Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

ISSN1925-931X

Putting Rubrics to the Test: The Effect of Rubric-Referenced Peer Assessment on EFL Learners' Evaluation of Speaking

Masoume Ahmadi, Naser Sabourian Zadeh

Department of Linguistic, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran.

Abstract

This study attempted to shed some light on the effect of rubric-referenced peer assessment on EFL learners' speaking skill and on the cultivating the learners' awareness of having appropriate criteria for speaking, as one of the four major skills. This study explored the effect of rubrics on peer assessment of 18 Iranian EFL learners. First, learners assessed their classmates speaking performance based on their own presuppositions and assumptions. Subsequently, a spoken language rubric was introduced to them. They re-assessed their classmates' performances through using this rubric. Quantitative data analysis revealed significant difference between the results. In-depth qualitative analyses of comments and marginal notes written down by learners revealed that peers heed not only to institutional components specified in scoring scales but also to other irrelevant criteria such as the result of the speech act performed. The study has suggested that the use of a combination of peer assessment and rubric-referenced assessment encourages students to become more rationally responsible and reflective and has shown positive formative effects on student achievement and attitudes. The article concludes with some guidelines for practitioners. The findings of this study also provide insight into the effective assessment and recommendations for future research and practice are made.

Key Words: Scoring scale, Peer Assessment, Speaking skill, Rubric-referenced assessment

I. Introduction

The need for the mastery of English for educational and professional purposes has heightened with the phenomenal spread of this language as a global lingua franca. Recently, the focus of attention and concern in EFL/ESL programs has shifted from emphasize on general language proficiency to emphasize on the students' development of skills, especially in advanced classes. Consequently, speaking module has caught the attention more than it probably ever did. As a result the ability to speak in a second/foreign language is highly valued in educational settings and this ability is seen as a key to entry into the academic discourse community.

White (1985, mentioned in Connor & Mbaye, 2002) has concluded that because testing strives to bridge the gap between teaching and learning in an educational context, the resurgence of interest in EFL/ESL teaching has translated into resurgence in attention to issues of assessing skills. Regarding this added attention to assessment, teachers and instructors' awareness of important testing issues such as validity, reliability , test types, test purpose , specific methods of assessment has heightened in comparison with the past. Assessment is a critical activity in any instructional operation. It is of essence for both learners and teachers to be involved in and have

Journal of Academic and Applied Studies
ISSN1925-931X

control over the assessment methods, procedures and outcomes, as well as their underlying rationale (Jafarpur, 1991). A bulk number of studies (Boud, Cohen, & Sampson; 1999; DiPardo& Freedman, 1988; Dochy, Segers&Sluijsmans, 1999; Liu & Hansen, 2002; Long & Porter, 1985; Masten, Morison, Pellegrini, 1985, among others) have accentuated the benefits of incorporating peer assessment, as one form of alternative assessment, into the regular assessment procedures. Topping (1998) defines peer assessment as 'an arrangement in which individuals consider the amount, level, value, worth, quality of success of the products or outcomes of learning of peers of similar status' (p. 250). Peer assessment encourages reflective learning through observing others' performances and becoming aware of performance criteria (Falchikov, 1986; also claimed in different contexts by Topping, 1998; Tornow, 1993; Somervell 1993). However, the benefits which peer assessment may bring into a language classroom cannot be guaranteed unless students are capable of implementing the assessment (Hu, 2005; Liu & Hansen, 2002; Saito, 2008). In this sense, several research studies on peer writing response groups have investigated the capacity issue by looking at training effects and have found benefits of training for the revision process (Stanley, 1992; Berg, 1999). Although the advantages of training for peer writing responses appear to be confirmed, no study has required students' use of a rating scale of any kind (Saito, 2008) on assessing speaking performance of the learners. Hence, in order to bridge this gap, the present study will make use of a scoring rubric to see whether it makes any difference in students' scores compared to training. Rubrics have become popular with teachers as a means of communicating expectations for an assignment, providing focused feedback on works in progress, and grading final products (Andrade 2000). Although educators tend to define the word 'rubric' in slightly different ways, a commonly accepted definition is a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor (Andrade 2000). Some studies on peer review tasks (for example, Nelson & Carson, 1998) have also reported that students often do not view their peers' feedback on grammar and lexis as effective and hence do not pay much attention to it in their subsequent revisions (Nelson & Murphy, 1993). Thus, in the present study, the effect of using a scoring rubric will be investigated in peer assessment to also see whether it neutralizes learners' inclination toward implementing some extraneous criteria.

