19 Rubrics for Assessment: Their Effects on ESL Students' Authentic Task Performance

Radhika De Silva
Open University of Sri Lanka
krsil@ou.ac.lk

Abstract: Alternative assessment or authentic assessment has gained popularity in the field of second language teaching during the past few years. The main purposes of alternative assessment include assessing students beyond the traditional testing and encouraging students to participate in open discussions about standards and criteria of successful performance (Brown & Abeywickrama, 2010). Anecdotal evidence of ESL teachers who try out alternative assessment shows that students' performance in authentic speaking and writing tasks is well below their expectations. Some teachers also claim that most students are not satisfied with the score or grade they receive for an assignment. The present study thus attempted to address this problem by studying the effects of provision of rubrics used for assessing to students before they do their assessment tasks. The study was done using secondary school students in a government school in Sri Lanka. A questionnaire was administered to students and a sub-sample of students and the respective teachers were interviewed to collect data on their perceptions of provision of rubrics prior to authentic assessment. Students' performance at pre-test (without knowledge of rubrics) was compared with that of the post-test (with prior knowledge of rubrics). The performance of students was also compared according to the level of exposure to rubrics.

Key Words: rubrics; alternative assessment; authentic assessment; ESL, task performance

Introduction

The term 'rubrics', has been used in the field of assessment for more than two decades now. However, the meaning of the term is not clear to many teachers and they misunderstand that a rubric is any scoring criterion. How does a rubric differ from general scoring criteria? Andrade (2000) defines a rubric as 'a scoring tool that lists the criteria for a piece of work' and one that 'articulates gradations of quality for each criterion, from excellent to poor' (p.1). According to Andrade, rubrics are useful for teachers and learners for five reasons:

1. They are useful as tools of teaching and assessment.

2. They help students become more thoughtful judges of the quality of their work and those of their peers.

3. They save teachers' time spent on evaluating student work.

4. They accommodate heterogeneous classes.

5. They are easy to use.

Assessment in schools and higher education institutions is facing a shift from traditional testing practices towards more authentic assessment of students' learning (Dochy, Gijbels, & Segers, 2006). In authentic assessment, the test-takers are required to demonstrate that they are able to complete a particular task that resembles something that they are likely to have to perform in the target situation (Mueller, 2003; O'Malley & Valdez Pierce, 1996). As rubrics are often used to assess students' performance in such authentic tasks, Jonsson and Svingby (2007) stress that "effective design, understanding and competent use of rubrics is crucial, no matter they are used for high-stake or classroom assessments" (p.131).

A major problem that arises when carrying out authentic assessment is the maintenance of validity and reliability (Jonsson & Svingby, 2007). The reliability of a detailed analytic scoring rubric for writing was determined by

East (2009) in a study which used 47 high school students (17-18 year olds) from 11 schools in New Zealand. The participants wrote two argumentative essays in German and these essays were scored by two independent raters using a scoring rubric specifically designed for this task. Inter-rater consensus and inter and intra rater consistency were calculated. The findings revealed that the analytic scoring rubric was highly reliable in measuring writing proficiency of foreign language learners. The placement test and the scoring rubric were found to be useful in differentiating well between ability levels of students. The scoring rubric was specifically designed for German and the reliability was decided based on the marks of two raters. The study discusses the transferability of scoring rubrics across different languages and claims that it is more successful when languages are similar but is less successful when the languages have marked differences.

The importance of rubrics in enhancing students' ability to communicate their ideas effectively, especially in writing, is stressed by Jaidev (2011) and according to Jaidev, 'knowledge of writing rubrics also helps students become more accountable for their own writing, and it allows them to gain a greater sense of ownership of what they have written' (p.1).

Value of instructional rubrics in teaching, learning and assessment is stressed by Andrade (2005), who shows that if carefully designed, rubrics help students to understand the goal of an assignment and support teachers in unbiased grading, giving feedback and assigning more challenging work to students. However, as 'rubrics are not self-explanatory' teachers need to explain those to students (Andrade, 2005, p.29). A study by Schafer, Swanson, Bené and Newberry (2001) investigated the effects of teacher knowledge of rubrics on students' achievement in biology and algebra and showed that when teachers clearly explain the expected achievement levels in their instruction it leads to higher performance on tests. The importance of students' understanding of rubrics is stressed again in a study by Andrade, Du and Wang (2008) which tested the effects of a scoring rubric on self-assessment of students' writing. The sample consisted of 116 volunteer participants in seven public elementary schools in the northeastern United States. The study found that the quality of students' writing was positively related to their use of a rubric for self-assessment. It stressed the value of involvement of students in generating criteria for assessment using model papers. The studies reported above show positive effects of rubrics on student's performance.

