

ASSESSING PERFORMANCE USING SCORING RUBRICS: AN OVERVIEW

Made Frida Yulia

Universitas Sanata Dharma Yogyakarta
Mrican, Tomol Pos 29, Yogyakarta, Telp. (0274) 513301

ABSTRACT

One of the strategies of Contextual Teaching and Learning is authentic assessment, which can provide valid and accurate information about what students really know and are able to do, or about the quality of educational programs. One of its forms is performance assessment. It requires students to demonstrate the application of knowledge to a particular context. The prevailing philosophy of performance assessment is that it provides ways to improve achievement, demonstrate exactly what a student does or does not understand, relate learning experiences to instruction, and combine assessment with teaching. Despite the benefits, some teachers are hesitant to implement it because they do not have enough knowledge about how to assess a student's performance properly. Since in performance assessment there are degrees to which a student is successful or unsuccessful, the performance should be evaluated in such a way that it allows teachers to consider those degrees, that is by creating rubrics. This paper attempts at discussing basic steps in implementing performance assessment in general and developing scoring rubrics in particular.

Key words: *performance, assessment, scoring rubric, analytic, and holistic*

1. Introduction

One of the strategies of Contextual Teaching and Learning is the adoption of authentic assessment. The aim of authentic assessment is to provide valid and accurate information about what students really know and are able to do, or about the quality of educational programs. Hence, it can be used to monitor students' progress and inform teaching practices.

Authentic assessment provides students with opportunities and direction for improvement. Moreover, it authentically allows a student to demonstrate his ability to perform tasks, solve problems, or expressing know-

ledge in ways that simulate situations which are found in real life.

Authentic assessment includes project, product, peer, self, and performance assessment. This paper concentrates on performance assessments and one way to assess them; that is, using scoring rubrics.

2. Performance Assessments

As mentioned previously, one form of authentic assessment is performance assessment. According to Wiggins (1993) (in <http://oregonstate.edu/instruction/ed555/zone5/zone5hom.htm>), performance assessment is developed to 'test' the ability of students to

demonstrate their knowledge and skills (what they know and can do) in a variety of 'realistic' situations and contexts. In a word, it requires students to demonstrate the application of knowledge to a particular context.

Tasks used in performance-based assessment include essays, oral presentation, open-ended problems, hands-on problems, real world simulations and other authentic tasks. Such tasks are concerned with problem solving and understanding.

The prevailing philosophy of performance assessment is that it provides ways to improve achievement, demonstrate exactly what a student does or does not understand, relate learning experiences to instruction, and combine assessment with teaching. Seeing the benefits, however, some teachers are hesitant to implement it because they do not know enough about how to fairly assess a student's performance.

There are some basic steps to be taken when developing performance assessment. An initial step in planning performance assessment is defining the purpose of the assessment. To help teachers do it, Stiggins (1994 in Moskal, 2003) suggests several questions as guidelines:

- 1) What concept, skill, or knowledge am I trying to assess?
- 2) What should my students know?
- 3) At what level should my students be performing?
- 4) What type of knowledge is being assessed: reasoning, memory, or process?

The answers to the above questions will guide the teachers to do the second step, selecting an appropriate assessment activity. Some considerations prior to this are time constraints, availability of resources in the classroom, and how much data is necessary in order to make an informed decision about the quality of a student's performance.

The selected performance should reflect a valued activity or resemble the type of activities that are known to take place in the work-

place. Additionally, the completion of performance assessments should provide a valuable learning experience; that is, an increase in the teacher's understanding of what students know and can do and an increase in the students' knowledge of the intended content and constructs.

Next, teachers should define the criteria, i.e., elements of the project/task that determine the success of a performance. Having clearly defined criteria makes it easier for the teachers to remain objective during the assessment. They will know exactly which skills or concept they are supposed to be assessing.

In defining criteria (Airasian (1991) in Brualdi, 1999), teachers should list the important aspects of the performance to be assessed. They should try to limit the number of performance criteria, so they can all be observed during a student's performance. If possible, then, they ask groups of other teachers think through the important behaviors included in a task. It is worth remembering that the performance criteria must be expressed in terms of observable behaviors and must be arranged in the order in which they are likely to be observed. Additionally, ambiguous words that cloud the meaning of the performance criteria should be avoided. Accordingly, teachers must use specific and clear language that the students understand.

