

Designing Scoring Rubrics for Assessing English Conversation Teachers' Performance¹

Nurdiana
Universitas Bunda Mulia
nurdiana@bundamulia.ac.id

Abstract

Studies on language tests and assessment have been discussed for years. Most of them deal with how to assess students' performance, or how to write a good test. However, we lack investigations on teachers' performance assessment, particularly on English Conversation teachers'. One of the most common assessment tools used to evaluate teachers performance is Communicative Orientation of Language Teaching (COLT)—an observation scheme deployed in a classroom; nevertheless, it is not specifically devised for conversation classes. We also have other kinds of teacher evaluation rubrics, yet they cover dimensions of how teachers perform in class in general. Thus, it is imperative to design scoring rubrics for assessing English conversation teachers' performance. This paper discusses preliminary research on scoring rubrics designed to assess English conversation teachers' performance. It includes what dimensions should be assessed in the scale and the score for each dimension. It would be like a band scale used to assess students' oral communication skills. This research deployed qualitative research as the data analysed were the non-numerical ones. The results of the study showed that it was not an easy assignment to design scoring rubrics to assess conversation teachers' performance as teachers' response to the designed scoring rubrics varied greatly.

Keywords: Scoring rubrics, English conversation classes, and teachers' performance

1. Introduction

Scoring rubrics are a scale used to assess one's performance. In English language teaching, they are normally used to examine students' oral communication and writing skills. Since scoring rubrics function as an assessment tool, I believe they can also be deployed to evaluate teachers' performance in a classroom, particularly English conversation teachers'. The idea of designing scoring rubrics for the assessment of English conversation teachers performance came up when I was doing self reflection. I was thinking about a more appropriate assessment tool for English conversation teachers'.

¹ This journal entry was presented as a scientific paper for the 2nd ELTLT International Conference 2013: English Language Teaching, Literature, and Translation.

Indeed, we have Communicative Orientation of Language Teaching (COLT)—an observation scheme which is commonly used in a classroom; nevertheless, it is not specifically applied in speaking or conversation classes.

There are some other kinds of teacher evaluation rubrics, yet they cover overall teachers' performance. Thus, from my vantage point, designing scoring rubrics for assessing English conversation teachers' performance is paramount.

2. Literature Review

2.1 Principles of Communicative Approach

These scoring rubrics are devised by referring to the principles of Communicative Approach which are implemented in Communicative Language Teaching or CLT. Why CLT? CLT is commonly applied in conversation classes as its tenets are appropriate to carry out in oral communication classes. According to Freeman (1986), in Communicative Approach, teachers function as a facilitator of students' learning. They are the managers of classroom activities. One of their role is to promote communication, to be an advisor, to answer students' questions and to monitor their performance.

Some of the techniques or materials associated with the Communicative Approach or CLT are:

a. Authentic Materials

Authentic materials are one of the techniques recommended in Communicative Approach as they contain authentic and natural language use which represent the real life situations. Some examples of authentic materials are movies, TV programs, news, magazines, etc.

b. Communicative Classroom Activities

Class activities such as games, role play, discussion, simulations, presentation, skits, and debates are suggested activities in Communicative Approach (Freeman, 1986 and Richards and Rogers, 2001). Thus an English conversation teacher should apply these class activities in his or her conversation classes.

c. Accuracy and Fluency

In Communicative Approach, errors are tolerated and fluency is the primary goal as Richards and Rogers (2001) and Freeman (1986) state that “Errors are tolerated and seen as outcome of the development of communication skills. Fluency and acceptable language is the primary goal while accuracy is judged not in the abstract but in context.”

2.2 Scoring Rubrics

Scoring rubrics are defined as scoring schemes described by teachers or evaluators to assess students’ performance or efforts (Brookhart, 1999 as cited in Moskal, 2000). Boston (2002) states that a rubric is a rating system which can be used to examine students’ level of proficiency of tasks performed or knowledge displayed.

