16.7	19.5	97.6
	no difference	1	2.1	2.4	100.0
	Total	41	85.4	100.0	
Missing	9	7	14.6		
	Total	48	100.0		

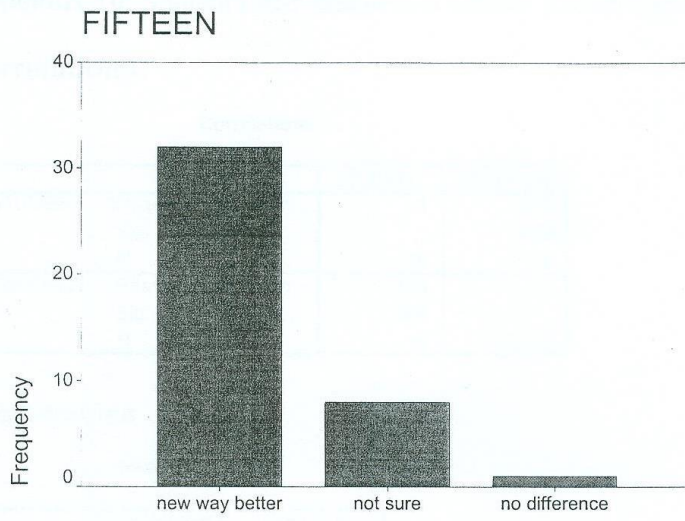
Bar Chart



THIRTEEN



FOURTEEN



FIFTEEN

Appendix 10: Student's self-assessment and the teacher's evaluation grade correlation

Correlations

Correlations

		STUDENT	TEACHER
STUDENT	Pearson Correlation	1	.439
	Sig. (2-tailed)	.	.069
	N	18	18
TEACHER	Pearson Correlation	.439	1
	Sig. (2-tailed)	.069	.
	N	18	18

Frequencies

Statistics

		STUDENT	TEACHER
N	Valid	18	18
	Missing	0	0

Frequency Table

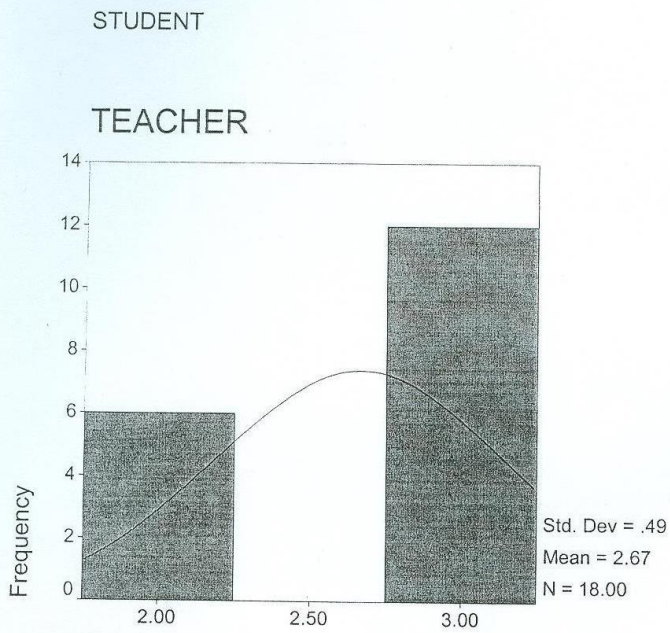
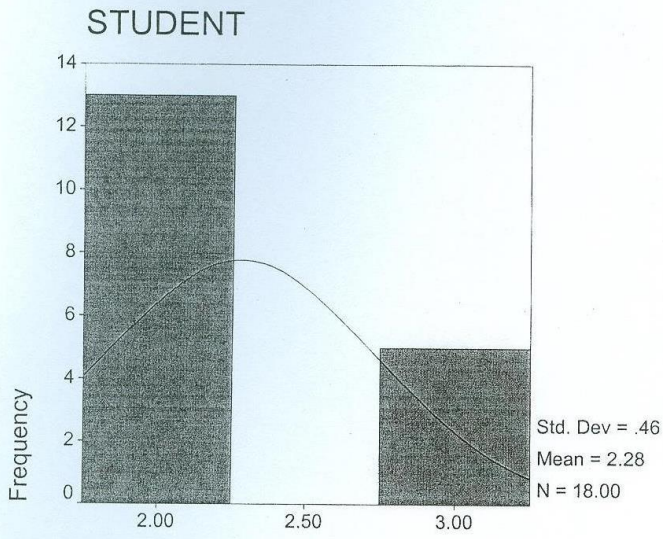
STUDENT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Satisfactory	13	72.2	72.2	72.2
	Very good	5	27.8	27.8	100.0
	Total	18	100.0	100.0	

TEACHER

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Satisfactory	6	33.3	33.3	33.3
	Very good	12	66.7	66.7	100.0
	Total	18	100.0	100.0	

Histogram



TEACHER