"Although the rubric has emerged as one of the most popular assessment tools in progressive educational programs, there is an unfortunate dearth of information in the literature quantifying the actual effectiveness of the rubric as an assessment tool *in the hands of the students"* (Hafner&Hafner, 2003, p. 1509).

Research Questions

Regarding the gap in the literature, this study takes over the mission of addressing the following questions:

Q1) Does rubric-referenced peer assessment incorporate any significant difference on the scores of EFL learners' speaking performance?
Q2)How can the utilization of a scoring rubric affect the criteria in which learners lay emphasis on?

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

Journal of Academic and Applied Studies
ISSN1925-931X

II. Methodology

Participants

The study employed a convenience sample of 18EFL learners in an intact class who were placed in intermediate level at Kish Air Language Institute, Tehran, Iran. The Participants aged between 20 and 28 with the median age of 25.8. Half of the participants (50%) were females, and the remaining percent (50%) were males. The participants took part in the study voluntarily. The reason for using volunteer participants was to make sure that they would participate willingly because such an assessment is a difficult and time-consuming job for learners and they do not enjoy it. Another reason was to make sure that they would participate in both phases, that is, peer assessment with and without referring rubrics. The present researcher decided to research these two phases in just one class in order to eliminate the individual variation among individuals and to guarantee the internal validity of the study.

Instruments

A. Role plays

EFL learners were supposed to make and play a conversation about a topic "shopping"proposed by the teacher. Role plays were recorded and learners were to assess their classmates' speaking skill.

B. Rating scale

A spoken language rubric administered by Council of Europe (2001, see Appendix) was employed for the purpose of this study. This scale was originally containing scoring rubrics and a concise descriptor for every key word and component related to speaking skill (i.e., range, accuracy, fluency, interaction, coherence).

Data collection Procedure

In order to fulfill the objective of the present study, certain steps were followed. First of all, raters were given instructions which clarified what they were supposed to do (e.g. making conversation, playing roles, attending to their classmates' conversation, assessing the speaking skill of their friends with and without referring the rubric, etc.) They also were informed that they were supposed to have their conversations within 4 minutes. After that, they started making conversations in 10 minutes. While playing the conversation, the teacher recorded their conversations. Recording didn't threat the internal validity of the study due to the fact that classes are equipped by the cameras and learners are accustomed to them. Subsequently, the learners rated the spoken performance of each other and wrote their ideas down. After collecting the learners' ideas, the scoring rubric was introduced to them and they received abundant information about the use of such a scoring scale. The teacher exemplified for the learners how

Journal of Academic and Applied Studies
ISSN1925-931X

to use them and to what each category refers. The recorded episode was played back and, in this turn, learners were supposed to assess their classmates' speaking skills through employing rubrics.

Scripts were analyzed for five major categories of range, accuracy, fluency, interaction, coherencewith bands ranged from A1 to C2, A1 being lowest and C2, the highest band. Each band in each category has an accompanying profile description which essentially stipulates key features to focus on regarding particular categories and their corresponding bands. Soring scales also contained a part-blank space-to be filled with score and scoring criteria by rater. In order to facilitate comparison between scores in these two phases, the researchers assigned the value 1 to A1, 2 to A2, 3 to B1, 4 to B2, 5 to C1, and finally 6 to C2. In this way, the total number of scoring was equal to thirty in both phases. Subsequently, the collected data were classified and analyzed. Paired samples t-test analysis was run to find the answers for the first fore-mentioned research question. In order to plumb the details to which learners paid attention, the criteria considered by learners while peer-assessing were investigated qualitatively.