As Andrade, Du and Wang (2008) point out, there is a dearth of studies done in the area of rubrics and hence there is 'limited empirical evidence' to support the claim that rubrics can promote learning and achievement. The number of empirical studies that can be found is limited to three (Andrade, 2001; Cohen, Lotan, Scarloss, Schultz & Abraham, 2002 cited in Andrade et al., 2008; Andrade et al.; 2008) and these studies are also limited to the skill of writing. No study to the researcher's knowledge has focused on the effects of knowledge of rubrics on the performance of students' speaking. Hence, empirical research on the effects of the knowledge of rubrics on speech performance is necessary to address this gap in research. Research done in English as a second language contexts is also rare and hence how far the secondary school students of English as a second language could make use of rubrics for self-assessment and the effects of rubrics on the development of language skills, mainly the productive skills i.e. writing and speaking, are worth investigating.

The present study
The objectives of the present study were to determine the effects of rubrics on ESL students' performance in authentic writing and speech tasks, to evaluate the effects of level of exposure to rubrics on students' speech and writing performance and to explore the effects of exposure to rubrics on students' self-assessment.

Research Questions
What impact does the extent of exposure to rubrics have on ESL students' speech and writing performance?
Does their understanding of rubrics have an impact on their self-assessment?

Methodology
The study adopted a pre-post quasi experimental design. The sample consisted of 70 secondary school students in a national school in Sri Lanka. Two classes were selected for this purpose and students and their teacher of English were briefed about the study. A pre-test was conducted which tested students' speech and writing performance prior to their exposure to rubrics (see Appendix C for tasks).
The rubrics for speech and writing were used by two markers who are graduate teachers of English with over 10 years of experience in teaching to score tasks done by a sample of 30 students and inter-rater reliability was calculated. The inter-rater reliability was high for both Speaking and Writing. Pearson's correlation coefficient for Speaking scores was .91 while it was .93 for Writing.
The pre-test was then assessed by the teacher and the students were also asked to self-assess their work. Focus group interviews were held with the two classes and an open-ended questionnaire was administered to check their understanding of criteria for assessing writing and speech. Class A was then given the rubrics used to assess the tasks and the rubrics were explained to students. Students were encouraged to ask questions to clarify their doubts about the rubrics. They were

asked to recall and write down the assessment criteria (rubrics) before they did the tasks. Then the tasks (post-test) were done in class and were evaluated by the teacher and self-assessed by the students.

Class B was also given the set of rubrics prior to their writing and speaking tasks but the rubrics were not explained to them. They were then assigned the same tasks that were given to Class A and were scored by the same teacher using the rubrics and the students self-assessed their work.

Data Analysis

Students' pre-post scores in the two groups were compared using the statistical package SPSS Version 16 and the open-ended questionnaire and Class A students' notes on how they were scored prior to the post-test were analysed for evidence of knowledge of rubrics. Correlation coefficients were calculated for teacher assessment and students' self-assessment. Since the data showed non-normal distribution ($p<0.05$ in Kolmogorov- Smirnov and Shapiro-Wilk tests), non-parametric statistics were used for data analysis.

Findings

As shown in Table 1, Class A (Rubrics+ Explanation) which received more scaffolding on rubrics than did Class B (Rubrics- Explanation) performed better in both writing and speaking tasks at the post test.

significantly higher speaking task performance ($Z= 4.6$, $p<0.01$, Cohen's $d = 1.31$) for the R+E group than that for the R-E group. The effect size Cohen's d between the R+E and R-E groups at pre-test was .4 while it was 1.31 at post-test.

