Rubrics Revisited

Ali A. Al-Rabai, PhD

Abstract

Measuring teaching effectiveness and or evaluating students' achievement are basic process to Educators and students alike. Henceforth, tests, portfolios, reflections, interviews, presentations, debates, project and so on and so forth are all means of measurement and evaluation that have been used by schools and universities all over the world. Some major concerns are common to all means of measurement and evaluation such as validity and reliability. Validity entails that the evaluation or assessment measure is sound and clear and that it tests, evaluates or assesses what it is meant to test, evaluate or assess. Therefore, recent trends have been emphasizing the importance of Rubrics for the validity of the evaluation, assessment or measurement process. A Rubric 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, Stiggins 2001, Arter and Chappuis 2007). Or it can be defined as " a descriptive tool for determining the level of performance or quality of a piece of work" (http://www.nuigalway.ie/celt/teaching_and_learning/Rubrics). The most important criteria for Rubrics is: sound evaluation Criteria, quality definition and the a scoring Strategy (Popham 1997). Yet, a crucial element in a good rubric is the language appropriateness. The language of Rubrics has been left with only a minimal research interest although the clarity of rubrics is the most important characteristic for its comprehension and application (Tierney and Simon 2004, Moni, Beswick, and Moni 2005, Payne 2003, Lapsley and Moody 2007).

The researcher, although a good supporter of the use of rubrics for quality education and better instruction, believes that the Art of Making Rubrics requires vigilant linguistic knowledge and expert expertise that can guarantee rubric validity and accountability.

Keywords: Rubrics, validity, reliability, accountability, clarity, quality, assessment, measurement, evaluation.

Introduction

I was first introduced to rubrics in 2008. Although I have to admit that I was initially skeptical of their use and value, I gradually have become rather enthusiastic on the idea as I have built a better awareness and henceforth practice on how to make rubrics and how to use them significantly. As a university TESOL instructor, I have to coach, test, and assess a multitude of language skill-based assignments that come in various shapes and forms corresponding to the various English language skills that are taught in my classrooms. Writing in particular has always been my most troubling area when evaluation and grading are concerned as I have always found myself going subjective in evaluating my students' work. Rubrics were so helpful in my EFL career over the past five years. I have learnt how to develop rubrics and how to use them for accurate, fair and sound grading and evaluation. Recently, I have also started using instructional rubrics for developing my teaching and instruction techniques (Andrade 2001).

With practice and daily use, rubrics have become more and more helpful and valuable to my career. I started using them for all sorts of teaching activities and evaluation processes. Quite recently, I was introduced to rubrics for accreditation and evaluation of large-scale projects and programs.

What remains of a particular interest to me is the language of a rubric. Language is vitally important in the process of making a rubric. I have seen so many samples and examples of rubrics that can hardly be applied in a fair and consistent manner. Word choice, in particular, can strongly damage the accuracy and accountability of a rubric.

It is common practice for teachersto plan their lessons, units and programs at large. Classroom sessions begin with specific objectives and goals to be attained. However, there has always been a great deal of uncertainty when it comes to making sure that the plans are accurately applied and the objectives are satisfactorily achieved. Tests, quizzes, direct questions and many other techniques have been used to evaluate our students' achievement and the quality of our teaching. However, in grading our students' work we have been usually accused of inconsistency or inequity. Similarly, we have always found out that one or more of our objectives are not assessed or catered for in the general evaluation process. These facts have been a source of trouble for educators, assessors and evaluators for years. How can a teacher of Essay Writing class that has 40 students keep fair and consistent in grading his students work? How can one, two or even more exams cover all course objectives and instructional outcomes? What can students' results in tests or assignments tell us about the students and about ourselves? What can grades say about our instruction and teaching practices? Such questions and more could not be answered in a traditional testing or evaluation process. Rubrics give answers or possibilities.