Data Analysis

To analyze the collected data and test the hypothesis of the study, paired samples t-test was employed through the use of 16th version of statistical package for social sciences (SPSS). Qualitative analysis will be also carried out in order to investigate the rubric-referenced peer assessment in details.

III. Results and Analysis

Q1) Does rubric-referenced peer assessment incorporate any significant difference on the scores of EFL learners' speaking performance?

A dependent paired samples t-test was run to compare the mean scores of the EFL students speaking performance on the peer assessment and rubric-referenced peer assessment. The t-observed value is 4.35 (Table 1). This amount of t-value is higher than the critical value of 2.12 at 16 degrees of freedom.

The results of this study are both statistically significant and meaningful. The effect size for the t-value of 4.35 is .54. Normally, an effect size of .14 and above is considered strong. Based on these results, it can be concluded that there is a significant difference between the mean scores of the EFL students speaking performance on the peer assessment and rubric-referenced peer assessment. Table 2 displays the mean scores for the students on the mean scores of the EFL students speaking performance on the peer assessment and rubric-referenced peer assessment.

Q2) How can the utilization of a scoring rubric affect the criteria in which learners lay emphasis on?

In order to address the second question, concerning the criteria to which raters dealt with in peer assessments, in-depth qualitative analysis were run. The results revealed that learners, by themselves are not fully aware of the criteria of speaking performance assessment. Although a bulk of research (Berg, 1995; Boud, Cohen, & Sampson,1999; Dochy, Segers&Sluijsmans,

ISSN1925-931X

1999; Hu, 2005;Masten, Morison, Pellegrini, 1985; to name a few)has emphasized the effective role of peer assessment, it seems not to be enough alone. When learners are not fully familiar with the criteria of an effective performance, peer assessment will not suffice. Analyses revealed that EFL learners are concerned about the following criteria while assessing performances of their peers:

1) *Accuracy*
2) *Fluency*
3) *Range*
4) *Coherence*
5) *Interaction*
6) *Pragmalinguistics*
7) *Loudness*

Here are some examples for each category (no change is undergone):

Pragmalinguistics*: In my opinion she talks very politely and uses more polite expressions.*
Fluency: *He talks with some hesitation and has many pauses during his conversation.*
Range: *the range of using different words is very broader in his speech.*
Interaction: *there is no intonational cue.*
Pragmalinguistics: *He uses enough and satisfactory reasons to convince the seller to change the item.*
Loudness: *She is not loud enough.*
Accuracy: *Make a mistake not do a mistake.*
Pragmalinguistics: *No greeting.*

Among these criteria, the two last ones, i.e. pragmalinguistics and loudness, can be considered as extraneous criteria for assessing performance. The last example clarifies the point perfectly. *"No greeting"* reflects the personal attitude of the learner which reflects the cultural venue they have been in contact with. Acquainted with rubrics, learners paid attention to the criteria which are of essence in a speaking performance assessment. When rubrics are used to guide peer-assessment, students become increasingly able to spot and solve problems in other's work. The results of thisstudy corroborate the findings of Hafner and Hafner (2003) who claimed that the general form and evaluative criteria of the rubric are clear and that the rubric is a useful assessment tool for peer-group (and self-) assessment by students.

IV. Conclusions

Peer assessment as a desirable technique to be applied to raters is a value–laden approach to learning, teaching and rating which seeks to involve raters in decision making about assessment process and how to make judgments on their own and their colleagues. Overall, this research showed discrepancies among scoring criteria for raters in peer assessment and rubric-referenced

Journal of Academic and Applied Studies
ISSN1925-931X

peer assessment. The results would be of great help to learners to improve their comprehension of what each skills really contains. The effectiveness of the training programs on modifying the view point of teachers has been documented in the literature and it is claimed that training programs result in increasing intergroup consistency and reducing but not eradicating biasedness and severity. Cumming (1990) presented the significance of rater training in reducing the variability in raters' judgments. According to findings of this study, one further step can be added to peer-assessment programs. Future studies can invest on other techniques to enhance the effectiveness of alternative assessment and in particular, peer-assessment and shed some lights on types of peer-assessment which can benefit the most.