Table 2: Frequency Counts for References to Contents of the Rubric in the Answers to the Open Ended Questionnaire at the Post-Test – Writing

Criterion	Frequency Class A (Rubrics+ Explanation)	Frequency Class B (Rubrics- Explanation)
Relevance and Content	27	7
Organisation	13	5
Vocabulary	8	3
Grammar	28	9
Spelling	12	0
Mechanics	12	4
Presenta-tion-Handwriting	4	9
Presenta-tion-Neatness	16	7
Writing Style	1	0

Table 1: Descriptive Statistics for Task Performance of Class A and Class B

Condition	Task	N	Mean Pre (SD)	Mean Post (SD)	Mean Gain	Cohen's d (Within Subject)
Class A Rubrics+Explanation	Writing	34	10.76 (2.20)	13.82 (2.75)	3.06	1.23
Class B Rubrics-Explanation	Writing	36	9.69 (3.63)	9.88 (2.90)	0.19	0.06
Class A Rubrics+Explanation	Speaking	34	11.97 (3.38)	13.05 (3.69)	1.08	0.31
Class B Rubrics-Explanation	Speaking	36	10.55 (2.88)	8.50 (3.26)	-2.05	-0.66

A Mann-Whitney U test on the pre-test data showed no significant difference between the Rubrics+Explanation (R+E) group and the Rubrics-Explanation (R-E) group in writing task performance ($Z= 1.41$, $p>0.05$, Cohen's d =.36) A Mann-Whitney U test on post-test data showed significantly higher writing task performance ($Z= 4.81$, $p<0.01$, Cohen's $d = 1.39$) for the R+E group than that for the R-E group. The effect size Cohen's d between the R+E and R-E groups at pre-test was .36 while it was 1.39 at post-test.

A Mann-Whitney U test on the pre-test data showed no significant difference between the Rubrics+Explanation group and the Rubrics-Explanation group in speaking task performance ($Z=1.72, p>0.05$ Cohen's $d = .40$). A Mann-Whitney U test on post-test data showed

As answers to the question 'What do you think your teacher takes into consideration when marking your essays?' students listed different criteria and they were categorized under the areas given above. Another question in the open-ended questionnaire asked them 'What do you think are the qualities of an essay which will get the highest marks/lowest marks from the teacher? Given below are some excerpts from students' answers.

If the essay has been written using correct hand writing and language that will be a successful essay. The most important thing is the vocabulary used in that essay. It is necessary to use correct language at the correct place. And also the starting of that essay should be very creative. We should end the essay

creatively too. If the word range is given we should not exceed that range' –Student 1 from Class A

The writer should write only about the topic. He shouldn't write all the things he knows but he should summarize the facts he knows and limit it to the number of words. He should write the sentences grammatically and make meaningful sentences. The writer should think about the content and order and have a good start, body paragraphs and a good ending. The writer's spelling mistakes and mechanisms (mechanics) of writing should be good enough for a perfect essay. –Student 4 from Class A

Study the essay topic well and use correct grammar and write within the given word limit. –Student 6 from Class B

'I think most of the teachers concern more about grammar when marking an essay. Although teachers show concern about grammar think teachers should show concern to the related points than grammar'. –Student 14 from Class B

To the open ended question 'What do you think your teacher takes into consideration when marking your speech?', most of the students in Class A who were given explanation of the rubrics, mentioned most of the criteria in the rubrics while the group which received no explanation had only a few students mentioning the contents of the rubrics.

Table 3: Frequency Counts for References to Contents of the Rubric in the Answers to the Open Ended Questionnaire at Post-Test – Speaking

Criterion	Frequency Class A (Rubrics+Explanation)	Frequency Class B (Rubrics-Explanation)
Content and Relevance	15	8
Structure	11	7
Audibility	20	12
Gestures	9	3
Eye Contact	8	1
Confidence	6	3

Correlation between Students' Self-Assessment and Teacher Assessment

At pre-test, Spearman's correlation co-efficient was calculated for students' average self-assessment scores and teacher's scores. For Rubrics+Explanation group it was .56 and $p<.01$ for writing and .27 and $p=.12$ for Speaking. For Rubrics-Explanaton group, it was .11 and $p=.56$ for writing and .03, $p=.85$ for speaking.

At post-test, Spearman's correlation co-efficient for R+E group writing was .88, $p<0.01$ while it was .76, $p<0.01$ for speaking. For R-E group, correlation co-efficient was

.03 and $p>0.05$ for writing while for speaking it was .30, $p>0.05$.