According to Perlman (2003), a scoring rubric has some components which include:

- a. one or more dimensions regarding the performance which will be assessed or rated.
- b. definitions and examples illustrating the attributes which will be measured.
- c. a rating scale for each dimension.

Perlman (2003) further discusses the selecting tasks for performance assessment. He describes the criteria to be considered which he adapted from Herman, Aschbacher, and Winters (1992, as cited in Perlman, 2003).

- a. Does the task truly match the outcomes or standards you are trying to measure? The answer to this question is related to whether or not the task relevant to the outcome.
- b. Does the task require the students to use critical thinking skills? This question is related to bloom’s taxonomy which require the students to perform a task based on the levels measured in the taxonomy such as analyzing, drawing inferences or conclusion, evaluating, synthesizing, creating, and comparing.
- c. Is the task a worthwhile use of instructional time? Questioning how much time students would spend on a task should be taken into account.

- d. Does the assessment use engaging tasks with real-world applications? Tasks that are imagery or are not taken from real-life situations will be less interesting and students could be demotivated.
- e. Are the tasks fair and free from bias? The tasks should be less subjective or do not discriminate students. In other words, the tasks should be free from gender discrimination, races, or cultures.
- f. Is the task clearly defined?
- g. Is the task feasible?
- h. Will the task be credible?

There are two types of scoring rubrics namely holistic and analytic. According to Nitko (2001, as cited in Mertler, 2001), “a holistic rubric requires the teacher to score the overall process or product as a whole, without judging the component parts separately” while “an analytic rubric, the teacher scores separate, individual parts of the product or performance first, then sums the individual scores to obtain a total score.” (Moskal, 2000; Nitko, 2001 as cited in Mertler, 2001).

The following is the example of holistic scoring rubrics taken from *The Test of Spoken English band descriptors for Overall features* (ETS, 2001b: 30 as cited in Luoma, 2004)

60 Communication almost always effective: task performed very competently. Speaker volunteers information freely, with little or no effort, and may go beyond the task by using additional appropriate functions.

- Native-like repair strategies
- Sophisticated expressions
- Very strong content
- Almost no listener effort required

50 Communication generally effective: task performed competently.

Speaker volunteers information, sometimes with effort; usually does not run out of time.

- Linguistic weaknesses may necessitate some repair strategies that may be

slightly distracting

- Expressions sometimes awkward
- Generally strong content
- Little listener effort required

40 Communication somewhat effective: task performed somewhat competently. Speaker responds with effort; sometimes provides limited speech sample and sometimes runs out of time.

- Sometimes excessive, distracting, and ineffective repair strategies used to compensate for linguistic weaknesses (e.g. vocabulary and/or grammar)
- Adequate content
- Some listener effort required

30 Communication generally not effective: task generally performed poorly. Speaker responds with much effort; provides limited speech sample and often runs out of time.

- Repair strategies excessive, very distracting, and ineffective
- Much listener effort required
- Difficult to tell if task is fully performed because of linguistic weaknesses, but function can be identified

20 No effective communication: no evidence of ability to perform task.

Extreme speaker effort is evident; speaker may repeat prompt, give up on task, or be silent.

- Attempts to perform task end in failure
- Only isolated words or phrases intelligible, even with much listener effort
- Function cannot be identified

An example of analytic scoring rubrics can be seen below. (Source: Analytic descriptors of spoken language, Councils of Europe, 2001, 28-29 as cited in Luoma, 2004)

	<i>Range</i>	<i>Accuracy</i>	<i>Fluency</i>	<i>Interaction</i>	<i>Coherence</i>
C2	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it.	Can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly. Can interweave his/her contribution into the joint discourse with fully natural turntaking, referencing, allusion making, etc.	Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.
C1	Has a good command of broad range of language allowing him/her to select a reformulation to express him/herself clearly in an appropriate style on a wide range of	Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of	Can select a suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2	Has a sufficient range of language to be able to give clear descriptions and express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can produce stretches of language with fairly even tempo: although he/she can be hesitant as he/she searches for patterns and expressions. There are a few noticeably long pauses.	Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting others in, etc.	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness' in a long contribution.
B1	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversations on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2	Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Uses some simple structures correctly, but still systematically makes basic mistakes.	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.	Can answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can link groups of words with simple connectors like 'and' and 'but' and 'because'.
A1	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorised repertoire.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.	Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	Can link words or groups of words with very basic linear connectors like 'and' or 'then'.