Acknowledgements

We would like to express our special thanks of gratitude to the Council of Europe for developing such a comprehensive speaking rubric.

References

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality.*Journal of Second Language Writing,* 8, 225–241.
Boud, D., Cohen, R., & Sampson, J. (1999).Peer Learning and Assessment.*Assessment and Evaluation in Higher Education,* 24(4), 413-426.
Connor, U. &Mbaye, A. (2002). Discourse approaches to writing assessment. *Annual Review of Applied Linguistics,* 22, 263–278.
Cumming, A. (1990). Expertise in evaluating second language compositions.*Language Testing,* 7(1), 31–51.
DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: Theoretic foundations and new directions. *Review of Educational Research,* 58, 119–149.
Dochy, F, Segers, M. &Sluijsmans, D. (1999).The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education,* 24(3), 331-350.
Elder, C. Barkhuizen, G. Knoch, U. and Randow,J. V. (2007). Evaluating rater responses to an online training program for L2 writing assessment.*Language Testing,* 24(1), 37–64.
Fahim, M. &Bijani, H. (2011).The effect of rater training on rates' severity and bias in second language writing assessment.*Iranian Journal of Language Testing,* 1(1), 1–16.
Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment & Evaluation in Higher Education,* 11, 146–165.
Hafner, J. &Hafner, P. (2003).Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education,* 25(12), 1509-1528.
Hu, G. (2005). Using peer review with Chinese ESL student writers.*Language Teaching Research,* 9, 321–342.
Masten, A. S., Morison, P., Pellegrini, D. S. (1985).A revised class play method of peer assessment. *Developmental Psychology,* 21(3), 523-533.

Journal of Academic and Applied Studies
ISSN1925-931X

Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation and Research in Education* 5, 145-57.
Liu, J., & Hansen, J. G. (2002).*Peer response in second language writing classrooms*. Ann Arbor, MI: The University of Michigan Press.
Long, M. H., & Porter, P. A. (1985).Group work, interlanguage talk, and second language learning.*TESOL Quarterly*, 19, 207–228.
Nelson. G. I. & Carson. J. G. (1998). ESL students' perception of effectiveness in peer response groups.*Journal of Second Language Writing*, 7(2), 113-131.
Nelson. G. I. & Murphy. J. M. (1993). Peer response groups: do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27(1), 135-141.
Prins, F. R., Sluijsmans, D. M. A., Kirschner, P. A. &Strijbos, J. (2005). Formative peer assessment in a CSCL environment: a case study. *Assessment and Evaluation in Higher Education*, 30(4), 417-444.
Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25, 553
Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18, 221–233.
Spandel, V., & R.J. Stiggins. (1997). *Creating writers: Linking writing assessment and instruction*. 2nd ed. New York, NY: Longman.
Stanley, J. (1992). Coaching student writers to be effective peer evaluators.*Journal of Second Language Writing*, 1, 217–233.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-76.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

Journal of Academic and Applied Studies
ISSN1925-931X

TABLES

TABLE I
PAIRED-SAMPLES T-TEST PEER ASSESSMENT AND RUBRIC-REFERENCED PEER ASSESSMENT

Paired Differences				t	df	Sig. (2-tailed)
Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
		Lower	Upper			
1.11144	.18787	.61821	1.38179	4.353	16	.030