Discussion and Conclusion

The data analysis given above showed that the group which received rubrics plus explanation of what is meant by each criterion and how the grading is done performed significantly better than the group which received the rubric without any explanation. The R+E group showed a high positive effect of the knowledge of rubrics on students' speaking and writing performance at post-test. As Andrade (2005) points out, 'rubrics are not self-explanatory. Students need help in understanding rubrics and their use'. The findings of the present study support this view. The teachers who were interviewed also claimed that students (especially the low achievers) do not show any interest in any novel practice like using rubrics for self-assessment unless they are specifically trained to do so. According to the teacher who taught the R+E group, the students showed high interest in using the rubrics and they seemed to be aiming for the highest level in the rubric. This shows that rubrics help students in goal-setting and planning which are crucial metacognitive strategies (Anderson, 2003) which support students' learning. The decrease in the mean scores for speaking in the R-E group at post-test may be a result of overuse of rubrics by some of the students in the group. As the teacher pointed out, some students overused gestures and their eye contact was not natural and some tried to speak very loudly.

Another important finding of the study was that students' self-assessment which was far different from teachers' assessment of their work at the pre-test, changed considerably in the R+E group at the post-test. The self-assessed scores of this group which received detailed explanation of the rubrics showed a high correlation with the scores of the teacher at the post-test. The R+E groups' written products and speech had evidence of their attempts to use the rubrics given and the group became more motivated and more focused than did the members of the R-E group in their activities in the English as a Second Language classroom. This finding supports the findings of Andrade et al. (2008) which claim that students' self-assessment using a rubric leads to improvement in writing. Hence, the study shows that careful designing of rubrics and thorough explanation and constant scaffolding and guidance given to the users of rubrics are necessary if they are to bring positive outcomes in teaching, learning and assessment. However, there are some limitations to this study. The study included only one secondary school and the participants were all girls and hence the influence of gender could not be gauged. Students' performance was measured based on a single type of task (an informal letter for writing and a public speech). Further research should focus on other language skills and different types of tasks. Longitudinal studies with

students at primary and tertiary levels and at different achievement levels would also be of utmost importance. Such studies with diverse populations and in different subjects and skill areas are necessary to support the claim that knowledge of rubrics and using rubrics for self-assessment have the potential to improve learning and achievement.

References

1. Anderson, N. J. (2003). Metacognition in writing: Facilitating writer awareness. In A.Stubbs & J.Chapman (Eds.), *Rhetoric, uncertainty, and the university as text: How students construct the academic experience.* (pp. 10-30). Regina: Canadian Plains Research Center, University of Regina.
2. Andrade, H.G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5): 13-18.
3. Andrade, H.G. (2005). Teaching with rubrics. C*ollege Teaching, 53*(1) 27-30.Retrieved from http://www.uri.edu/assessment/uri/guidance/documents/Andrade_2005_Teachingwithrubrics.pdf on 17th May 2013.
4. Andrade, H.L., Du. Y., & Wang, X. (2008). Putting rubrics to the test: the effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice 27*(2): 3-13.
5. Brown, H. D. & Abeywickrama, P. (2010). *Language assessment, principles and classroom practices* (2nded.). White Plains, NY: Pearson Education, Inc.
6. Dochy, F., Gijbels, D., & Segers, M. (2006). Learning and the emerging new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts & S.Vosniadou (Eds.), *Instructional psychology: Past, present and future trends* (pp.191-206). Oxford, Amsterdam: Elsevier.
7. Jaidev. R. (2011). Rubrics-Based Writing: Liberating Rather Than Restricting in Many Contexts. *ELTWO Online.* Retrieved on 16th May 2013 from http://blog.nus.edu.sg/eltwo/2011/03/17/rubrics-based-writing-liberating-rather-than-restricting-in-many-contexts/
8. Jonsson, A., & Svingby, G. (2007). The use of rubrics: Reliability, validity and educational consequences. *Educational Research Review 2*: 130-144.
9. O'Malley, J.M., & Valdez, P.L. (1996). *Authentic Assessment for English Teachers: Practical approaches for teachers.* London: Addison-Wesley.
10. Schafer, W., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education, 14*(2), 151-170.