Rubrics, if well-constructed, can say something about whether the class objectives have been met or not; they can give feedback on students' particular areas of weakness ; they can tell the teacher something about his/her instruction effectiveness; they also can give the teacher hints on what areas need more emphasis and attention. Rubrics can also guarantee quite a good level of fairness in grading students' work and a more sound judgment on students' abilities.

The Literature Review

Although rubrics have been around for relatively a short time, the research efforts have touched on major issues such as the effectiveness of rubrics (Andrade and Du 2005), their validity and reliability (Moskal and Leydens 2000,Popham 1997, Jonsson and Svingby 2007), their value as indicators of students' progress and achievement (Andrade 2001,Simon and Forgette-Giroux 2001) and their power in developing teachers' teaching effectiveness (Andrade 2000, Parkes 2006). These researchers and many others have investigated the value, use and accountability of rubrics to the instruction process, to the instructors, to the students and to the educational institution in general. Some have looked at rubrics from the students' point of view and have found out that students' motivation is increased and their achievement is fostered when students knew what was expected of them through clear and well-constructed rubrics (Stiggins 2001, Andrade and Du 2005). The fact that rubrics have the power of increasing students' interest and motivation have also been investigated by researchers in various educational disciplines (Roblyer&Wiencke 2003, Reddy& Andrade 2010, Blacka&Wiliam 1989). In almost all cases, students show a rather higher level of motivation and involvement when they either participate in defining learning goals and objectives or in the process of assessment including knowing the criteria for good and weak performance.

Other researchers have studied the validity and accountability of rubrics and concluded that in most cases assessment based on well-designed rubrics has shown more consistency and fairness in the evaluation process and more accountability (Hafner&Hafner 2003, Andrade 2005). Some researchers, however, have found some evidence that supports the negative effect of ill-formed rubrics particularly when issues of validity and reliability were present (payne 2003, Andrade 2005). These researchers place high emphasis on rubric validity, clarity and reliability. They emphasize the fact that rubrics need to be made easy for students' comprehension so that students feel confidently aware of the value of using rubrics for assessment. Further, they encourage negotiating rubrics with colleagues, experts and with the students themselves. When rubrics are incomprehensible or vague, students, even the highly motivated ones, tend to become more tardy and uninterested. In some cases the researchers point out the importance of rubrics sensitivity to gender, ethnicity and social background as a means of sustaining equity in evaluation and judgment.

What remains poorly researched,however, is the language of rubrics. It is true that researchers talked about clarity but they did not fully explain the role of linguistic competence in making rubrics, a fact which the researcher assumes is of a special importance when accountability and clarity of rubrics is a concern. The following part of this paper shall tackle the issue of language as a major concern in the making of rubrics.

Presentation and Analysis:

In this part of the paper, the researcher will present examples of various rubrics in different disciplines with explanatory discussion focusing on the language of the rubrics presented in an attempt to highlight the importance of linguistic competence for rubric writers on the one hand the general effect of language in the making of comprehensible and clear rubrics for assessors and students alike.

A rubric has four components: criteria, standards, feedback, and an outcome. The criteria defines the quality of work which deserves the grade; the standard is the level of performance for each criterion or what has to be achieved against each criterion; the feedback is the total accomplishment of each assignment compared to the criteria and standards; and the outcome is the given grades for each assignment or piece of work.

In as much as the mathematical issue is concerned, many rubric makers may find the matter easy especially that most tend to use a zero to five grade distribution or a zero to four distributed over five or four standards.

Example:

Standard& Gradecriteria	Poor 0	Fair 1	Good 2	Very good 3	Excellent 4
Spelling					

Standard& Gradecriteria	Poor 1	Fair 2	Good 3	Very good 4	Excellent 5
Spelling					

Some might even make a two-scale rubric that classifies the work into pass or fail. Others may use a three –standard rubric or more. Some rubrics are general holistic and others are very detailed and specific. For each type there are advantages and of course disadvantages.

In the analysis below, I will be quoting sample rubrics for linguistic analysis purposes. My interest is to show to what extent vocabulary can cause incomprehensibility for students and for raters alike.