TABLE II
DESCRIPTIVE STATISTICS OF PEER ASSESSMENT AND RUBRIC-REFERENCED PEER ASSESSMENT

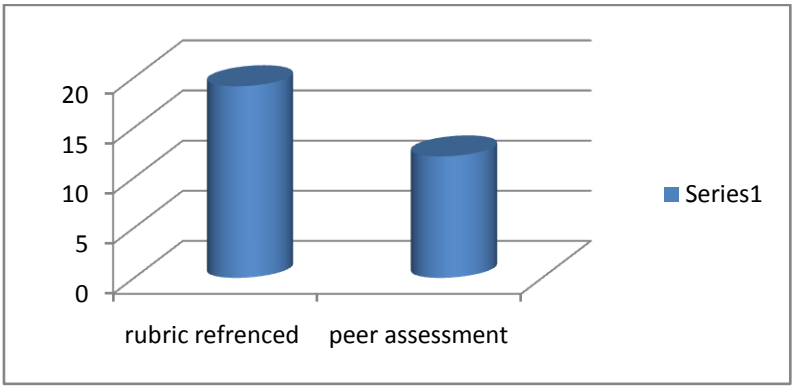
	Mean	N	Std. Deviation	Std. Error Mean
Rubric-referenced peer assessment	19.1143	18	.83213	.14066
Peer assessment	12.1143	18	.93215	.15756

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

Journal of Academic and Applied Studies

ISSN1925-931X

Figures



Graph. 1. Peer Assessment and Rubric-referenced Peer Assessment

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

Journal of Academic and Applied Studies

ISSN1925-931X

Appendix

Spoken language Rubric (Council of Europe, 2001; 28-29)

Dear rater, please listen to each recording carefully and assign a score you think is appropriate to each student's speaking ability (including yours) based on the band descriptors (range, accuracy, fluency, interaction, coherence). You can also write any further comments on the space provided following the rubric. Thanks for your time and consideration.

Analytic descriptors of spoken language (Council of Europe, 2001; 28-29)

Range

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.

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 general, academic, professional or leisure topics without having to restrict what he/she wants to say.

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.

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.

A2
Uses basic sentence patterns with memorized phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.

A1
Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.

Accuracy

C2
Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).

C1
Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.

B2
Shows a relatively high degree of grammatical control.Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.

B1
Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.

A2
Uses some simple structures correctly, but still systematically makes basic mistakes.

A1
Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire.

Fluency

C2
Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that interlocutor is hardly aware of it.

Journal of Academic and Applied Studies
ISSN1925-931X

C1
Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.
B2
Can produce stretches of language with fairly even tempo: although he/she can be hesitant as he/she reaches for patterns and expressions. There are a few noticeably long pauses.
B1
Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair s very evident, especially in longer stretches of free production.
A2
Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.
A1
Can manage very short, isolated, mainly prepackaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.

Interaction

C2
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.
C1
Can select a suitable phrase a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/her own contributions skillfully to those of other speakers.
B2
Can initiate discourse, take his/her turn when appropriate and end conversation he/she needs to, though he/she may not always do this elegantly. Can help the discussion long on familiar ground confirming comprehension, inviting others in, etc.
B1
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.
A2
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.
A1
Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing, and repair.

Coherence

C2
Can create coherent and cohesive discourse making full and appropriate use of a variety of organizational patterns and a wide range of connectors and other cohesive devices.
C1
Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organizational patterns, connectors and cohesive devices.
B2
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
Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2
Can link groups of words with simple connectors like 'and' and 'but' and 'because'.

Journal of Academic and Applied Studies
(Special Issue on Applied Sciences)
Vol. 3(11) November 2013, pp. 1-12
Available online @ www.academians.org

ISSN1925-931X

A1
Can link words or groups of words with very basic linear connectors like 'and' or 'then'.

Further comments to be added (optional)

……
……
……
……
……
……
……

Write the name of each student on each blank and assign the scores accordingly.

1) ..... *Range……accuracy……fluency……interaction……coherence……*
2) ..... *Range……accuracy……fluency……interaction……coherence……*
3) ..... *Range……accuracy……fluency……interaction……coherence……*
4) ..... *Range……accuracy……fluency……interaction……coherence……*
5) ..... *Range……accuracy……fluency……interaction……coherence……*
6) ..... *Range……accuracy……fluency……interaction……coherence……*
7) ..... *Range……accuracy……fluency……interaction……coherence……*
8) ..... *Range……accuracy……fluency……interaction……coherence……*
9) ..... *Range……accuracy……fluency……interaction……coherence……*
10) ..... *Range……accuracy……fluency……interaction……coherence……*