Sometimes the development of assessment criteria may involve students. It has an advantage of making the students think about the criteria and create the language for the scoring guide. Also, students interpret the criteria and put it in a language that is meaningful to them. Nevertheless, though beneficial, such a process is time consuming (<http://oregonstate.edu/instruction/ed555/zone5/zone5hom.htm>).

In performance assessments, there is no clear-cut right or wrong answers. Instead, there are degrees to which a student is successful or unsuccessful. Thus, the performance needs to be evaluated in a way that allows

teachers to take those varying degrees into account; that is, by creating rubrics (Moskal, 2003). The results of the performance assessment are recorded using *Rating Scale Approach*; teachers assign a numerical scale to indicate to what degree the standards are met.

3. Scoring Rubric: What, When and Why

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 (Brookhart, 1999). They consist of specific pre-established performance criteria, used in evaluating student work on performance assessments (Mertler, 2001). Using rubrics, teachers can define the different levels of proficiency for each criterion. The pre-defined scheme will also reduce the subjectivity in scoring.

Scoring rubrics are typically employed when a judgement of quality is required and may be used to evaluate a broad range of subjects and activities. They may be used to evaluate writing samples, group activities, extended projects and oral presentations. Where and when a scoring rubric is used does not depend on the grade level or subject, but rather on the purpose of the assessment.

Scoring rubrics contribute two benefits in the evaluation process. First, they support the evaluation of the extent to which criteria has been met because they are based on descriptive scales. Second, they provide feedback to students concerning how to improve their performances (Moskal, 2003). Since there is a description at each level as to what is expected as well as numerical values, students are enabled to understand why they receive the score and what they need to do to improve their future performance.

4. Developing Scoring Rubrics

Mertler (2001) proposed a step-by-step procedure in developing scoring rubrics to assess performance.

Step 1:	Re-examine the learning objectives to be addressed by the task. This allows us to match our scoring guide with our objectives and actual instruction.
Step 2:	Identify specific observable attributes that we want to see (as well as those we do not want to see) our students demonstrate in their product, process, or performance. Specify the characteristics, skills, or behaviors that we will be looking for, as well as common mistakes we do not want to see.
Step 3:	Brainstorm characteristics that describe each attribute. Identify ways to describe above average, average, and below average performance for each observable attribute identified in Step 2.
Step 4a:	For holistic rubrics, write thorough narrative descriptions for excellent work and poor work incorporating <u>each attribute</u> into the description. Describe the highest and lowest levels of performance combining the descriptors for all attributes.
Step 4b:	For analytic rubrics, write thorough narrative descriptions for excellent work and poor work for <u>each individual attribute</u> . Describe the highest and lowest levels of performance using the descriptors for each attribute separately.
Step 5a:	For holistic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work <u>for the collective attributes</u> . Write descriptions for all intermediate levels of performance.
Step 5b:	For analytic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work <u>for each attribute</u> . Write descriptions for all intermediate levels of performance for each attribute separately.
Step 6:	Collect samples of student work that exemplify each level. These will help we score in the future by serving as benchmarks.
Step 7:	Revise the rubric, as necessary. Be prepared to reflect on the effectiveness of the rubric and revise it prior to its next implementation.

5. Types of Scoring Rubrics

There are two types of scoring rubrics, holistic and analytic. A **holistic rubric** requires the teacher to score the overall process or product as a whole, without judging the component parts separately (Nitko (2001) as cited in Mertler, 2001). It focuses on the overall quality, proficiency or understanding of the specific content and skills. It is appropriate when there is no definitive correct answer in the performance tasks and they require students to create some sort of response. Besides, it is suitable when the purpose of the performance assessment is summative in nature. The scoring process using such a rubric is faster, but the feedback provided is limited.

Table 1: Template for Holistic Rubrics

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.

In contrast, an **analytic rubric** enables teachers to score separate, individual parts of the product or performance first, then sum the individual scores to obtain a total score (in Mertler, 2001). It is preferred when a fairly focused type of response is required. It means there may be one or two acceptable responses and creativity is not an essential feature of the students' responses. The advantage of using this type of rubric is the degree of feedback to students and teachers is significant. However, the drawbacks are it is time consuming and makes the scoring process substantially slower.

Regardless of which type of rubric is selected, an initial step to development is specific performance criteria and observable in-

dicators must be identified. Teachers then must consider whether to use holistic or analytic rubric, by taking the purpose of the assessment into account. The choice has some implications. Teachers must consider first how they intend to use the results. If an overall, summative evaluation is desired, a holistic scoring approach would be more desirable. Conversely, if formative feedback is the goal, an analytic scoring rubric should be used. Other implications include the time requirements, the nature of the task itself, and the specific performance criteria being observed (Mertler, 2001).