To begin with, I will try to use a simplified definition of rubric components for assessment or grading. The components of a rubric are basically: what the student is supposed to do= the task, what should be included in the task = what should be considered for grading = criteria, what is the quality of the task= standard or= levels of performance and what grade is assigned to each level of performance. Defining the criteria is the first step. Say for example that the teacher of a certain writing class wants the students to write an essay, a five paragraph essay. This task is usually graded for certain components or criteria: title, organization, thesis statement, mechanics, grammar etc.When the criteria are defined then the teacher has to think of the standard of work that students might produce. Here, it could be easier to say that students will hand in three levels of work quality: poor, good and excellent. Or a teacher may expect students to hand in work that comes in closer but still easily differentiated qualities: poor, fair, good, very good, excellent and outstanding. In either case, the teacher has to follow a certain systematic strategy in describing these standards or quality indicators so that there is no overlapping or confusion for the student and for the grader. This is where language becomes vitally important. In many cases, the vocabulary chosen to describe the level of performance falls into one of the following descriptors: vague, confusing, unclear, can't be measured etc.

Example 1:

Standard / Task	High distinction 85 and above	Distinction 75-84	Average 65-74	Pass 50-54	Fail 0-49
A writing Assignment (holistic grading)	a student demonstrates **exceptional** ability in the assignment and writes with EITHER no spelling or grammar errors OR only minor grammar/spelling errors OR errors that are **potentially** accidental.	a student demonstrates an **excellent** ability in the assignment and writes with EITHER no spelling or grammar errors OR only minor grammar/spelling errors OR errors that are **potentially** accidental.	a student is **average** in the responses to the required task and the writing **may** contain a few more grammatical or spelling errors, but not to the extent that the meaning of **numerous** sentences are impaired.	student responses are **satisfactory** and **adequate** and the writing **may** contain a few more grammatical or spelling errors, but not to the extent that the meaning of **numerous** sentences are impaired.	student responses to task are **poor** and unsatisfactory and the writing has **significantareas** of **unintelligible material** due to **poor** spelling or grammar.

A quick look at the table above and a special emphasis on the highlighted words shows how word choice can cause problems of comprehension and therefore problems of inconsistency and inequity in applying this rubric. The word *exceptional* under the first standard **High Distinction** needs to be explained specially that it is put in face to face comparison with the word *excellent* in the following column **Distinction**. Unless words like distinction and high distinction are first **negotiated** with students or with colleague s and unless several examples from previous students' work samples are presented to show a clear distinction between the two quality descriptors, there will be a huge discrepancy between raters in applying such a rubric. The same can be said for the words *average* and *adequate* or *satisfactory*.

The second problem in word choice is shown in the use of the descriptor " *potentially*" in the first two columns. How would a rater differentiate between a *potential accidental error* and a real/essential/ true error? When can we say that a certain error is accidental or real? And why do we use two descriptors instead of one? Why can't the word accidental be used alone without the word possibly or potentially.

In the third and fourth columns, the rubric has the words "*may*" and "*numerous*" which are so vague for a rater or a student. How would a student realize the use of "may" in the description of the quality of the task in a sentence like, " the writing "*may" contain few more grammatical or spelling errors*" ? What does "may" signify? Does it mean that the writing has some errors? Or does it mean that the essay can have some errors? If "may" is compared to "few" or "no" errors in the first and second standard, then why are these errors so important so as to place the student's work under an Average or Pass standard specially that the errors do not impair the meaning of the sentences? And what is the word " Numerous" supposed to mean in these tow standards?

Finally, when a student's work is considered an F (Fail), what kind of feedback is reflected in the use of the words like "*poor*" and "*significant areas of unintelligible material*"? I personally found this description derogatory and humiliating as well as incomprehensible! For what would "*Poor*' mean to a student who is given an F? and what does a phrase like "*significant areas of intelligible material*"" indicate? Does it mean "Many pieces of work that are not clear? Can't the words " many unclear sentences or paragraphs" be used instead? And why use a word like "*material*" in describing a writing assignment? What does *material* refer to in this context?