Appendix A: Rubrics for Assessing Writing

	4	3	2	1	0
Relevance and Content	Fully satisfies the requirements of the given task. Includes all relevant information.	Mostly covers the requirements of the given task. Includes most of the relevant information.	Addresses some of the requirements. Includes some relevant information but not clearly focused.	Attempts to address the topic but few relevant information. Digresses often from the topic.	Does not attempt the task/ the answer is completely irrelevant.
Organisation	Includes an inviting introduction and a satisfactory conclusion. Skillfully manages paragraphing. Logical arrangement of ideas. Manages all aspects of cohesion well.	Includes an introduction, body and conclusion. Uses paragraphing successfully. Uses a range of cohesive devices but may look mechanical.	Attempts to include an introduction, body and conclusion. Main idea is not clearly supported with details. Less attention given to organization. Rare use of transitions.	Begins abruptly. No paragraphing or inappropriate paragraphing. No attempt to maintain logical arrangement of ideas.	No clear message is communicated.
Vocabulary and Word Choice	Uses a sophisticated range of vocabulary which is appropriate for the purpose and audience. May use figurative language.	Uses accurate vocabulary which suits the audience and purpose with a mixture of precise and general words. Occasional errors in word choice.	Uses a fair range of vocabulary to express ideas. May be inappropriate for the audience and purpose at some occasions.	Uses a limited range of vocabulary. Mostly inaccurate for the purpose and audience.	Inappropriate and inaccurate vocabulary.
Grammar, Spelling and Punctuation	Uses a variety of grammatically correct sophisticated sentence structures. Perfect spelling and accurate punctuation.	Uses mostly correct sentence structures with 1-2 mistakes. Correct spelling and appropriate punctuation with occasional errors.	Uses basic sentence structures with some errors. Uses spelling and punctuation with some errors.	Uses mostly sentence fragments with frequent errors. Frequent errors in spelling and punctuation.	Writing incomprehensible.
Presentation	Neat, easy to read, error free	Mostly readable, neat, minimum errors	Fairly readable. Some strain on the reader.	Not clear. Considerable strain on the reader.	Illegible.

Appendix B: Rubrics for Assessing Speech

	4	3	2	1	0
Content and Relevance Support Focus	Fully satisfies the requirements of the task. Understands the audience. Supports the points with enough examples. Clear focus is maintained throughout the speech.	Mostly covers all the requirements of the task. Understands the audience. Supports most of the points with examples. Clear focus.	Addresses some of the requirements of the task. Some points are supported with examples. Some digression from the topic.	Attempts to address the requirements but few relevant information. Digresses often from the topic. Structure not clear.	Does not attempt the task/ Insufficient and irrelevant information
Structure	Clear structure starting with an attention getter, introduction to the topic, body which supports the topic and a conclusion which summarises the main points.	Uses an introduction, body with some supporting examples, and conclusion but there may not be an interesting attention getter.	Attempts to structure with an introduction but fails to maintain structure.	Attempts an introduction. No clear structure.	No structure
Delivery Eye contact Audibility Pace Gestures	Maintains good eye contact. Speaks with appropriate volume and pace. Uses appropriate gestures.	Maintains eye contact. Audible. Uses appropriate gestures.	Audible but there may be little eye contact.	May have considerable strain on the listener. Lot of pausing and hesitation.	Speech incomprehensible
Word Choice	Word choice appropriate to the audience.	There may be occasional inappropriacies in word choice.	Some inappropriacies and inaccuracies in word choice.	Inappropriate and inaccurate word choice.	Totally inappropriate.
Preparedness and confidence	Appeared well prepared, confident and comfortable.	Seemed fairly comfortable and confident but needs a bit more practice.	Seems somewhat prepared but needs more practice to build confidence	Inadequate preparation. Needs practice to build confidence.	Not prepared

Appendix C: Tasks –Post-test

Writing Task

Your friend who left school three years ago and migrated to another country is interested in what happened in your school during the past three years. Write him/her a letter describing
- the present situation in the school
- events that took place during the past three years
- asking information about his/ her experience abroad

Your letter should be limited to 300 words and follow the appropriate format.

Speaking Task

The dengue menace is on the increase with the monsoon rains. Give a speech at the school assembly on the Role of School Children in Prevention of Dengue Fever. Your speech should be 5 minutes long and it should include
- a description of the problem
- steps that can be taken in prevention and control
- role of the school in prevention and control of Dengue fever.

About the Author

Radhika De Silva is a Senior Lecturer in the Department of Language Studies at the Open University of Sri Lanka. She obtained her PhD from University of Reading, UK. Her research interests include writing strategy instruction, EAP, language teacher education and testing and assessment.