Table 2: Template for Analytic Rubrics

	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	

(Mertler, 2001)

6. Issues of Validity and Reliability

Validity refers to the degree to which the evidence supports that these interpretations are correct and that the manner in which the interpretations used is appropriate (Moskal and Leydens, 2001). They state further that validity concerns the following evidences:

- a. *Content-Related Evidence*
It refers to the extent to which a student's responses to a given assessment instrument reflects his knowledge of the content area that is of interest. Content-related evidence is also concerned with the extent to which the assessment instrument adequately samples the content domain.
- b. *Construct-Related Evidence*
Construct-related evidence is the evidence supporting that an assessment instrument is completely and only measuring the intended construct. Many different facets of the constructs should be identified and carefully considered so as to be displayed and provide convincing evidence of the students' underlying processes.
- c. *Criterion-Related Evidence*
It is evidence that supports the extent to which the results of an assessment correlate with a current or future event.

To ensure validity, the teacher should use the stated purpose and objective to guide

the development of the scoring rubric. In order not to be misled, they need to state the purpose and objective of the assessment clearly and develop scoring criteria that address each objective.

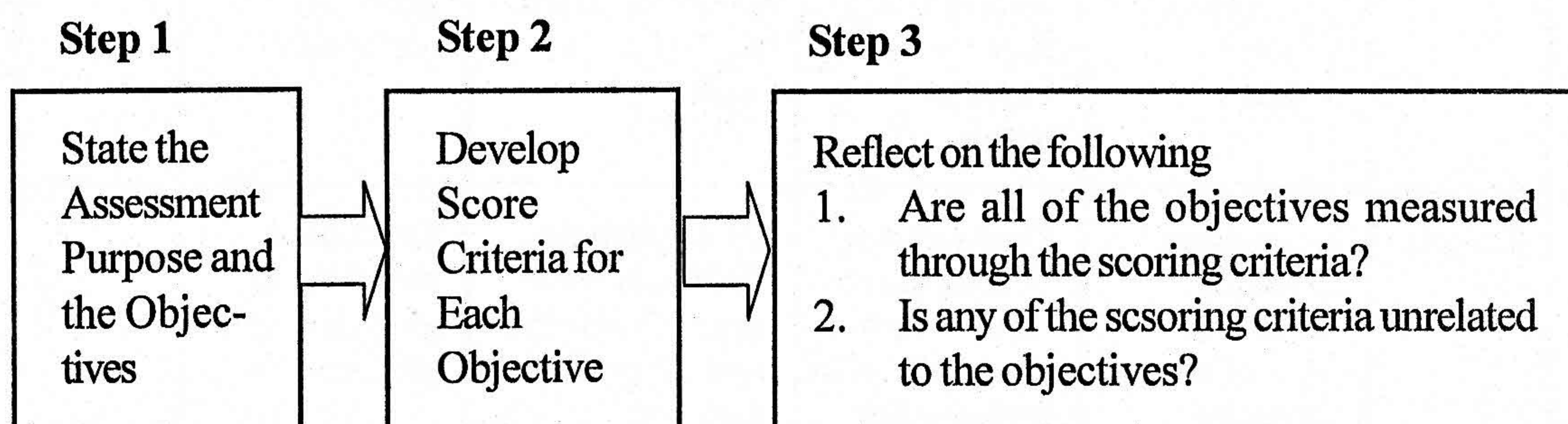
The following chart describes the general steps taken to ensure validity (Moskal and Leydens, 2001).

Specifically, the teacher must then choose which forms of evidence: content, construct, or criterion, is to be given consideration, in accordance with the purpose and objective of the assesment.

The list of questions below are to examine each type of validity evidence (Moskal and Leydens, 2001).

Reliability refers to the consistency of assessment scores regardless of when the student completes the assessment, when the response is scored, and who scores the response. Reliability in rubric development refers to rater reliability; i.e., the consistency of scores that are assigned by two independent raters and that are assigned by the same rater at different points in time (Moskal and Leydens, 2001). It is of two kinds:

- 1) *Interrater reliability*
It refers to the concern that a student's score may vary from rater to rater. Since the criteria are formalized at each score level, scoring rubric can reduce the occurrence of the discrepancies or variations between raters.