The above are examples are meant to illustrate how difficult it is for novice rubric-writers to construct a rubric that is intelligible for both students and raters unless they have a good linguistic background. Furthermore, when we apply a rubric, we try to do some sampling in order to maintain equity.So we start collecting similar pieces or

responses of students' writing and putting them together so that we can later hand in assignments to students with fairly similar feedback notes depending on the category or standard of the final grade assigned to the assignment: the High Distinction together, the Distinction together, the Average together and so on and so forth. With applying a rubric such as the one cited above, one will find it hard to do the sorting process simply because the language used for the descriptors is not making this easy. Ultimately, a rater using this rubric would end up mixing categories or grouping standards: the HDs and the Ds will be considered the same and the Average and the Pass will be assigned a similar grade. The Descriptions of the standards or the quality of the performances is not really creating a clear cut distinction among the various standards. This is a mistake that many rubric makers feel short to notice.

Example 2:

In describing one of the elements of a writing assignment namely Organization, the following rubric gives an introductory statement that aims at clarifying the rubric for both the students and the raters and a short brief description of the grade or the quality of the work. In doing that the raters and the students become familiar with the general standard that guides the grading process in determining the quality of the work, a step that shows some concern about the rubric clarity. However, the rubric falls into the same issue of linguistic ambiguity or vagueness when the specific quality descriptors are introduced and placed parallel to the grades assigned.

Organization	Grade	Description
		General clarification: the ideal essay has an introduction to pull the readers in and prepare them for the main point, a body of sufficient length and structure to explore and support the hypothesis, a conclusion that wraps up the point and shines a light out into the world, and transitions that move the reader smoothly from one part of the essay to another.
	0	No Organization- the essay is a **grab bag** of ideas
	1	Two or more elements mentioned above are **insufficient**
	2	Generally well organized, but lacks **some connection or flow**
	3	**Meets** the ideas stated above
	4	**Meets** the ideas stated above **exceptionally well**

As seen in this table, the rubric uses words that are either incomprehensible, vague and ambiguous or words that may carry negative messages to the students. What would the choice of " Grab Bag" as a descriptor signify? How would a student feel if the work is described as a Grab Bag of ideas?

A "grab bag" is " n. 1. A container filled with articles,…, to be drawn unseen. 2. Slang A miscellaneous collection," (www.thefredictionary.com).

Can a rubric writer choose more descriptive and less humiliating words? Would the description of a Zero grade just say " the ideas are not organized, or the writer doesn't present the ideas according to the standard organization pattern? Moreover, if the students gets 1 point for organization then the description of the work quality says that *'two or more elements mentioned above are insufficient*." The elements mentioned in the clarification are: an introduction, a body, a conclusion and transitions. So if two or more of these elements are insufficient, what is sufficient? It is clear that the rubric wanted to say that the quality of two or more elements is insufficient but the language used can be interpreted differently. Besides, what does the word "*insufficient*" mean? How can a rater determine what is sufficient and what s insufficient? Worse still is how would the student interpret this? Further, abetter off student who may get a 2 for the essay organization will have a description that says "*generally well organized*" and definitely that will create a feeling of joy which would only last for less than a second because the word *but* will kill the joy. What follows *but* is going to be more dramatic because it is so incomprehensible and vague: "(it) *lacks some connection and flow.*" So if the student's writing is *well organized* in general would it fall within a mediocre level as to be given a 2? Specially that what comes after the word *but* only talks about one element in the general clarification at the beginning of the rubric- transitions. In doing so the rubric is including contradictions within the same descriptor that tend to be more confusing rather than educating to the student and hard to be equitably followed by the raters. Lacking "*some connection and flow*" doesn't really state how many connectors are missing if they are present at all! It does not say how can an essay be *well organized* but doesn't have a flow! And what would be hard to grasp is why would an essay that is *well organized* lack *connection and flow*? If it does *lack connection and flow*, it is better described as (poorly) *organized* or (not well) *organized!*

Still, the two top ranking categories (3 & 4) are facing a major word choice problem. In describing each category the rubric uses the phrase:"*Meets the ideas stated above.*"However what ranks one piece of work better than another is the word modifier used in the description of the (4) category: "*exceptionally well.*"How can some writing *meet* the standards while another *meets* them *exceptionally well*? How would a rater notice the difference and where would the rater draw the line? How would I convince the student that the 3-level work quality is just *meeting* the standards while the 4-level one is doing that *exceptionally well*?