According to Stevens and Levi (2005), Allen (2004) , and Huba and Freed (2000, as cited in [http://assessment.uconn.edu/docs/How to Create Rubrics.pdf](http://assessment.uconn.edu/docs/How_to_Create_Rubrics.pdf)), a rubric involves four components:

Part 1: Task Description

- It is about students' performance.
- The tasks could be specific assignments such as a paper, a poster, a presentation.
- The tasks can be used to assess students' behavior; e.g., participation, use of proper lab protocols, behavioral expectations in the classroom.

Part 2: Scale

- Describes how well or poorly students do the tasks given.
- Terms such as “Mastery”, “Partial Mastery”, “Progressing”, “Emerging” are applicable.
- Nonjudgmental or noncompetitive language: “High level”, “Middle level”, “Beginning level”.
- The following labels are also common to use:
 - Sophisticated, competent, partly competent, not yet competent
 - Exemplary, proficient, marginal, unacceptable
 - Advanced, intermediate high, intermediate, novice
 - Distinguished, proficient, intermediate, novice
 - Accomplished, average, developing.
- We can use 3-5 level in the scale.

The higher the level is, the better the performance is.

Part 3: Dimensions

Dimensions describe the items assessed through the scale. For example, in the scoring rubrics I have designed, the dimensions are authentic materials, class activities, accuracy/fluency, and classroom management.

Part 4: Description of the Dimensions

- A rubric should contain descriptions of each dimension. For example, if the score is one, the descriptions of dimension authentic materials are teachers do not use authentic materials at all.

Stevens and Levi (2005), Allen (2004) , and Huba and Freed(2000, as cited in <http://assessment.uconn.edu/docs/How to Create Rubrics.pdf>) further said that there are four stages in constructing a rubric.

“1. Reflecting. In this stage, we take the time to reflect on what we want from the students, why we created this assignment, what happened the last time we gave it, and what our expectations are.

- a) Why did you create this assignment?
- b) Have you given this assignment or a similar assignment before?
- c) How does this assignment relate to the rest of what you are teaching?
- d) What skills will students need to have or develop to successfully complete this assignment?
- e) What exactly is the task assigned?
- f) What evidence can students provide in this assignment that would show they have accomplished what you hoped they would accomplish when you created the assignment?
- g) What are the highest expectations you have for student performance on this assignment overall?
- h) What is the worst fulfillment of the assignment you can imagine short of simply not turning it in at all?

2. Listing. In this stage, we focus on the particular details of the assignment and what specific learning objectives we hope to see in the completed assignment.

Answers to (d)-(e)-(f) above regarding skills required, the exact nature of the task, and the types of evidence of learning are most often a good starting point to generate this list. Once the learning goals have been listed, you add a description of the highest level of performance you expect for each learning goal. These will later contribute to the “Descriptions of Dimensions” on a finished rubric.

3. Grouping and Labeling. In this stage, we organize the results of our reflections in Stages 1 and 2, grouping similar expectations together in what will probably become the rubric dimensions. Start with the highest performance expectations completed in Stage 2 and group together items which are related. Once the performance descriptions are in groups of similar skills, read them and start to find out what is common across the

group and label it. These labels will ultimately become dimensions on the rubric – it is important to keep them clear and neutral; e.g., “Organization”, “Analysis”, or “Citations”.

4. Application. In this stage, we apply the dimensions and descriptions from Stage 3 to the final form of the rubric, utilizing the matrix/grid format.”

3. Subjects of the research

The subjects of the research were English conversation teachers who teach in a language centre. 14 English conversation teachers were involved in this study.