Content	Construct	Criterion
<ol style="list-style-type: none"> 1. Do the evaluation criteria address any extraneous content? 2. Do the evaluation criteria of the scoring rubric address all aspects of the intended content? 3. Is there any content addressed in the task that should be evaluated through the rubric, but is not? 	<ol style="list-style-type: none"> 1. Are all of the important facets of the intended construct evaluated through the scoring criteria? 2. Is any of the evaluation criteria irrelevant to the construct of interest? 	<ol style="list-style-type: none"> 1. How do the scoring criteria reflect competencies that would suggest success on future or related performances? 2. What are the important components of the future or related performance that may be evaluated through the use of the assessment instrument? 3. How do the scoring criteria measure the important components of the future or related performance? 4. Are there any facets of the future or related performance that are not reflected in the scoring criteria?

2) **Intrarater reliability**

The scoring process of a given rater changes because of internal influence of the rater such as fatigue and mood. Scoring rubrics help establish a description of the scoring criteria in advance, which can be revisited to maintain the consistency.

To improve the two kinds of reliability above, we need to clarify the scoring rubrics. The questions below serve to check the appropriateness of the designed scoring rubrics:

- 1) Are the scoring categories well defined?
- 2) Are the differences between the score categories clear?
- 3) Would two independent raters arrive at the same score for a given response based on the scoring rubric?

If the answer to any of these questions is “no”, then the unclear score categories should be revised.

Teacher might also use anchor papers (a set of scored responses that illustrate the nuances of the scoring rubric), then ask other teachers to try out the scoring rubric. Any differences in interpretation and adjustments should be negotiated. Whenever possible, the rubric should be shared with the students in

advance in order to allow them the opportunity to construct the response with the intention of providing convincing evidence that they have met the criteria. It is also helpful for the students because they know what is expected of them (Moskal and Leydens, 2001).

7. Application in Speaking Assessment

Traditionally, speaking skills is associated with public speaking. Recently, however, definitions of speaking have been expanded (Crown (1981) in Mead and Rubin, 2003). One trend has been to focus on communication activities that reflect a variety of settings: one-to-many, small group, one-to-one, and mass media. Second trend has focused on using communication to achieve specific purposes: to inform, to persuade, and to solve problems. A third trend has focused on basic competencies needed for everyday life (giving directions, asking for information, or providing basic information in an emergency situation, etc.). Many of these broader views stress that oral communication (speaking) is an interactive process in which an individual alternatively takes the role of speaker and listener and hence involves comprehension and production (Hughes, 1989:101).

There are two methods for assessing speaking skills. The first is *Observational approach*, in which students' behavior is observed and assessed unobtrusively. The second is *Structured approach*. In this approach students are asked to perform one or more specific oral communication tasks, and then the performance on the task is evaluated. Students may perform in a one-on-one setting or in a group or class setting. In either setting, students should feel that they are communicating meaningful content to a real audience.

Both approaches use a variety of rating systems (<http://ericae.net/edo/ED263626.htm>). A holistic rating captures a general impression of the students' performance. Analytic scales capture the students' performance on various aspects of communication, such as delivery, organization, content, and language.

The methods used for assessing speaking skills depend on the purpose of the assessment. Thus, in the first place, teacher needs to define the purpose of the assessment. As it is about speaking assessment, the purpose is to find out whether students have the ability to interact successfully in the language.

Having defined the purpose, teacher may then select an appropriate assessment activity (the task). Some possible formats of performance assessment for speaking are role-play and discussion. Tasks should focus on topics that all students can easily talk about or at least they should be given an opportunity to collect information on the topic. In line with the purpose stated before, say, the teacher decides to have role-play as the performance to be assessed. If that is the case, it means that the teacher employs structured approach.

Before going into defining the criteria, the teacher must decide whether he will use holistic or analytic rubric. In order to provide significant feedback for students' improvement in their future performances, the teacher selects analytic rubric. It will assess various aspects of students' performance, such as deliv-

ery, organization, content, and language

Afterwards, the teacher defines the performance criteria and creates the scoring rubric. The following is an example.