Similarly, a simple question that says: State any five of the seven characteristics of a good essay that you have learnt in this course. Can be easily corrected and fairly judged by any teacher using a traditional evaluation procedure simply because the

question structure is designed to test memorization or the ability to name five characteristics of a good essay from seven that were previously studied. A simple look and count of the student's response is so easy and straight- forward. Some raters might pay some close attention to the language of the response and thus take some points off. However, most raters will look directly to the number and names of the characteristics. If there are five different characteristics, then the student gets a full mark.

In attempting a rubric for a simple question like this, an inexperienced rater might fall into a trap like the following:

Level of performance Criteria	Excellent 5	Very Good 4	Good 3	Fair 2	Poor 1	Fail 0
Content of the Answer	Five characteristics are mentioned in *clear and straightforward language with <u>no</u> grammatical mistakes at all*.	Five characteristics are mentioned in *clear and straightforward language with <u>few</u> grammatical mistakes.*	Five characteristics are mentioned *in clear and straightforward language with <u>some</u> grammatical mistakes.*	Five characteristics are mentioned *in clear and straightforward language with <u>many</u> grammatical mistakes.*	Five characteristics are mentioned *in a language <u>full</u> with grammatical mistakes.*	*The five elements are <u>not</u> mentioned*

Even though the rubric above might seem attractive at the first glance, when applied on students' responses the rater will have a very difficult task getting the corrected answers into their proper sampling (files). It would be difficult to grade the fours and the threes and consequently to by the same token the twos and ones simply because the quantifiers used as descriptors are not easily definable. "few" is insufficient, scarce, a small number of , one or two, not many, a handful of, hardly any etc. according to dictionaries. Similarly, "some" is several, a number of, a few, many, numerous etc. while "many" is defined as numerous, more than few, a number of etc. And the word "full" is jam-packed, crowded, crammed, etc. These quantifiers are easily used by many novice rubric makers even though they don't give a definite description of the response that can create equity and or standardized judgment among students. The second pitfall the rubric above falls into is the shift of purpose in defining the objective or the purpose of the question in the first place. A first look at the question would indicate that the question was meant to test students' recall ability – stating five of the seven characteristics studied in the course. The rubric however, doesn't test that. It moves from testing or rating the students memory to the rating of the students' writing and grammar. Only does the last standard (F= Fail=0) talk about the students' ability to remember the five characteristics but it also fails to mention how many were remembered!

Conclusion;

Rubrics are very helpful tools for teachers and students alike. They are wonderful tools for determining the quality of the work and defining the judgment criteria but one should be wary of the fact that rubrics must be constructed with so much patience, vigilance and hard work. It is true that rubrics are becoming more and more popular in schools and universities for the values mentioned earlier. Some schools and educational institutions are even using them for developing instruction and training instructors rather than just as grading tools for the rich feedback that the give to the teacher and the students (Black and Wiliam 1998, shepard 2000). However, with the rising tendency to use rubrics, it is noticeable that more and more teachers and raters are taking an opposite, negative attitude to the use of rubrics. Some have tried rubrics and found them hard to be followed consistently in large classes and within a multitude of assignments. Others have tried building rubrics and found out that the process was even more nerve taking than the process of following the rubrics during the evaluation process. Still other teachers believe that rubrics are making their teaching decisions more fragmental and less dynamic.(Rudner & Schafer 2002)

Finally,the researcher believes that rubrics are sound and powerful tools for evaluation (grading) and assessment (developing instruction) at the same time (Walser, on-line). It is the art of making rubrics that really needs more practice and training. It is true that rubrics look easy on the surface and many a teacher may attempt constructing one in no time. However, when looked at linguistically, a rubric may end up been more unjust, pejorative, confusing and incomprehensible than many traditional evaluation measures.