4. Rationale of the dimensions for the scoring rubrics

There are some reasons why the researcher included authentic materials, class activities, accuracy/fluency, and classroom management as the dimensions in these scoring rubrics. First of all, authentic materials, class activities, and accuracy/fluency represent Communicative Approach which is implemented in Communicative Language Teaching. As described in the literature review, these three dimensions are techniques deployed in Communicative Language Teaching.

Secondly, classroom management is an important element in English language teaching as it affects the effectiveness of class activities. Classroom management includes seating arrangement, the use of classroom facilities such as boards, LCDs, laptops, and markers. It also covers how the students should work on class activities—whether or not they should work in pairs, in groups, or individually.

The scoring rubrics I designed are the analytic ones. Analytic scoring rubrics, in my opinion, are “friendly” to use and the dimensions are more clear cut than those in holistic ones. To design these scoring rubrics, I adapted all of the theories on scoring rubrics described in literature review.

5. Research method and data collection

To obtain the data, I distributed the scoring rubrics I designed and below the rubrics, there are two open-ended questions that the respondents should answer. The responses to the questions helped the researcher to picture the

respondents' ideas of an ideal scoring rubrics which are applicable in conversation classes.

As the data were the non-numerical ones, this study is categorized as a qualitative research (Dornyei, 2007). The following is the scoring rubrics I devised and the two open-ended questions that the subjects of the research had to answers.

Scale Descriptions	Authentic Materials	Class Activities	Accuracy/Fluency	Classroom Management
4	Teachers always use authentic materials in class.	Teachers use more various classroom activities e.g. games, role play, presentation, group discussions, scrambled sentences, and debates.	Teachers put emphasis on students' accuracy and fluency. They assess students' fluency without neglecting their accuracy.	Teachers apply good classroom management, e.g. effective/efficient use of board, able to arrange the seats effectively, use various teaching aids, and able to manage teachers and students talking time.
3	Teachers often use authentic materials e.g. songs, movies, stories, and realia.	Teachers use a lot more class activities, e.g. games, role play, and presentation.	Teachers are concerned with students' fluency, but accuracy is still their priority.	Teachers are more knowledgeable of classroom management e.g. they know how to use the board efficiently/effectively, how to arrange classroom seats, and use teaching aids.
2	Teachers hardly ever use authentic	Teachers use more various class activities,	Teachers are more concerned with students' accuracy,	Teachers have little knowledge of

	materials, e.g. songs only.	e.g. games and role play.	but fluency is not neglected.	classroom management. For example, they know how to use the board efficiently/effectively, but they do not know how to arrange seats in class, and do not use teaching aids.
1	Teachers do not use authentic materials at all.	Teachers use less various class activities, e.g. games only.	Teachers put emphasis on accuracy or fluency only.	Teachers do not know how to manage the class, e.g. ineffective/inefficient use of board, ineffective/inefficient seating arrangement, ineffective/inefficient teachers and students talking time, and no teaching aids at all.

Please answer the following questions.

1. Are there any other components/elements that should be assessed? If so, please provide your reasons.
2. Please give comments/suggestions on each scale and description above used to assess English conversation teachers performance. Your comments/suggestions will be very advantageous for the improvement/revision of this scoring rubrics.

6. Results and discussion

The responses to question 1 varied greatly and a lot of suggestions were given to the improvement of this scoring rubric. One of them is I should have assessed the use of the materials in class. In other words, the four rubrics assessed

will not be useful if the teachers do not use the materials covered in the book used in the class. Teachers may use authentic materials, various class activities, and apply good classroom management, but it will be useless if they do not teach the lessons that they are supposed to teach. Thus in this case, whether or not the teachers use the syllabus when delivering the materials should be assessed too.

Another component that should be taken into account when assessing English conversation teachers' performance is teachers' knowledge of English, such as their speaking skills and spoken language, pronunciation, fluency, and their accuracy. It is imperative to assess teachers' knowledge of English as they are the main source of language input for the students. Other research findings regarding the components that should be assessed are the clarity of instructions given for all class activities, rapport building, and lesson plans. According to the respondents, teachers need to be able to build rapport to the students so that the students are willing to get involved in all class activities and they have courage to speak up their mind.