Proficiency Descriptions

Accent

1. pronunciation frequently unintelligible
2. frequent gross errors and a very heavy accent make understanding difficult, require frequent repetition
3. 'foreign accent' requires concentrated listening, and mispronunciation lead to occasional misunderstanding and apparent errors in grammar or vocabulary
4. marked 'foreign accent' and occasional mispronunciations which do not interfere with understanding
5. no conscious mispronunciations, but would not be taken for a native speaker
6. native pronunciation, with no trace of foreign accent

Grammar

- 1 grammar almost entirely inaccurate phrases
- 2 constant errors showing control of very few major patterns and frequently preventing communication
- 3 frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding
- 4 occasional errors showing imperfect control of some patterns but no weakness that cause misunderstanding
- 5 few errors, with no patterns of failure
- 6 no more than two errors during the interview

Vocabulary

- 1 vocabulary inadequate for even the simplest conversation
- 2 vocabulary limited to basic personal and survival areas (time food, transportation,

- family, etc.)
- 3 choice of words sometimes inaccurate; limitations of vocabulary prevent discussion of some common professional and social topics
 - 4 professional vocabulary adequate to discuss special interests, general vocabulary permits discussion of any non-technical subject with some circumlocutions
 - 5 professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations
 - 6 vocabulary apparently as accurate and extensive as that of an educated native speaker

Fluency

- 1 speech is so halting and fragmentary that conversation is virtually impossible
- 2 speech is very slow and uneven except for short or routine sentences
- 3 speech is frequently hesitant and jerky; sentences may be left uncompleted
- 4 speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words
- 5 speech is effortless and smooth, but perceptibly non-native in speech and evenness

- 6 speech on all professional and general topics as effortless and smooth as a native speaker's

Comprehension

- 1 understands too little for the simplest type of conversation
- 2 understands only slow, very simple speech on common social and touristic topics; requires constant repetition and rephrasing
- 3 understands careful, somewhat simplified speech when engaged in a dialogue but may require considerable repetition and rephrasing
- 4 understands quite well normal educated speech when engaged in a dialogue, but requires occasional repetition and rephrasing
- 5 understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech
- 6 understands everything in both formal and colloquial speech to be expected of an educated native speaker

(Adapted from Hughes, 1989: 113)

Scoring Rubric							
Name: _____							
Class: _____							
	Scale						
Accent	1	2	3	4	5	6	
Grammar	1	2	3	4	5	6	
Vocabulary	1	2	3	4	5	6	
Fluency	1	2	3	4	5	6	
comprehension	1	2	3	4	5	6	
total							

8. Conclusion

Performance assessment can be assessed using scoring rubrics. They can define different levels of proficiency and the use of such pre-defined performance criteria may reduce the subjectivity in scoring. Since they are based on descriptive scales, scoring rubrics can support the evaluation of the extent to which criteria has been met. Also, the feedback they provide is significant, which can help students improve their future performances.

Different types of scoring may be used, analytic or holistic, depending on the purpose of the assessment. Whichever type is used, it should adhere to the measurement principles of reliability and validity.

Following the discussion of implementing performance assessment and developing scoring rubrics provided in this paper, it is hoped that teachers will gain benefits and will not be hesitant anymore to implement performance assessment and use scoring rubric to assess it.

BIBLIOGRAPHY

- Brualdi, Amy. 1999. *Implementing Performance Assessment in the Classroom*. <http://edresearch.org/pare/getvn.asp?v=5&n=7> (accessed on September 27, 2003)
- Hughes, Arthur. 1989. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Mead, Nancy A and Donald L. Rubin. *Assessing Listening And Speaking Skills*. <http://ericae.net/edo/ED263626.htm> (accessed on September 27, 2003)
- Mertler, Craig A. 2001. *Designing Scoring Rubrics for Your Classroom*. <http://edresearch.org/pare/getvn.asp?v=7&n=25> (accessed on September 27, 2003)
- Moskal, Barbara M. 2002. *Recommendations for Developing Classroom Performance Assessment and Scoring Rubrics*. <http://edresearch.org/pare/getvn.asp?v=8&n=14> (accessed on September 27, 2003)
- Moskal, Barbara M. 2003. *Scoring Rubrics: What, When And How?* <http://edresearch.org/pare/getvn.asp?v=7&n=3> (accessed on September 27, 2003)
- Moskal, Barbara M. and Jon A. Leydens. 2001. *Scoring Rubric Development: Validity and Reliability*. <http://edresearch.org/pare/getvn.asp?v=7&n=10> (accessed on September 27, 2003)
- _____. *What Is Contextual Teaching And Learning?* <http://www.cew.wisc.edu/teachnet/ctl> (accessed on December 19, 2003)
- _____. *Assessing Student Performance and Understanding*. <http://oregonstate.edu/instruction/ed555/zone5/zone5hom.htm> (accessed on September 27, 2003)
- _____. *Alternative/Performance-Based Assessment*. [http://www.emtech.net/Alternative Assesment.html](http://www.emtech.net/Alternative_Assesment.html) (accessed on September 27, 2003)