References

Andrade, H. G. 2000. Using rubrics to promote thinking and learning. *Educational Leadership* 57 (5): 13-18

------------------ 2001. The effects of instructional rubrics on learning to write. *Current Issues in Education* 4 (4), http://cie.ed.asu.edu/volume4/number4.

------------------ 2005. Teaching with rubrics: The good, the bad and the ugly. *College Teaching* 53, no.1:27-30.

 Andrade, H., and Y. Du. 2005. Student perspectives on rubric-referenced assessment. *Practical Assessment Research & Evaluation* 10, no. 5: 1-11.

Arter, J., and J. Chappuis. 2007. *Creating and recognizing quality rubrics*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Arter, J., &McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Thousand Oaks, CA: Corwin Press, Inc.

Black, P., and Wiliam, D. 1998. Inside the black box: Raising standards through classroom assessment. *Assessment and Evaluation in Higher Education* 11 (2): 146-66.

Hafner, J. C., and P. M. Hafner. 2003. Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education* 25, no.12: 1509-28.

Jonsson, A., and G. Svingby. 2007. The use of scoring rubrics: Reliability, Validity and educational consequences. *Educational Research Review* 2: 130-44.

Lapsley, R., and R. Moody. 2007. Teaching tip: Structuring a rubric for online course discussion to assess both traditional and non-traditional students. *Journal of American Academy of Business* 12, no.1: 167-72.

Moni, R. W., E. Beswick, and K. B. Moni. 2005. Using student feedback to construct an assessment rubric for a concept map in Physiology. *Advances in Physiology Education* 29:197-203.

Moskal, B., and J. Leydens. 2000. Scoring rubric development: Validity and Reliability. *Practical Assessment, Research and Evaluation* 7 (10), http://ericae.net/pare/getvn.asp?v=7&n=10

Parkes, K. A. 2006. The effect of performance rubrics on college level applied studio grading. PhD diss., University of Miami. UMI No. 3215237.

Payne, D. A. 2003. *Applied Educational Assessment*. 2nd ed. Belmont, CA: Wadsworth/Thomson Learning.

Popham, J. W. 1997. What's Wrong-and what's right- with rubrics. *Educational Leadership* 55 (2):72-75.

 Reddy, Y. M., & Andrade, H. 2010. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. doi:10.1080/02602930902862859

Roblyer, M. D., &Wiencke, W. R. 2003. Design and use of rubric to assess and encourage interactive qualities in distance courses. *American Journal of Distance Education*, 17, 77-99.

Rudner, L. and W. Schafer. 2002. What Teachers Need to Know About Assessment. Washington, DC: National Education Association. From the free on-line version.

Shepard, L. 2000. The Role of Classroom Assessment in Teaching and Learning. CSE Technical Report 517. CRESST/University of Colorado at Boulder.

Simon, M., and R. Forgette-Giroux. 2001.A rubric for scoring postsecondary academic skills. *Practical Assessment, Research and Evaluation* 7, no. 18. http://pareonline.net/getvn.asp?v=7&n=18.

Stevens, D. D., & Levi, A. J. (2005). Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning. Sterling, VA: Stylus.

Stiggins, R. J. 2001. *Student- involved classroom assessment*. 3rd ed. Upper Saddle River, NJ: Prentice- Hall.

Tierney, R., & M. Simon (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2).

Walser, T. 2010. Using a Standard Rubric to Promote High Standards, Fairness, Student Motivation, and Assessment for Learning. Mountain rise, *The International Journal of the Scholarship of Teaching and Learning* Vol. 6, Issue 3.

Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Portsmouth, NH: Heinemann.