Lesson plans, according to some of the respondents, need to be paid attention to. Systematic and well planned lessons will affect the learning process. If they are not well planned, teachers will have no directions what to teach and how to teach. Thus they will be kind of cluttered which make the teachers unable to achieve the objectives of the lessons.

Responses to question 2 given in the questionnaire are even more various than those to question 1. One of the respondents, for example, thinks that authentic materials should not be assessed from how often the teacher uses them. He thinks that authentic materials should be used effectively, and should be used when necessary only. The use of authentic materials should also be assessed from the appropriateness. For example, if a teacher always uses authentic materials, but the authentic materials don't have any clear purposes related to the lessons, then they would be a waste of time.

The same comment would also be given to the class activities. The assessment should cover not only how many or how various the class activities are, but also how effective they are. Do the class activities have clear purposes to make students practice, or just to make the class more lively or fun? Another

comment on class activities is the word “various” in class activities rubric. It should be replaced with one, two, three, or four to make it more explicit as the term “various” has no clear indicators for class activities.

Regarding accuracy/fluency, these two components should be assessed separately. In other words, rubric column for accuracy should not be in the same column with fluency. That means each of these components should have its own descriptions that can be used for the assessment of English conversation teachers’ performance. Some other suggestions for this rubric are accuracy/fluency may not be separated, but there should be the word “attention” before accuracy/fluency as the assessment puts emphasis on whether or not the teachers pay more attention to the accuracy or fluency when they assess their students’ performance and when they are delivering the materials. Hence, the rubric will be “attention to accuracy/fluency” not “accuracy/fluency”. The assessment of accuracy/fluency, in addition, should depend on the students’ level. Therefore, the scoring rubrics used to assess the teachers’ performance in lower level classes (elementary) would be different from those used to assess the teachers’ performance in higher level classes (intermediate).

Some comments and suggestions are given to the last rubric, which is about classroom management. According to the respondents, grouping should be assessed as it can show teachers’ ability and creativity to group the students in many ways with different purposes. In addition, teachers’ talking time should be assessed in different rubric (not in the classroom management component) as teachers’ talking time and students’ talking time play an important role in a conversation class.

Other comments given to this scoring rubrics are “What is the exact criteria for “do not know”, “little”, and “good” in classroom management? What are the indicators of ineffective and inefficient in classroom management? and less various, more various in class activities should be replaced by 1,2, 3 or more activities.

7. Conclusion and suggestions

To conclude, designing scoring rubrics to assess English conversation teachers' performance is not an essay assignment as there are many criteria and rubrics should be listed and observed carefully before designing the rubrics. A further study needs to be conducted to improve and develop the scoring rubrics to assess the performance of English conversation teachers by reviewing the suggestions and comments given by the respondents. Moreover, the researcher should test its reliability in order to make sure that these scoring rubrics are reliable enough to assess English conversation teachers' performance.

References

- Boston, Carol. (2002). *Understanding scoring rubrics: A Guide for Teachers*. Washington D.C.: ERIC Clearinghouse on Assessment and Evaluation.
- Dornyei, Zoltan. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press
- Freeman-Larsen, Diane. (1986). *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press
- Luoma, Sari. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- Mertler, Craig A. (2001). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research & Evaluation*, 7(25). Retrieved 29 June 2014 from <http://PAREonline.net/getvn.asp?v=7&n=25> .
- Moskal, B.M. (2000). Scoring rubrics: what, when and how?. *Practical, Assessment, Research & Evaluation*, 7(3). Retrieved 29 June 2014 from <http://PAREonline.net/getvn.asp?v=7&n=3>
- Perlman, C.C. (2003). *Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics*. U.S. Department of Education: Office of Educational Research and Improvement: Education Resources Information Centre (ERIC)
- Richards, J.C. and Theodore S. Rogers (2001). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- [http://assessment.uconn.edu/docs/How to Create Rubrics.pdf](http://assessment.uconn.edu/docs/How%20to%20Create%20Rubrics.pdf) (retrieved on 15 September 2013)